O-MINIMALITY AS AN APPROACH TO THE ANDRÉ-OORT CONJECTURE

by

Thomas Scanlon

Abstract. — Employing a proof technique suggested by Zannier and first successfully implemented by Pila and Zannier to give a reproof of the Manin-Mumford conjecture on algebraic relations on torsion points of an abelian variety, Pila presented an unconditional proof of the André-Oort conjecture when the ambient Shimura variety is a product of modular curves. In subsequent works, these results have been extended to some higher dimensional Shimura and mixed Shimura varieties. With these notes we expose these methods paying special attention to the details of the Pila-Wilkie counting theorem.

Résumé **(O-minimalité comme approche à la conjecture d'André-Oort)**
En utilisant une technique de preuve, suggérée par Zannier et utilisée avec succès par Pila et Zannier, pour prouver la conjecture de Manin-Mumford sur les relations algébriques sur les points de torsion d'une variété abélienne, Pila a présenté une preuve inconditionnelle de la conjecture de André-Oort, lorsque la variété de Shimura ambiante est un produit de courbes modulaires. Ces résultats ont ensuite été étendus à d'autres variétés de Shimura et variétés de Shimura mixtes. Nous exposons ici ces méthodes, en accordant une attention particulière aux détails du théorème de comptage de Pila et Wilkie.

1. Introduction

In the paper [**53**] Pila gave the first unconditional proof of the André-Oort conjecture for mixed Shimura varieties expressible as products of curves. This fact on its own is a remarkable development, but the method of proof, coming as it does from the theory of o-minimality, constitutes a major breakthrough. Zannier had proposed that a theorem of Pila and Wilkie on counting rational points in definable sets in combination with suitable estimates on sizes of Galois orbits could be used to prove theorems in the vein of the André-Oort conjecture, and, indeed, in joint work with Pila [**58**] he implemented this strategy to reprove the Manin-Mumford conjecture. Subsequent work by many authors [**25, 41, 39, 38, 45, 55, 15**] has borne out the

Partially supported by NSF grants FRG DMS-0854998 and DMS-1001550.

promise of this strategy and the pace of the continuing developments suggests that the project has not been played out.

These notes are based on a pair or lecture series I delivered in May 2011, one in Luminy to an assembled group of experts on the André-Oort conjecture with the aim of expositing the applications of the Pila-Wilkie counting theorem to diophantine geometric problems and then a second lecture series in Lyon to model theory students participating in a special Maloa (Mathematical Logic and its Applications) semester with the goal of explaining in detail the counting theorem itself. I have prepared two other accounts of these theorems **[65, 67]** to which the reader is referred for gentler introductions. In this paper, I will follow the proofs of the original papers fairly closely resisting the temptation to "simplify" those arguments. I do not claim any of the results explicated in this paper as my own, though, of course, any errors I may have inadvertently introduced are mine. The principal innovation is to have assembled in one place the key steps in these proofs.

The subject has progressed during the three years since the bulk of this paper was written. Most notably, Tsimerman has completed an unconditional proof of the André-Oort conjecture for $\mathcal{A}_g$, the coarse moduli space of principally polarized abelian varieties of dimension g, using the Pila-Zannier method [**74**]. The present text retains the structure and emphases of its 2012 version, but we conclude with a short section describing the current state of the art.

This paper is organized as follows. In Section 2 we outline the Pila-Zannier strategy. We follow with Section 3 in which we review the basic theory of o-minimality. In Section 4, the technical heart of this paper, we expose in detail the Pila-Wilkie counting theorem. Finally, in Section 5 we present some of the details of the proofs of the diophantine geometric theorems proven with these methods.

Many people have assisted me in writing these notes. I thank in particular Jonathan Pila and Umberto Zannier for their detailed comments on earlier drafts. I thank Matthias Aschenbrenner, Philipp Habegger, and Emmanuel Ullmo for enlightening discussions. I thank the anonymous referees for their close reading and for suggesting innumerable improvements.

2. Overview of the Pila-Zannier strategy

In this section we shall outline the main steps of the Pila-Zannier strategy for proving diophantine geometric theorems. Since the surveys [**65**] and [**67**] are devoted exactly to such outlines, we shall be brief here.

We are interested in proving theorems to the effect that if X is a "special" variety (a Shimura variety, an abelian variety, a moduli space for abelian varieties *et cetera*) and $Y \subseteq X$ is an irreducible closed subvariety containing a Zariski dense set of "special" points (special points in the sense of the theory of Shimura varieties, torsion point, CM-moduli points, *et cetera*), then Y is a "special" subvariety (variety of Hodge-type, group subvariety, submoduli variety, *et cetera*). In practice, we must specify the meaning of the term *special* (as we have suggested parenthetically). The Pila-Zannier strategy takes advantage of the theory of o-minimality which is essentially a theory of real geometry. As such, the technique applies only over the complex numbers, but one could speculate about extensions of the relevant counting theorems to analytic geometric situations over other local fields. Indeed, Comte, Cluckers and Loeser have formulated and proved a version of the Pila-Wilkie counting theorem for sets defined using p-adic analytic functions [**11**]. Subsequently, Chambert-Loir and Loeser have shown how to use this nonarchimedian counting theorem to prove functional transcendence results for maps coming from p-adic analytic uniformizations [**10**].

The first step in the Pila-Zannier strategy is to realize the complex algebraic variety $X(\mathbb{C})$ analytically as a coset space. That is, we seek some complex homogenous space $\mathfrak{X}$ for the action of some (open subgroup of a) real algebraic group $G(\mathbb{R})$ by analytic automorphisms so that some analytic function $\pi : \mathfrak{X} \to X(\mathbb{C})$ represents $X(\mathbb{C})$ as the quotient $\Gamma\backslash\mathfrak{X}$ where $\Gamma \leq G(\mathbb{R})$ is an arithmetic subgroup. For example, if X is an abelian variety over $\mathbb{C}$ of dimension g, then $X(\mathbb{C})$, being a complex torus, may be expressed as $\mathbb{C}^g/\Lambda$ for some lattice Λ. In this case, we would take $\mathfrak{X} = \mathbb{C}^g$ and $G = \mathbb{G}_a^{2g}$ acting via an appropriate real analytic trivialization of $\mathbb{C}^g$ as $\mathbb{R}^{2g}$ for which Λ is identified with $\mathbb{Z}^{2g}$. In the case that $X = \mathbb{A}^1$ regarded as the j-line, then we could take

$$\mathfrak{X} = \mathfrak{h} := \{\tau \in \mathbb{C} \ : \ \mathrm{Im}(\tau) > 0\}$$

to be the upper half plane, $G = \mathrm{PSL}_2$ to be the special linear group acting via fractional linear transformations, $\Gamma := \mathrm{PSL}_2(\mathbb{Z})$ and $\pi := j : \mathfrak{X} \to \mathbb{A}^1(\mathbb{C})$ to be the j-function. The reader might object that as the irreducible closed subvarieties of $\mathbb{A}^1$ are not particularly complicated, being either points or the full space, the problem considered here is trivial. Treat instead $X = \mathbb{A}^N$ regarded as the moduli space of products of N elliptic curves, $\mathfrak{X} = \mathfrak{h}^N$ and $\pi : \mathfrak{X} \to \mathbb{A}^N(\mathbb{C})$ given by

$$(\tau_1, \ldots, \tau_N) \mapsto (j(\tau_1), \ldots, j(\tau_N))$$

Of course, we need to be somewhat careful about how we choose the analytic covering $\pi : \mathfrak{X} \to X(\mathbb{C})$. In particular, we wish to have that the special points in $X(\mathbb{C})$ come from arithmetically simple points in $\mathfrak{X}$. What is meant by *arithmetically simple*? We shall ensure that $\mathfrak{X} \subseteq \mathbb{C}^M$ is an open subset of some complex affine space. Thus, it would make sense to ask whether a point in $\mathfrak{X}$ is rational or algebraic. In practice, we might like for the special points in $X(\mathbb{C})$ to be the images of the rational points in $\mathfrak{X}$, or possibly just algebraic points in $\mathfrak{X}$ of some bounded degree. With our example of X a complex abelian variety, the set of torsion points on $X(\mathbb{C})$ is exactly the image of $\mathbb{Q}^{2g}$ under the analytic covering map. In the case of the j-function giving a covering of $\mathbb{A}^1(\mathbb{C})$ by $\mathfrak{h}$, the set of special points, the j-invariants of elliptic curves with complex multiplication, is the image of the quadratic imaginary numbers. In the general applications of this method, we shall arrange that the set of preimages of special points under the covering map be the set of algebraic points in $\mathfrak{X}$ of degree at most d over $\mathbb{Q}$ for some fixed natural number d.

Once we have found the desired analytic covering map, the problem of describing the set of special points on the algebraic subvariety $Y \subseteq X$ may be converted to the problem of calculating the set of algebraic points of degree $\leq d$ on the analytic variety $\mathfrak{Y} := \pi^{-1}Y$. On the face of it, such a move converts a difficult problem to an intractable one as there is very little in general that one can say about the algebraic points on an analytic variety and the known theorems about the rational points on algebraic varieties are amongst the deepest in all of mathematics. To exploit this translation from special points on an algebraic variety to rational (or algebraic of bounded degree) points on an analytic variety we use the theory of definability in o-minimal structures. The covering map $\pi : \mathfrak{X} \to X(\mathbb{C})$ is almost never definable in a logically tame structure in any sense, but if we were to restrict π to an appropriate fundamental domain $\mathfrak{D} \subseteq \mathfrak{X}$ then the whole situation is often definable in an o-minimal expansion of the real numbers.

In the cases we have been considering, o-minimal definability takes on a very concrete form. Using real and imaginary parts we identify $\mathbb{C}$ with $\mathbb{R}^2$, and hence, $\mathbb{C}^N$ with $\mathbb{R}^{2N}$. By a semialgebraic set we mean a subset of $\mathbb{R}^{2N}$ defined by a finite boolean combination of conditions of the form $f(x_1, \ldots, x_{2n}) \geq 0$ where f is a polynomial with real coefficients. In the cases we have been considering, the fundamental domain $\mathfrak{D}$ may be taken to be semialgebraic. Indeed, when X is a complex abelian variety, then the natural choice for $\mathfrak{D}$ would be $[0,1)^{2g}$. In the case of the covering of the affine line by the j-function, the usual fundamental domain,

$$\mathfrak{D} := \{z \in \mathbb{C} \ : \ \frac{-1}{2} \leq \mathrm{Re}(z) < \frac{1}{2} \text{ and } |z| \geq 1\} \ ,$$

is easily seen to be semialgebraic. We say that a function is *restricted analytic* if it is the restriction of a real analytic function on some open set to a compact box. By an *explicitly definable* function we mean the restriction to a semialgebraic domain of a function built as a composition of polynomials, restricted analytic functions, and the real exponential function. The covering maps we have been considering are explicitly definable. In the case of the covering of an abelian variety $\pi : \mathbb{C}^g \to X(\mathbb{C})$ since π is globally analytic and the fundamental domain $\mathfrak{D}$ is contained in a compact box, one sees that the restriction of π to $\mathfrak{D}$ is already the restriction of a restricted analytic function to a semialgebraic set. For the j-function, one sees from the q-expansion of j, that the restriction of j to $\mathfrak{D}$ may be realized as the restriction to a semialgebraic set of the composite of a restricted analytic function with a function built from restricted analytic functions and the real exponential function.

At this point we may invoke the Pila-Wilkie counting theorem (or one of its refinements) to say something about the distribution of algebraic points on $\widetilde{\mathfrak{Y}} := \mathfrak{D} \cap \mathfrak{Y} = (\pi \restriction \mathfrak{D})^{-1} Y(\mathbb{C})$. The counting theorem says that after accounting for rational points which might lie on semialgebraic sets, there are subpolynomially many rational points on a definable set. Let us be a little more precise.

For a rational number $\frac{a}{b}$ written in lowest terms we define the (multiplicative) height as follows.

$$H(\frac{a}{b}) \ := \ \begin{cases} 0 & \text{if } a = 0 \\ \max\{|a|, |b|\} & \text{otherwise} \end{cases}$$

For an n-tuple $x = (x_1, \ldots, x_n) \in \mathbb{Q}^n$ we define

$$H(x) := \max\{H(x_i) \ : \ i \leq N\} \ .$$

Given a set $X \subseteq \mathbb{R}^n$ and a number $t \geq 1$ we define

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

and

$$N(X, t) := \# X(\mathbb{Q}, t) \ .$$

Finally, we define the algebraic part of X, X^{alg}, to be the union of all connected, positive dimensional semialgebraic subsets of X. The Pila-Wilkie counting theorem asserts that for any $\epsilon > 0$ we have $N(X \smallsetminus X^{\mathrm{alg}}, t) = O(t^\epsilon)$. Various refinements of this theorem are known in which, for example, the set X may be allowed to vary in a family, the exceptional set may be taken to be smaller than the full algebraic part, and one can count points of some bounded degree over the rationals rather than merely rational points. We shall delay our discussion of these refinements until Section 4.

Applying the Pila-Wilkie counting theorem (or its refinement for algebraic points of bounded degree) to our definable set $\widetilde{\mathfrak{Y}}$, we see that there are few, meaning sub-polynomially many, points on $\widetilde{\mathfrak{Y}}$ which are preimages of special points unless $\widetilde{\mathfrak{Y}}^{\mathrm{alg}}$ is large. It is possible for the algebraic part to be large. For example, in the case where X is an abelian variety, if $Y \subseteq X$ is an algebraic subgroup, then $\mathfrak{Y}$ is a linear subspace of $\mathbb{C}^g$. As such, $\widetilde{\mathfrak{Y}}^{\mathrm{alg}} = \widetilde{\mathfrak{Y}}$ (as long as $\dim(Y) > 0$). To continue the argument one must show that this is essentially the only way for $\widetilde{\mathfrak{Y}}^{\mathrm{alg}}$ to be large. Such results are formalized as what we call Ax-Lindemann-Weierstraß theorems, since they generalize the classical theorem of Lindemann and Weierstraß that if $\alpha_1, \ldots, \alpha_n$ are $\mathbb{Q}$-linearly independent algebraic numbers, then $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are algebraically independent, via statements in line with Ax's formal version of Schanuel's conjecture that if $\alpha_1, \ldots, \alpha_n \in t\mathbb{C}[[t]]$ are $\mathbb{C}$-linearly independent power series with zero constant term, then the field $\mathbb{C}(\alpha_1, \ldots, \alpha_n, \exp(\alpha_1), \ldots, \exp(\alpha_n))$ has transcendence degree at least $n + 1$ over $\mathbb{C}$.

Finally, one concludes these arguments by playing a lower bound on the size of the Galois orbit of a special point against the upper bounds coming from the counting theorem. That is, we find a field of definition k for Y and X, some number $\epsilon > 0$, a constant C and some natural number valued measure of complexity of a special point $c(\zeta)$ so that $[k(\zeta) : k] > Cc(\zeta)^\epsilon$ for all special points ζ. For example, in the case that X is an abelian variety one could measure the complexity of a torsion point by its order for which such lower bounds on the degree of a torsion point are known. The trick is to relate the complexity of a special point to the height of a point in $\mathfrak{D}$ mapping to it by π. In the case of torsion points, these quantities are nearly identical, but in general there may be a polynomial distortion. In any case, since every Galois conjugate (over k) of a special point on Y must also lie on Y, the existence of a Zariski dense set of special points on Y would imply that there are more rational (or algebraic of bounded degree) points on $\widetilde{\mathfrak{Y}}$ than the counting theorem would allow unless the algebraic part of $\widetilde{\mathfrak{Y}}$ is large. From an appropriate Ax-Lindemann-Weierstraß theorem we then conclude that some special relations hold on $\widetilde{\mathfrak{Y}}$. Possibly after passing to a quotient, we then conclude that Y must be special.

While the argument outlined above contains many lacuna and elides some essential subtleties, it does accurately reflect the general lines of the o-minimal approach to the André-Oort problems. We shall return to these arguments supplying some details in Section 5.

3. O-minimality

O-minimality was isolated by van den Dries [**79**] as the requisite condition to prove the basic structural results of semialgebraic geometry and then by Steinhorn and Pillay (partially in collaboration with Knight) [**59, 33, 60**] in its logical form. The reader should consult the book by van den Dries [**81**] for a fuller account of o-minimal geometry and the notes for Wilkie's Bourbaki seminar [**87**] for a discussion of the range of geometric theories to which o-minimality applies. In this section, we shall take a middle route, focusing on a specific o-minimal structure, namely the expansion of the ordered field of real numbers by restricted analytic functions and the real exponential function, while recounting some of the general theorems in o-minimality, especially the cell decomposition theorem, and discussing what it would mean to consider o-minimal structures on nonstandard models. This last point may be a significant departure from most expository articles on o-minimality aimed for a general mathematical audience in that part of the characteristically logical nature of the subject will be revealed.

An o-minimal structure is a totally ordered structure (in the sense of first-order logic) all of whose definable, with parameters, subsets are finite unions of points and intervals. The notions of *structure in the sense of first-order logic* and *definability* may very well be somewhat unfamiliar to the reader. Since the idea of *definability* is central to the Pila-Zannier strategy, we shall discuss it in depth. On the other hand, since the general principles of first-order logic and the admittedly stultifying details of the constructions of terms, formulae, languages, interpretations, *et cetera* are spelled out in innumerable textbooks on this subject, we shall content ourselves now with a short summary specialized to our applications.

3.1. A review of first-order logic.

— In specifying a first-order structure one must indicate both a language and an interpretation of this language.

Definition 3.1. — A language $\mathcal{L}$ is given by the data of a set $\mathcal{C}$ of *constant* **symbols**, a set $\mathcal{R}$ of *relation* **symbols**, a set $\mathcal{F}$ of *function* **symbols** and functions $n : \mathcal{R} \to \mathbb{Z}_+$ and $n : \mathcal{F} \to \mathbb{Z}_+$ giving the *arity* of each relation and function symbol.

Remark 3.2. — We have been intentionally vague in saying that a language is *given by* the data listed above. The tuple $\langle \mathcal{C}, \mathcal{R}, \mathcal{F}, n \rangle$ is sometimes called a *signature* or a *vocabulary* and the language is the set of formulae constructed from the signature. Such niceties are of little concern to us here, but the reader may encounter these distinctions in the literature.

Remark 3.3. — We have emphasized that the elements of the sets giving the language are *symbols*. Since in our intended applications these symbols will transparently correspond to actual functions and relations, one might reasonably conclude that such a distinction is merely scholastic, but it is precisely this move in separating syntax from semantics which grounds mathematical logic. In our applications, we shall deduce properties of the sets defined via the standard interpretations of the symbols from results proven about nonstandard interpretations.

Definition 3.4. — Given a language $\mathcal{L}$, an *$\mathcal{L}$-structure* $\mathfrak{M}$ consists of a nonempty set M together with interpretations of each of the symbols from $\mathcal{L}$. That is, for each constant symbol $c \in \mathcal{C}$ we have a distinguished element $c^{\mathfrak{M}} \in M$, for each relation symbol $R \in \mathcal{R}$ we are given a set $R^{\mathfrak{M}} = R(M) \subseteq M^{n(R)}$, and for each function symbol $f \in \mathcal{F}$ we have an actual function $f^{\mathfrak{M}} : M^{n(f)} \to M$.

The structures we shall consider all arise from expansions of the ordered field of the real numbers.

Example 3.5. — The real numbers as an ordered field, often denoted as $\overline{\mathbb{R}}$, is the structure whose underlying universe is the set of real numbers, interprets the single relation symbol $<$ as the usual ordering relation (so that $n(<) = 2$), has a constant symbol r for each $r \in \mathbb{R}$ with $r^{\overline{\mathbb{R}}} = r$, has two function symbols $+$ and $\cdot$, each of arity 2, interpreted by the usual addition and multiplication functions, respectively, and one more function symbol $-$ with $n(-) = 1$ to be interpreted as the additive inverse operation.

Example 3.6. — We may regard the ring $\mathcal{C}(\mathbb{R}, \mathbb{R})$ of continuous real valued functions on $\mathbb{R}$ as an $\mathcal{L}$-structure in the language of Example 3.5 by interpreting the symbols $+$, $\cdot$, and $-$ in the usual way, each constant $r \in \mathbb{R}$ as the corresponding constant function, and $<$ by the partial order $f < g \Longleftrightarrow (\forall x \in \mathbb{R}) f(x) < g(x)$.

While we have only explicitly presented addition and multiplication as distinguished operations in $\overline{\mathbb{R}}$, there are, of course, other functions which we can build from these operations. For example, a polynomial function may be explicitly represented as a composition of the basic functions. In order to maintain the separation of syntax from semantics, we should say that the polynomial may be expressed as a formal composition of the basic function symbols, variables and distinguished constant symbols (that is, it is a *term*) while the function defined by the polynomial is the interpretation of that term which is an honest composition of functions.

Given an $\mathcal{L}$-structure one may define sets within that structure by evaluating equalities and the distinguished relations on the functions defined by terms functions. More precisely, for any $\mathcal{L}$-structure $\mathfrak{M}$, sets of the following kind are the *basic definable sets.*

– Given a pair of terms t and s built from the variables $x_1, \ldots, x_n$, the set
$$(t = s)(\mathfrak{M}) := \{(a_1, \ldots, a_n) \in M^n \ : \ t^{\mathfrak{M}}(a) = s^{\mathfrak{M}}(a)\}$$
is a basic definable set.

– Given a basic relation symbol R of arity m and $t_1, \ldots, t_m$ terms built from the variables $x_1, \ldots, x_n$, then the set
$$(R(t_1, \ldots, t_m))(\mathfrak{M}) := \{(a_1, \ldots, a_n) \in M^n \ : \ (t_1^{\mathfrak{M}}(a), \ldots, t_m^{\mathfrak{M}}(a)) \in R(\mathfrak{M})\}$$
is a basic definable set.

Remark 3.7. — When we write $t^{\mathfrak{M}}$ we mean the honest function on M^n given by composing the basic functions as described in the formal presentation of the term t. The expressions $t = s$ and $R(t_1, \ldots, t_m)$ are called *atomic formulae*. One should indicate the ambient space with the notation for the definable set defined by a given formula since one could always allow further dummy variables.

The class of the quantifier-free definable sets is obtained from the basic definable sets by closing under finite boolean operations. Likewise, the quantifier-free formulae are obtained from the atomic formulae by closing under finite boolean operations. To obtain all of the definable sets, we close off the class of definable sets under the operations of coordinate projections and finite boolean combinations. At the level of syntax, coordinate projections correspond to existential quantifiers. That is, if the set $X \subseteq M^{n+1}$ is defined by the formula $\phi(x_1, \ldots, x_{n+1})$ and $\pi : M^{n+1} \to M^n$ is the projection map $(x_1, \ldots, x_{n+1}) \mapsto (x_1, \ldots, x_n)$, then the image of X under π is defined by the formula $(\exists x_{n+1})\phi$.

Remark 3.8. — What distinguishes *first-order* logic from other forms of symbolic logic is that quantification is permitted only over the *elements* of the structure. That is, one is not permitted to form a formula using syntax like "for every subset X of the structure …" or "there is a finite sequence of elements such that …". Consequently, many sets which are obviously "definable" in the sense that in a natural mathematical language one may specify the set through a rigorous definition are not definable sets. For example, in the structure $\mathbb{R}$ the set of rational numbers is not definable.

Remark 3.9. — Our definition of *definable set* gives the notion of a set definable *without* parameters. In the basic example we have been considering, namely $\overline{\mathbb{R}}$, since every element of the structure is already named by a constant, there is no distinction between this concept of definability and the more general notion of parametrically definable sets though we shall encounter situations requiring this extension. Given a structure $\mathfrak{M}$ in some language $\mathcal{L}$ we may expand $\mathfrak{M}$ to a structure in a larger language having a constant symbol for each element of M. A set is parametrically definable if it is definable in this expanded language. We often drop the word "parametrically".

Example 3.10. — Returning to the real numbers considered as an ordered field, $\overline{\mathbb{R}}$, we see that for any subvariety of affine space, $X \subseteq \mathbb{A}^n_{\mathbb{R}}$, the set of $\mathbb{R}$-valued points, $X(\mathbb{R})$, is definable. Indeed, if the ideal of X is generated by the polynomials $f_1, \ldots, f_m \in \mathbb{R}[x_1, \ldots, x_n]$, then taking $\tilde{f}_i$ to be a term corresponding to the polynomial f_i we have

$$X(\mathbb{R}) = (\tilde{f}_1(x) = 0 \ \& \ \cdots \ \& \ \tilde{f}_m(x) = 0)(\overline{\mathbb{R}})$$

Example 3.11. — More generally, a basic definable set in $\overline{\mathbb{R}}$ is a basic semialgebraic set, that is a set of the form

$$\{(a_1, \ldots, a_n) \in \mathbb{R}^n \ : \ g(a_1, \ldots, a_n) \geq 0\}$$

for some polynomial $g \in \mathbb{R}[x_1, \ldots, x_n]$. A general quantifier-free definable set in $\overline{\mathbb{R}}$ is simply a semialgebraic set. A fundamental theorem of Tarski **[72]** asserts that *every* definable set in $\overline{\mathbb{R}}$ is semialgebraic.

Remark 3.12. — It makes sense to consider formulae with no free variables, also called *sentences*. Reading the definition of the interpretation of a formula, we see that if ϕ is a formula with no free variables and $\mathfrak{M}$ is an $\mathcal{L}$-structure, then $\phi(M)$ should be considered as a subset of M^0, a singleton, and is thus either empty or all of M^0. We say that ϕ is true in $\mathfrak{M}$ or that $\mathfrak{M}$ models ϕ, written $\mathfrak{M} \models \phi$, if $\phi(M) \neq \varnothing$. For a set Σ of sentences, we write $\mathfrak{M} \models \Sigma$ to mean that for every $\phi \in \Sigma$ we have $\mathfrak{M} \models \phi$.

The compactness theorem is the fundamental theorem of model theory. When we enter into some detailed arguments, the compactness theorem will be invoked explicitly to provide nonstandard models and to deduce uniformities from case by base arguments.

Theorem 3.13. — *If $\mathcal{L}$ is any first-order language and Σ is a set of $\mathcal{L}$-sentences having the property that for any finite subset $\Sigma_0 \subseteq \Sigma$ there is some $\mathfrak{M}_0$ with $\mathfrak{M}_0 \models \Sigma_0$, then there is some $\mathcal{L}$-structure $\mathfrak{M}$ such that $\mathfrak{M} \models \Sigma$.*

Various proofs of the compactness theorems are available. For example, it follows from Gödel's Completeness Theorem (N.B.: not the *In*completeness Theorem) **[23]**. Algebraists may be familiar with Łoś's proof using ultraproducts **[36]**.

For the sake of illustration, we show how the compactness theorem may be used to show that finiteness across a class of models implies boundedness.

Proposition 3.14. — *Let T be a theory in some language $\mathcal{L}$, that is, a set of $\mathcal{L}$-sentences, and $\phi(x;y)$ an $\mathcal{L}$-formula. Suppose that for every model $\mathfrak{M}$ of T and every parameter b from M, the definable set*

$$\phi(\mathfrak{M};b) := \{a \in M \ : \ \mathfrak{M} \models \phi(a;b)\}$$

is finite. Then there is a natural number N depending only on ϕ and T so that every such set has size at most N.

Proof. — Consider an expansion $\mathcal{L}'$ of the language $\mathcal{L}$ by new constant symbols $\{c_i : i \in \mathbb{N}\} \cup \{b\}$ and the set of $\mathcal{L}'$ sentences

$$T' := T \cup \{\phi(c_i, b) \ : \ i \in \mathbb{N}\} \cup \{c_i \neq c_j \ : \ i < j\}$$

If there were no bound N as in the statement of the proposition, then each finite subset of T' would have a model as if $T_0 \subseteq T'$ were finite then it would mention new constants from some finite set $\{b\} \cup \{c_i \ : \ i < N\}$ (for some N) and by hypothesis there is some $\mathfrak{M}_0 \models T$ with an element b for which $\phi(\mathfrak{M}_0, b)$ has size at least N. Hence, by the compactness theorem there would be some model $\mathfrak{M} \models T'$ which is, as an $\mathcal{L}$-structure, a model of T, but as $c_i^{\mathfrak{M}} \neq c_j^{\mathfrak{M}}$ for $i \neq j$ and $\{c_i^{\mathfrak{M}} : i \in \mathbb{N}\} \subseteq \phi(\mathfrak{M}, b)$, we see that $\phi(\mathfrak{M}, b)$ is infinite contradicting the hypothesis that every set defined by an instance of ϕ is finite. □

Remark 3.15. — Of course, if in Proposition 3.14 we assumed only that for some fixed structure $\mathfrak{M}$ every definable set of the form $\phi(\mathfrak{M}, b)$ were finite, then we could not conclude that the cardinality of these sets is bounded. It is essential that the finiteness of the definable sets of the form $\phi(\mathfrak{M}, b)$ be known for all models of the theory.

We shall also see the compactness theorem used to produce nonstandard models, that is given an infinite structure $\mathfrak{M}$ we shall find a different structure $\mathfrak{N}$ which satisfies exactly the same sentences.

Definition 3.16. — By the theory of an $\mathcal{L}$-structure $\mathfrak{M}$ we mean the set

$$\mathrm{Th}(\mathfrak{M}) := \{\phi \ : \ \mathfrak{M} \models \phi\} \ .$$

We say that two $\mathcal{L}$-structures $\mathfrak{M}$ and $\mathfrak{N}$ are elementarily equivalent, written $\mathfrak{M} \equiv \mathfrak{N}$, if $\mathrm{Th}(\mathfrak{M}) = \mathrm{Th}(\mathfrak{N})$.

Example 3.17. — Tarski's theorem that every definable set in $\overline{\mathbb{R}}$ is semialgebraic holds in every real closed field, that is, in every ordered field for which one-variable polynomials enjoy the intermediate value property: for each one-variable polynomial $P(x)$, if $a < b$ and $P(a) < 0 < P(b)$, then there is some $c \in (a, b)$ with $P(c) = 0$. Indeed, every real closed field is elementarily equivalent to $\overline{\mathbb{R}}$.

Proposition 3.18. — *If $\mathfrak{M}$ is any infinite $\mathcal{L}$-structure, then there is another structure $\mathfrak{N}$ elementarily equivalent to $\mathfrak{M}$ but of greater cardinality.*

Proof. — Let X be a set of cardinality strictly greater than the cardinality of $\mathfrak{M}$. Let $\mathcal{L}'$ be the expansion of $\mathcal{L}$ by new constants $\{c_x \ : \ x \in X\}$ and consider the theory

$$T' := \mathrm{Th}(\mathfrak{M}) \cup \{c_x \neq c_y \ : \ x \neq y \text{ from } X\}$$

As in Proposition 3.14, any finite subset of T' may be realized in $\mathfrak{M}$ by interpreting the finitely many new constants as distinct elements. Hence, by the compactness theorem there is a model $\mathfrak{N} \models T'$ which satisfies $\mathrm{Th}(\mathfrak{M})$ and has cardinality at least that of X as the elements $c_x^{\mathfrak{N}}$ are distinct. □

We shall end our review of basic model theory here though we shall return to the compactness theorem with the proof of the Pila-Wilkie counting theorem, Theorem 4.6, in Section 4.

3.2. Basic theory of o-minimality. — Let us recall the notion of o-minimality with a formal definition.

Definition 3.19. — Let $\mathcal{L}$ be some language having a binary relation symbol $<$ and possibly other distinguished relation, function and constant symbols. We say that an $\mathcal{L}$-structure $\mathfrak{R} = (R, <, \ldots)$ is *o-minimal* if $<^{\mathfrak{R}}$ is a total order on R and every definable (with parameters) subset of R is a finite union of singletons and intervals.

Remark 3.20. — Since we shall only consider structures in which the interpretation of $<$ defines a total order, we shall drop the superscript and write $<$ rather than $<^{\mathfrak{R}}$ for the total order itself on R.

Remark 3.21. — By an *interval* in a totally ordered set $(R, <)$ we mean a set of the form $(-\infty, a) := \{x \in R \ : \ x < a\}$, $(a, b) := \{x \in R \ : \ a < x < b\}$, $(b, \infty) := \{x \in R \ : \ b < x\}$, or $(-\infty, \infty) := R$ for some elements a and b of R with $a < b$. It is not enough to ask that the set be convex. For example, the set

$\{x \in \mathbb{Q} \ : \ 0 < x < \pi\}$ of rational points in the real interval $(0, \pi)$ is not an interval in $(\mathbb{Q}, <)$.

Remark 3.22. — One might define a structure $(R, <, \ldots)$ to be *strongly* o-minimal if for any elementarily equivalent structure $(R', <, \ldots) \equiv (R, <, \ldots)$ the structure $(R', <, \ldots)$ is also o-minimal. It is a nontrivial theorem, due to Knight, Pillay and Steinhorn [**33**], that if one assumes that the ordering $<$ is dense, then strong o-minimality is equivalent to o-minimality.

In the direct applications of o-minimality considered in this paper, the o-minimal structures will be expansions of the ordered field of real numbers. However, somewhat more general structures, namely elementary extensions of the real numbers, will appear in the proof of the Pila-Wilkie counting theorem. Thus, while the reader would correctly apprehend the intended geometric consequences of o-minimality by restricting attention to real geometry, we cannot impose such a restriction without severely limiting the scope of the arguments.

Let us consider some examples of o-minimal structures on the real numbers.

Theorem 3.23 **(Tarski)**. — *The ordered field of real numbers,* $\overline{\mathbb{R}} := (\mathbb{R}, <, +, \cdot, -)$, *is o-minimal.*

O-minimality had not been defined at the time of Tarski's work on real geometry, but Theorem 3.23 is an immediate corollary of his quantifier elimination theorem for real closed fields: in $\overline{\mathbb{R}}$ every definable set may be defined by a formula without any quantifiers. In one variable, the basic definable sets have the form $\{x \in \mathbb{R} \ : \ f(x) = 0\}$ or $\{x \in \mathbb{R} \ : \ f(x) > 0\}$ for some polynomial $f \in \mathbb{R}[x]$. The zero set of a polynomial is either all of $\mathbb{R}$ if f is the zero polynomial or is finite. By continuity of f and the completeness of $\mathbb{R}$, sets of the second kind are finite unions of open intervals having as their endpoints $\pm\infty$ and some of the zeros of f.

Theorem 3.24 **(Wilkie [85])**. — *The ordered field of real numbers given together with the real exponential function*

$$\mathbb{R}_{\exp} := (\mathbb{R}, <, +, \cdot, \exp)$$

is o-minimal.

As one might expect, the proof of Theorem 3.24 is substantially more difficult than that of Theorem 3.23. From examples constructed by Osgood one knows that there are sets definable in $\mathbb{R}_{\exp}$ using existential quantifiers but which cannot be defined using only quantifier-free formulae [**6**]. For example, consider the following set.

$$\begin{aligned}\{(y_1, y_2, y_3) \in \mathbb{R}^3 \quad : \quad & (\exists x_1 \in \mathbb{R})(\exists x_2 \in \mathbb{R}) \ \ [y_1 = x_1 \ \& \ y_2 = x_1 x_2 \\ & \& \ y_3 = x_1 x_2 e^{x_2} \ \& \ x_1^2 + x_2^2 \leq 1]\}\end{aligned}$$

However, further quantifiers are not necessary. That is, Wilkie established that the theory of $\mathbb{R}_{\exp}$ is *model complete*: every definable set may be defined by a formula of the form $(\exists y_1) \ldots (\exists y_n)\phi$ where no quantifiers appear in the formula ϕ.

The proof of Theorem 3.24 passes through a study of Pfaffian functions and is based in crucial respects on earlier work of Khovanski [**28**] on so-called few-nomials. By a *Pfaffian chain* we mean a finite sequence $f_1, \dots, f_m$ of functions $f_i : \mathbb{R} \to \mathbb{R}$ so that for each i there is some polynomial $P(t, y_1, \dots, y_i) \in \mathbb{R}[t, y_1, \dots, y_i]$ so that the function f_i satisfies the differential equation $f_i'(t) = P(t, f_1(t), \dots, f_i(t))$. A function is *Pfaffian* if it belongs to some Pfaffian chain. The differential equation $y' = y$ satisfied by the exponential function expresses it as a Pfaffian function. As we noted in the previous paragraph, Theorem 3.24 is deep; to do justice to it would lead us too far afield.

A key result used in the proof of Theorem 3.24 is the o-minimality of the theory of the ordered field of real numbers with the *restricted* exponential function, that is, of the structure $(\mathbb{R}, <, +, \cdot, \exp \upharpoonright [0, 1])$. More generally, the theory of the ordered field of real numbers with all restricted analytic functions is o-minimal.

Definition 3.25. — By a *restricted analytic function* we mean a function $f : [0, 1]^n \to \mathbb{R}$ for which there is some open neighborhood $U \supseteq [0, 1]^n$ and a real analytic function $\widetilde{f} : U \to \mathbb{R}$ with $f = \widetilde{f} \upharpoonright [0, 1]^n$.

Remark 3.26. — One might instead define a restricted analytic function to be any function on a compact box which extends to a real analytic function in some neighborhood. One could represent any such function as the composite of a restricted analytic function in the sense of Definition 3.25 with a linear change of variables.

Theorem 3.27 **(van den Dries [80] via Gabrielov [18])**
The structure

$$\mathbb{R}_{\mathrm{an}} = (\mathbb{R}, <, +, \cdot, \{f\}_{f:[0,1]^n \to \mathbb{R} \textit{ restricted analytic}})$$

is o-minimal.

Remark 3.28. — Strictly speaking, if $f : [0, 1]^n \to \mathbb{R}$ is restricted analytic, then when we regard f as an n-ary function symbol, its interpretation in $\mathbb{R}_{\mathrm{an}}$ should be a function whose domain is all of $\mathbb{R}^n$. Since in practice we only care about the interpretation of f on the box $[0, 1]^n$ we are at liberty to define its interpretation outside the box as we see fit and we shall impose the condition that on arguments outside of the box, f takes the value 0.

The counterexample to quantifier elimination for $\mathbb{R}_{\exp}$ mentioned above also provides a counterexample to quantifier elimination for $\mathbb{R}_{\mathrm{an}}$. However, as a Euclidean counterpart to their work on p-adic analytic geometry (from which they deduced some remarkable theorems on the rationality of some Poincaré series attached to p-adic analytic functions) Denef and van den Dries [**16**] proved Theorem 3.27 as a consequence of a quantifier elimination theorem for a variant of $\mathbb{R}_{\mathrm{an}}$ given together with a division operation. Deviating slightly from their formalism in which the universe of the structure is the interval $[0, 1]$ rather than the full set of real numbers, we would allow one new binary function $D : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$D(x,y) \quad := \quad \begin{cases} \frac{x}{y} & \text{if } 0 \leq \frac{x}{y} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Clearly, the function D is already definable, but by including it as a basic function symbol, we allow for more complicated terms. For example, if $f(x)$, $g(z_1, \ldots, z_n)$ and $h(z_1, \ldots, z_n)$ are restricted analytic functions, then the function $f(\frac{g(z_1,\ldots,z_n)}{h(z_1,\ldots,z_n)})$ is represented by a term whereas without the D-function its natural definition would require a quantifier. In this expanded language, Denef and van den Dries prove quantifier elimination by interweaving two different steps: using the Weierstraß preparation and division theorems to reduce questions about the sign of analytic functions to the corresponding question about some associated polynomials and then applying Tarski's elimination of quantifiers theorem for real closed fields, they show how to eliminate an existential quantifier from an existential formula in the language of $\mathbb{R}_{\mathrm{an}}$ (without D) at the cost of introducing some applications of D and then they show how to remove the applications of D at the cost of introducing existential quantifiers. It would seem that that these steps might cancel each other, but the induction is organized so that the formulae become simpler with each iteration. Once the quantifier elimination theorem has been established, o-minimality follows from the Weierstraß division theorem.

Combining the structures $\mathbb{R}_{\mathrm{an}}$ and $\mathbb{R}_{\exp}$ we obtain $\mathbb{R}_{\mathrm{an},\exp}$, the expansion of the ordered field of real numbers by all restricted analytic functions and the global real exponential function. By work of van den Dries and Miller [**83**], this structure, too, is o-minimal. The o-minimality of $\mathbb{R}_{\mathrm{an},\exp}$ admits other proofs. For example, in work generalizing Wilkie's on the exponential function, Speissegger showed that the expansion of any o-minimal structure on the real numbers by Pfaffian functions is still o-minimal [**70**]. Inspired by Écalle's theory of transseries, van den Dries, Macintyre and Marker [**82**] constructed nonstandard models of $\mathbb{R}_{\mathrm{an},\exp}$ via logarithmic-exponential series.

Other o-minimal expansions of the field of real numbers exist. For example, in work of Rolin, Speissegger and Wilkie [**64**] it is shown that o-minimal structures may be constructed from certain classes of quasianalytic functions. Moreover, in the same paper it is shown that there is no maximal o-minimal structure on the real numbers. Here, a maximal o-minimal structure on $\mathbb{R}$ would have been an o-minimal expansion $\mathcal{R}$ of $\overline{\mathbb{R}}$ having the property that for any other o-minimal expansion R of $\overline{\mathbb{R}}$ all of the definable sets in R (and its cartesian powers) are already definable sets in $\mathcal{R}$. It is shown in [**64**] that it is possible to find two functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ so that the two structures $(\mathbb{R}, <, +, \cdot, f)$ and $(\mathbb{R}, <, +, \cdot, g)$ are o-minimal, but the structure $(\mathbb{R}, <, +, \cdot, f, g)$ is not o-minimal. In particular, there is no o-minimal expansion of $\overline{\mathbb{R}}$ in which both f and g are definable.

The known applications of o-minimality to diophantine geometric problems use $\mathbb{R}_{\mathrm{an},\exp}$ and to my knowledge there are no natural problems for which the o-minimality of some of these more exotic structures may be relevant, but nothing rules out the possibility.

Simply from the definition of o-minimality, knowing that a given structure is o-minimal can have striking consequences. For example, under various natural hypotheses, if

$$B := \{a \in \mathbb{R}^n \ : \ |a| \leq 1\}$$

and $f : B \to B$ is a real analytic function with a fixed point at the origin, then there is a function $\Phi : B \times [1, \infty) \to B$ definable in $\mathbb{R}_{\mathrm{an,exp}}$ so that for $a \in B$ and $n \in \mathbb{Z}_+$ one has $\Phi(a, n) = f^{\circ n}(a)$. It then follows that for such an f if $Y \subseteq B$ is a real analytic subvariety and $a \in B$, then either the f-orbit of a is eventually constrained to lie in Y or it meets Y in only finitely many points since the set $\{z \in [1, \infty) \ : \ \Phi(a, z) \in Y\}$ is a finite union of points and intervals so that if $\{n \in \mathbb{Z}_+ \ : \ f^{\circ n}(a) \in Y\}$ is infinite, it contains all numbers beyond some point [**66**].

While the immediate deductions from the definition of o-minimality may be interesting, the real strength of the condition of o-minimality is that from a hypothesis about the simplicity of the structure of definable subsets of the line one may deduce strong regularity properties of the definable sets in any dimension.

The fundamental theorem of o-minimality, the cell decomposition theorem, will appear both explicitly and implicitly throughout our account of the Pila-Wilkie counting theorem.

Convention 3.29. — To avoid some technical issues, we shall insist that every o-minimal structure we consider be an expansion of an ordered field. While there can be good reasons to drop this hypothesis, as, for example, tropical geometry may be regarded as the study of definablity in the structure $(\mathbb{R}, <, +, 0)$ of the real numbers considered as an ordered abelian group, some of the basic results in o-minimality require qualification when stated for weaker structures. Treating our o-minimal structure $(R, <, +, \cdot, \ldots)$ as a topological space with respect to the order topology, it makes sense to speak of the continuity of a definable function. Using the field structure and the usual definition of the derivative as a limit of ratios of partial differences, one sees that it makes sense to speak of the derivative of a definable function.

Before we delve into the details of the cell decomposition theorem, let us observe a very easy, but powerful, result about families of higher dimensional definable sets in an o-minimal field: the existence of definable choice, or *Skolem*, functions.

Proposition 3.30. — *If $(R, <, +, \cdot, -, \ldots)$ is an o-minimal expansion of a field and $Y \subseteq R^{n+m}$ is a definable set regarded as a definable family of definable subsets of R^m parametrized by R^n via*

$$Y_b := \{a \in R^m \ : \ \langle a, b\rangle \in Y\} \ ,$$

then there is a definable function $f : R^n \to R^m$ so that for each $b \in R^n$ if $Y_b \neq \varnothing$, then $f(b) \in Y_b$. (Such a function is called a Skolem function *for Y.)*

Let us illustrate arguments in o-minimality by giving a complete (if very easy) proof of Proposition 3.30.

Proof. — We work by induction on m with the case of $m = 0$ being trivial. For $m = 1$, we break into cases. If $Y_b = R$ or $Y_b = \varnothing$, then we define $f(b) := 0$. Otherwise, by o-minimality ∂Y_b, the boundary of Y_b, is a nonempty finite set and thus has a least element a which is clearly definable from b and Y. If $a \in Y_b$, then we define $f(b) := a$. If $(-\infty, a) \subseteq Y_b$ but $a \notin Y_b$, then define $f(b) := a - 1$, while if $Y_b = (a, \infty)$, define $f(b) := a + 1$. Finally, by o-minimality, if none of the above conditions holds, then for c the second point of ∂Y_b, $(a, c) \subseteq Y_b$ and we may define $f(b) := \frac{1}{2}(a + c)$. For the inductive step from m to $m+1$, consider first the definable set $Z \subseteq R^{n+m}$ defined by

$$Z := \{c \in R^{n+m} \ : \ (\exists y)\langle c, y\rangle \in Y\}$$

By induction, there is a Skolem function $g : R^n \to R^m$ for Z. Regard now Y as a family of subsets of R parametrized by R^{n+m}. From the case of $m = 1$ we obtain a Skolem function $h : R^{n+m} \to R$ for Y (so regarded). Define $f : R^n \to R^{m+1}$ as $f(b) := \langle g(b), h(b, g(b))\rangle$. □

Let us define what is meant by a cell.

Definition 3.31. — Let $(R, <, +, \cdot, \ldots)$ be an o-minimal expansion of an ordered field. We define the class of cells in R^n by recursion on n.

- The singleton R^0 is a cell in R^0.
- If $X \subseteq R^n$ is a cell and $f : X \to R$ is a definable, continuous function, then

$$\Gamma(f)_X := \{(x, y) \in R^{n+1} \ : \ f(x) = y\}$$

  which is the graph of f, is a cell in R^{n+1}.
- If $X \subseteq R^n$ is a cell, $f : X \to R$ and $g : X \to R$ are definable, continuous functions on X for which $f(x) < g(x)$ for all $x \in X$, then

$$(f, g)_X := \{(x, y) \in R^{n+1} \ : \ f(x) < y < g(x)\}$$

  is a cell.
- If $X \subseteq R^n$ is a cell and $f : X \to R$ is a definable, continuous function, then the infinite intervals $(-\infty, f)_X$, $(f, \infty)_X$, and $(-\infty, \infty)_X = X \times R$ are all cells.

With Definition 3.31 in place, we may state the cell decomposition theorem.

Theorem 3.32. — *Let $(R, <, +, \cdot, -, \ldots)$ be an o-minimal expansion of an ordered field and $Y_1, \ldots, Y_m$ be a finite sequence of definable subsets of R^n for some $n \in \mathbb{N}$. Then there is a cell decomposition of R^n subjacent to $Y_1, \ldots, Y_m$. That is, there is a finite set $\mathcal{C}$ of cells in $\mathbb{R}^n$ so that each of the sets Y_i is the disjoint union of those cells in $\mathcal{C}$ which have nonempty intersection with Y_i.*

Remark 3.33. — Theorem 3.32 admits various refinements. For example, if the sequence of sets $Y_1, \ldots, Y_m$ varies in a definable family, then the cell decompositions may also be chosen uniformly from some definable family. In the course of the proof of Theorem 3.32 one shows that every definable function is almost everywhere differentiable. It follows that in Theorem 3.32 for any given degree k of smoothness, one may choose the definable functions involved in the definition of the cells to be $\mathcal{C}^k$. It is not true in all generality that the functions may be taken to be $\mathcal{C}^\infty$. However, each of the structures we have discussed, $\overline{\mathbb{R}}$, $\mathbb{R}_{\text{exp}}$, $\mathbb{R}_{\text{an}}$ and $\mathbb{R}_{\text{an,exp}}$, has *analytic* cell decomposition by which we mean that the functions used to define the cells may be taken to be real analytic. The key step in proving analytic cell decomposition for these structures is to prove that every definable function $f : \mathbb{R} \to \mathbb{R}$ is analytic at all but finitely many points and that, moreover, if f varies in a definable family, then the number of points at which it is not analytic may be bounded uniformly across the family.

One proves Theorem 3.32 by induction on n. The nominal base case of $n = 0$ is trivial, while the true base case of $n = 1$ follows immediately from the definition of o-minimality. However, to carry out the induction one must interweave the proof of cell decomposition with a second theorem on the continuity of definable function for which the one variable case is the most difficult step.

Theorem 3.34. — *Let $(R, <, +, \cdot, -, \ldots)$ be an o-minimal expansion of an ordered field and $f : R^n \to R$ a definable function. Then there is a cell decomposition of the domain so that the restriction of f to each cell is continuous. In fact, for any given $k \in \mathbb{N}$ one may find a cell decomposition of the domain so that the restriction of f to each cell is k-times continuously differentiable. If $n = 1$, after removing finitely many points, the domain may be decomposed into finitely many open intervals on which f is continuous and either strictly monotone or constant.*

As we noted above, the most difficult part of the proof of Theorem 3.34 is to establish the piecewise monotonicity and continuity of univariate definable functions. One argues by noting that various conditions, for example, continuity or f, are definable so that if the lemma were false, then by o-minimality there would be open intervals on which the undesirable property, *e.g.* discontinuity, obtains. Of course, it is possible for a function to be everywhere discontinuous on an interval, but through further considerations of definability one shows that such a condition is impossible for a function definable in an o-minimal structure. Piecewise continuity (or even sufficient smoothness) in dimension $n > 1$ follows from a fairly routine application of cell decomposition in dimension n (which one must prove from Theorem 3.34 and Theorem 3.32 in dimension $< n$ and the monotonicity theorem).

Let us consider the inductive step from dimension n to $n + 1$ for Theorem 3.32. For the sake of illustration, let us work with a single definable set $Y \subseteq R^{n+1}$. We may regard as Y as a family of definable subsets of R indexed by R^n. That is, for $b \in R^n$ let

$$Y_b := \{a \in R \ : \ \langle b, a \rangle \in Y\}$$

By o-minimality, the boundary of each set Y_b, ∂Y_b, is finite. Thus, we have (partially defined) definable functions $f_i : R^n \to R$ (for $i \in \mathbb{Z}_+$) given by

$$b \mapsto \text{ the } i^{\text{th}} \text{ element of } \partial Y_b$$

and $g : R^n \to R$ given by $b \mapsto \max \partial Y_b$. From the dimension n case of Theorem 3.34 we know that g and the f_i's are piecewise continuous. The key to this step of the proof is to show that there is a finite bound on the cardinality of ∂Y_b independent of b. This is achieved by showing that for almost every b (meaning, outside the union of finitely many lower dimensional cells) there is a neighborhood of b over which Y is the disjoint union of the graphs of finitely many continuous functions. As in the proof of piecewise continuity, one argues that if this result were false, then there would be an open ball U in R^n so that this property would fail for every $b \in U$. With some work, one then derives a contradiction.

It is hard to overstate the importance of the cell decomposition theorem, especially for studying the geometry of sets definable in o-minimal structures. For example, it follows that if $(\mathbb{R}, <, +, \times, -, \dots)$ is an o-minimal expansion of the ordered field of real numbers, $Y \subseteq \mathbb{R}^{n+m}$ is a definable subset of $\mathbb{R}^{n+m}$ regarded as a definable family of definable subsets of $\mathbb{R}^n$ via

$$Y_b := \{a \in \mathbb{R}^n \ : \ \langle a, b \rangle \in Y\}$$

then there are only finitely many homemorphism types represented in this family and for each such Y_b the singular homology groups, for instance, are all finitely generated. (Why? Cell decompose $\mathbb{R}^{n+m}$ subjacent to Y. This cell decomposition induces uniform cell decompositions of the fibres and one can read off the homeomorphism type and algebraic topological invariants from a combinatorial description of the cell decomposition.) Working in $\mathbb{R}_{\text{exp}}$ one recovers, qualitatively, a theorem of Khovanski on bounding the number of connected components of systems of solutions of fewnomials, that is systems of equations in a fixed number variables which range over $\mathbb{R}_+$ having a fixed number of monomials but for which the coefficients and exponents are allowed to vary.

We may deduce a simple topological characterization of finite definable sets in any number of variables.

Proposition 3.1. — *Let $(R, <, +, \cdot, 0, 1, \dots)$ be an o-minimal structure and $n \in \mathbb{Z}_+$ a natural number. Then a definable set $X \subseteq R^n$ is finite if and only if X is discrete.*

Proof. — By the cell decomposition Theorem 3.32, X may be expressed as a finite union of cells. Clearly, X is finite if and only if each of those cells is finite and X is discrete if and only if each of the cells is discrete. Arguing by induction on n, one sees that if $C \subseteq R^n$ is a cell, then it is finite if and only if it is a singelton if and only if it is discrete. Indeed, for $n = 1$, a cell is either a singleton or an interval and this claim is obvious in either case. For $n + 1$, if $C = (f, g)_D$ for some cell $D \subseteq R^n$ and continuous definable functions $f : D \to R$ and $g : D \to R$ with $f < g$ on D, then C is infinite and non-discrete. On the other hand, if $C = \Gamma(f)_D$ for some cell $D \subseteq R^n$ and continuous definable function $f : D \to R$, then C is homeomorphic to D and is thus by induction is finite if and only if it is discrete. $\square$

Proposition 3.1 has the curious consequence that finiteness is definable for definable sets in o-minimal structures. More precisely, we have the following corollary.

Corollary 3.35. — *Let T be an o-minimal theory extending the theory of real closed fields in some language $\mathcal{L} = \mathcal{L}(<, +, \cdot, -, 0, 1, \dots)$ and $\phi(x; y)$ be an $\mathcal{L}$-formula in the variables $x = (x_1, \dots, x_m)$ and $y = (y_1, \dots, y_n)$ for some natural numbers m and n. Then there is a formula $\vartheta(y)$ so that in any model $R \models T$ letting $X = \phi(M) \subseteq M^m \times M^n$ be the set defined by ϕ, the set $F := \{b \in R^n \ : \ X_b \text{ is finite} \}$ is defined by ϑ.*

Proof. — By Proposition 3.1, $F = \{b \in R^n \ : \ X_b \text{ is discrete }\}$. We check that the condition of being discrete relative to the topology induced by the product topology of the order topology on the line is definable. Indeed, if $X = \varphi(R)$ where $\varphi(x; y)$ is the formula defining X and $x = (x_1, \ldots, x_m)$ is a tuple of variables ranging over R^m while y ranges over R^n, then F is defined by

$$\begin{aligned} \vartheta(y) \quad := \quad & (\forall x_1, \ldots, x_m)[\varphi(x; y) \to \\ & (\exists \epsilon > 0)(\forall z_1, \ldots, z_n)([\varphi(z; y) \ \& \ \bigwedge_i |z_i - w_i| < \epsilon] \to z = x)] \ . \end{aligned}$$

□

Combining Corollary 3.35 with Proposition 3.14, we deduce that in any o-minimal structure, if one has a uniformly definable family of finite sets, then then there is a uniform bound on the size of the sets in that family.

Let us isolate a specific consequence of the cell decomposition theorem required in the sequel: for those o-minimal structures admitting analytic cell decomposition, every definable set may be covered, up to finitely many points, by the images of nonconstant, definable, real analytic functions on $(-1, 1)$.

Lemma 3.36. — *If $X \subseteq \mathbb{R}^n$ is a set definable in an o-minimal expansion of the ordered field of real numbers admitting analytic cell decomposition, then for all but finitely many points $a \in X$ there is a nonconstant, definable, real analytic map $\gamma : (-1, 1) \to X$ with $a = \gamma(0)$.*

Proof. — We work by induction on n. Of course, if $n = 0$, then the result is trivially true. Analytically cell decompose X. We will show now that for each cell C in this decomposition all but finitely many points are contained in some definable real analytic curve entirely contained in C. If $\dim(C) = 0$, then C itself is a singleton and the result is trivial. Otherwise, C is the graph of a definable, real analytic function $f : C' \to \mathbb{R}$ where $C' \subseteq \mathbb{R}^{n-1}$ is a cell of positive dimension or $C = (f, g)_{C'}$ where $C' \subseteq \mathbb{R}^{n-1}$ is a cell, f and g are real analytic definable functions with domain C' for which $f(x) < g(x)$ for all $x \in C'$. In the first case, the images of the definable, nonconstant, real analytic curves cover all but finitely many points in C'. The curves of the form $t \mapsto (\gamma(t), f(\gamma(t)))$ where $\gamma : (-1, 1) \to C'$ is nonconstant, definable and real analytic will then cover all but finitely many points in C. In the latter case, every points in C is contained in a vertical line segment, which is clearly the image of a nonconstant linear function. □

Remark 3.37. — We shall use Lemma 3.36 only when X is semialgebraic, in which case we may take the real analytic curves $\gamma : (-1, 1) \to X$ to be semialgebraic.

4. Pila-Wilkie counting theorem

In this section, we shall discuss the Pila-Wilkie counting theorem in some detail breaking up the discussion into parts. We begin with a general presentation of the theorem and an outline of its proof. We then embark on a more detailed account of the parametrization theorem whereby it is shown that every bounded definable set may be covered by the images of a small number of balls under maps with small derivatives. We then discuss some results in diophantine approximation to the effect that the rational points on sets admitting such parametrizations are contained in a small number of hypersurfaces. Finally, we explain how to combine these results to deduce the counting theorem and some of its refinements.

Our proof sketch follows the methods from [**48**] rather closely, especially with regards to the parametrization theorem. The diophantine estimates required for this theorem appear in a sequence of papers beginning with a paper of Bombieri and Pila [**7**] and continuing in works of Pila [**47, 49, 50**]. A precursor to the general counting theorem for rational points in definable sets appears in the work of Wilkie [**86**] on diophantine properties of definable sets.

4.1. An overview of the counting theorem. — Let us begin by recalling what we mean by counting rational points. To do so we first recall the notion of the height of a rational number and then extend the height function to finite tuples.

Definition 4.1. — We define the multiplicative height of a rational number by $H(0) := 0$ and $H(\frac{a}{b}) := \max\{|a|, |b|\}$ when a and b are coprime integers. For a tuple $x = (x_1, \ldots, x_n) \in \mathbb{Q}^n$ we define $H(x) := \max\{H(x_i) \ : \ 1 \leq i \leq n\}$.

Definition 4.2. — Let $X \subseteq \mathbb{R}^n$ be any set and $t \in \mathbb{R}_+$ a positive real number. We define $X(\mathbb{Q}, t) := \{x \in X \cap \mathbb{Q}^n \ : \ H(x) \leq t\}$. Note that $X(\mathbb{Q}, t)$ is always a finite set. We define the counting function associated to X by $N(X, t) := \#X(\mathbb{Q}, t)$.

The Pila-Wilkie counting theorem almost says that if $X \subseteq \mathbb{R}^n$ is definable in some o-minimal expansion of the real field, then for any $\epsilon > 0$ there is a constant $C = C(X, \epsilon)$ so that $N(X, t) \leq Ct^\epsilon$ for $t \gg 0$. Of course, such an assertion is plainly false as, for example, the whole space $\mathbb{R}^n$ is definable and $N(\mathbb{R}^n, t)$ grows like a polynomial of degree n in t. There could be subtler reasons for X to contain many rational points. For example, it may happen that X contains a relatively open subset U of some semialgebraic set Y of dimension $m > 1$. If the semialgebraic set Y admits a parametrization by rational functions defined over $\mathbb{Q}$, then $N(U, t)$ will grow like a polynomial of degree k in t. More generally, it might happen that Y contains many rational points for reasons unrelated to rational parametrizations (though Lang's conjectures predict that the presence of many rational points on Y should be explained by the presence of an algebraic group. [See [**34**]]) We obtain the counting theorem once we omit all semialgebraic subsets from X.

Definition 4.3. — We say that the set $Y \subseteq \mathbb{R}^n$ is *semi-algebraic* if it is definable (with parameters) in the structure $\overline{\mathbb{R}} = (\mathbb{R}, <, +, \times)$ of the real numbers considered as an ordered field.

Definition 4.4. — Given a set $X \subseteq \mathbb{R}^n$ we define the algebraic part of X, written X^{alg}, to be the union of all infinite, connected semialgebraic subsets $Y \subseteq X$ of X. The transcendental part of X is $X^{\mathrm{tr}} := X \setminus X^{\mathrm{alg}}$.

If X is a connected, infinite semialgebraic set, then obviously $X = X^{\mathrm{alg}}$. There are other sets which are equal to their algebraic parts. For example, suppose that

$Z \subseteq \mathbb{R}^n$ is any definable set and $X := Z \times \mathbb{R} \subseteq \mathbb{R}^{n+1}$, then $X = X^{\text{alg}}$. In a different direction, if Y is an infinite, connected semialgebraic set and $X \subseteq Y$ is a relatively open subset of Y, then $X = X^{\text{alg}}$. Sets of this last kind play an important role in a refinement of the counting theorem, as we shall see at the end of Section 4.5.

Remark 4.5. — It need not be the case that the algebraic part of a definable set is itself definable. For example the following set is definable in $\mathbb{R}_{\text{exp}}$.

$$X := \{\langle x, y, z\rangle \in \mathbb{R}^3 \ : \ x > 0 \ \& \ x^y = z\}$$

However, its algebraic part consists of those triples $\langle x, y, z\rangle$ for which y is rational. Indeed, for each rational number r, the curve

$$\{\langle x, r, z\rangle \ : \ x^r = z\}$$

is clearly semialgebraic, infinite and connected. To see that every point in X^{alg} is contained in such a curve, one should invoke a functional version of the Gelfond-Schneider theorem.

With our definitions in place, we may state the counting theorem.

***Theorem 4.6* (Pila, Wilkie).** — *Let $X \subseteq \mathbb{R}^n$ be definable in some o-minimal expansion of the real field and let $\epsilon > 0$ be a positive real number. Then there exists a constant $C = C(X, \epsilon) > 0$ so that for $t \geq 1$ we have $N(X^{tr}, t) \leq Ct^\epsilon$.*

The proof of Theorem 4.6 passes through two main steps.

First, we prove a general theorem about parametrizations of definable sets in o-minimal expansions of real closed fields.

Definition 4.7. — Let $k \in \mathbb{Z}_+$ be a positive integer and let $X \subseteq R^n$ be a definable set. We say that the definable function $\phi : (0,1)^m \to X$ is a strong partial k-parametrization of X if $m = \dim(X)$, $\phi \in \mathcal{C}^k$ and $|\phi^{(\alpha)}(x)| \leq 1$ for all $x \in (0,1)^m$ and multi-indices α with $|\alpha| \leq k$. A *strong k-parametrization* of X is a finite set S of strong partial k-parametrizations of X for which $X = \bigcup_{\phi \in S} \phi(0,1)^m$.

Theorem 4.8. — *For every $k \in \mathbb{Z}_+$ every definable set $X \subseteq (0,1)^n$ admits a strong k-parametrization.*

Remark 4.9. — Theorem 4.8 is inspired by similar results of Yomdin [**90, 89**] (extended by Gromov [**24**]) about parametrizations in real algebraic geometry. Strictly speaking, the Gromov-Yomdin theorems are not applicable since we will remove the algebraic part of a definable set before counting the rational points, but as with many arguments in o-minimality, the proof methods in the semialgebraic case extend to the general o-minimal setting.

On the face of it, such a result for definable subsets of $\mathbb{R}^n$ is an easy consequence of the cell decomposition theorem. Indeed, since by the cell decomposition theorem we may express X as a finite union of cells, it suffices to consider the case that X is itself a cell defined by $\mathcal{C}^k$ definable functions. From the description of cells, it is easy to see how to express such a bounded cell as the image of a ball under a function with bounded derivatives. Decomposing the domain into finitely many pieces and making a change of variables, we may arrange that the bound for these derivatives is one.

However, in Section 4.3 we shall establish a version of this parametrization theorem for every o-minimal structure. We then deduce from the compactness theorem a uniform version which has nontrivial content even when we restrict to o-minimal expansions of the real numbers.

To achieve the counting theorem, we prove a general result about rational points on sets parametrized by functions with small derivatives.

Theorem 4.10. — *Given $m < n$, and $d \in \mathbb{Z}_+$ there is a positive integer k and positive constants $\epsilon = \epsilon(m,n,d)$ and $C = C(m,n,d)$ so that if $\phi : (0,1)^m \to \mathbb{R}^n$ is a $\mathcal{C}^k$ function with image X satisfying $|\phi^{(\alpha)}(x)| \leq 1$ for all multi-indices α with $|\alpha| \leq k$, then for any $t \geq 1$ the set $X(\mathbb{Q},t)$ is contained in the union of at most Ct^ϵ (not necessarily irreducible) hypersurfaces of degree d. Moreover, $\epsilon(m,n,d) \to 0$ as $d \to \infty$.*

On its own, Theorem 4.10 has nothing to do with o-minimal structures. However, we shall obtain Theorem 4.6 by applying Theorem 4.10 to the functions parametrizing the definable set X provided by Theorem 4.8.

4.2. Reductions in the parametrization theorem. — Let us now establish some basic reductions and then explain how Theorem 4.6 follows from Theorems 4.8 and 4.10.

As we noted above, if we knew Theorem 4.8 only for subsets of $\mathbb{R}^n$, then the result would be of very little value, but because it will be proven for definable sets in any o-minimal field, a uniform version follows. Let us prove now this formal implication.

In the following proposition, and, indeed, throughout the rest of this section, we work in a general o-minimal structure $(R, <, +, \cdot, -, 0, 1, \ldots)$ expanding a real closed field. In particular, the underlying real closed field R need not be isomorphic to $\mathbb{R}$.

Proposition 4.11. — *Theorem 4.8 implies that if $\{X_b\}_{b\in B}$ is a definable family of definable subsets of $(0,1)^n$ and $k \in \mathbb{Z}_+$ then there are finitely many definable families of definable functions*

$$\{\phi_{i,b} : (0,1)^{\ell_i} \to (0,1)^n\}_{b\in B}$$

so that for each $b \in B$ for some choice of I, $\{\phi_{i,b}\}_{i\in I}$ is a strong k-parametrization of X_b.

Proof. — We shall prove an ostensibly weaker result: there is some $N \in \mathbb{Z}_+$ and N families $\{\phi_{i,c}\}_{c\in C_i}$ of definable functions so that for any $b \in B$ there is some $I \subseteq \{1,\ldots,N\}$ and some $c_i \in C_i$ so that $\{\phi_{i,c_i}\}_{i\in I}$ is a strong k-parametrization of X_b.

We deduce the full result from this weaker version from the existence of Skolem functions, Proposition 3.30. Indeed, for a given $b \in B$ and $I \subseteq \{1,\ldots,N\}$ the condition on $(c_1,\ldots,c_N) \in C_1 \times \cdots \times C_N$ that $\{\phi_{i,c_i}\}_{i\in I}$ is a strong k-parametrization of X_b is definable. Hence, there is a definable function $\rho_I = (\rho_{I,1},\ldots,\rho_{I,N})$ so that for any $b \in B$ if one can find $c_1,\ldots,c_N$ so that $\{\phi_{i,c_i}\}_{i\in I}$ is a strong k-parametrization of

X_b, then $\{\phi_{i,\rho_{I,i}(b)}\}_{i\in I}$ is a strong k-parametrization of X_b. Define $\psi_{\langle i,I\rangle,b} := \phi_{i,\rho_{I,i}(b)}$. Then the conclusion of the proposition holds for $\{\psi_{\langle i,I\rangle,b}\}_{b\in B}$.

Returning to our compactness argument, suppose that the weak version of uniformity fails. Consider the following set Σ of sentences in the expanded language with a constant symbol b for an element of B.

- $\mathrm{Th}(R)$ — the theory of R, every formal first-order sentence true in R
- $b \in B$
- for each finite sequence of formulae $\vartheta_j(x_1,\ldots,x_{\ell_j},y_1,\ldots,y_m;z_1,\ldots,z_{s_j})$ (with $j \leq N$) the assertion $(\forall c_{1,1})\cdots(\forall c_{1,s_1})\cdots(\forall c_{N,1})\cdots(\forall c_{N,s_N})$ the sets defined by $\vartheta_1(x,y;c_1)$, $\ldots$, $\vartheta_N(x,y;c_N)$ do not give a strong k-parametrization of b.

Of course, the expressions in the last item for Σ are not explicitly presented as formal sentences. However, it is a routine matter to formalize such conditions as that a given formula defines the graph of a function, that the derivatives of that function up to some given finite order are bounded by 1 and that X_b is the union of the images of these functions.

Let us check that Σ is satisfiable.

Given any finite subset of Σ we would have to consider only finitely may formulae of the form $\vartheta(x_1,\ldots,x_\ell,y_1,\ldots,y_n,z_1,\ldots,z_s)$.

Let us note that given $\vartheta(x_1,\ldots,x_\ell,y_1,\ldots,y_n,z_1,\ldots,z_s)$ from the family of definable sets $\{\Theta_b\}_{b\in R^s}$, where $\Theta_c = \{\langle x,y\rangle \in R^{\ell+n} \ : \ R \models \vartheta(x,y,c)\}$ we may define the set

$$C := \{c \in R^s \ : \ \Theta_c \text{ is the graph of a function with domain } (0,1)^\ell\}$$

By our hypothesis, there is some $b \in B(R)$ for which no subcollection of the functions possibly defined by these finitely many ϑ give a strong k-parametrization of X_b. Therefore, Σ is consistent and we may find some elementary extension R^* of R and a point $b \in B(R^*)$ satisfying Σ.

The structure R^* is still o-minimal (though it is almost certainly non-archimedian). Hence, by Theorem 4.8 the set X_b admits a strong k-parametrization. For each definable function in R^* there is some formula $\vartheta(x,y,z)$ (here the variables may be tuples) and a choice of a parameter c from R^* for which $\vartheta(x,y;c)$ defines the graph the function. However, since b satisfies Σ one cannot find a finite list of formulae for which by specializing parameters in R^* one obtains a strong k-parametrization of X_b. With this contradiction, we conclude the proof. □

As a simpler reduction, we allow ourselves to replace the condition in the definition of a strong k-parametrization that the derivatives are bounded by one with the condition that they are merely bounded.

Definition 4.12. — We say that a set $X \subseteq R^n$ is *strongly bounded* if there is a positive integer $N \in \mathbb{Z}_+$ so that $X \subseteq [-N,N]^n$. We say that a function is strongly bounded if its graph is strongly bounded.

Remark 4.13. — If $R \subseteq \mathbb{R}$, then for $X \subseteq R$ there is no distinction between X being bounded (meaning that there is some $a \in R$ with $X \subseteq [-a,a]$) and X being strongly bounded. The difference becomes apparent only for non-archimedian models.

Definition 4.14. — Let $X \subseteq R^n$ be a definable set with $\dim(X) = \ell$. For a given positive integer $k \in \mathbb{Z}_+$ we say that the partial parametrization $\phi : (0,1)^\ell \to X$ is a partial k-parametrization if ϕ is $\mathcal{C}^k$ and for each multi-index $\alpha \in \mathbb{N}^\ell$ with $|\alpha| \leq k$ the map $\phi^{(\alpha)}$ is strongly bounded. A *k-parametrization* of X is a finite set of partial k-parameterizations of X the union of whose images covers X.

Arguing as above when we noted that through appropriate changes of variables that every bounded definable set in $\mathbb{R}^n$ admits a strong k-parametrization, we see that to prove the existence of strong k-parametrizations, it suffices to prove that k-parametrizations exists.

Proposition 4.15. — *A definable set $X \subseteq R^n$ admits a strong k-parametrization if and only if it admits a k-parametrization.*

Remark 4.16. — Proposition 4.15 will permit us some flexibility in our proof of Theorem 4.8. Ultimately, it is essential that we work with strong k-parametrizations as the condition that a finite sequence of functions gives a strong k-parametrization is definable in families, but the condition that it gives k-parametrization is not.

Let us now outline the key steps in the proof Theorem 4.8. In order to carry out an inductive argument, we simultaneously prove a theorem about the regularity of definable functions with an especially strong form of the regularity theorem for functions of a single variable. Recall that in the case of the cell decomposition theorem, the analogous regularity theorem took the form of the theorems that every unary definable function is piecewise monotone and in general that every definable function is piecewise continuous where the pieces of the domain are cells. For Theorem 4.8, the regularity theorem concerns *reparameterizations* of definable functions.

Definition 4.17. — Let $X \subseteq R^m$ be a definable set, $\Phi : X \to R^n$ a definable function and S a k-parametrization of X. We say that S is a *k-reparametrization* of Φ if for each $\phi \in S$ the function $\Phi \circ \phi$ is $\mathcal{C}^k$ and for each multi-index α with $|\alpha| \leq k$ the function $(\Phi \circ \phi)^{(\alpha)}$ is strongly bounded.

Remark 4.18. — The definition of a k-reparametrization of Φ is *almost* equivalent to asking that $\{\Phi \circ \phi : \phi \in S\}$ be a k-parametrization of the image of X under Φ. The difference appears when Φ collapses dimensions in the sense that the image of X under Φ has dimension strictly less than that of X.

The reparametrization theorem takes the following form.

Theorem 4.19. — *For every $k \in \mathbb{Z}_+$ every strongly bounded definable function admits a k-reparametrization.*

As we indicated above, the proof of Theorem 4.8 is organized as an induction with which we interweave a proof of Theorem 4.19 starting with a particularly strong form of Theorem 4.19 in dimension one. Let us outline the structure of the proof. For $n, m, k \geq 1$ consider the following assertions.

B(k) If $F : (0,1) \to M$ is a definable function, then there is a k-reparametrization of F so that for each ϕ in the reparametrization either ϕ or $F \circ \phi$ is a polynomial.

R(m,n,k) Every strongly bounded function $F : (0,1)^n \to M^m$ admits a k-reparametrization.

P(n,k) Every strongly bounded definable set $X \subseteq M^n$ admits a k-parametrization.

Note that B(k) is a refined form of R(1,1,k) and that Theorem 4.8 is the assertion that for every n and k the condition P(n,k) holds while Theorem 4.19 is the assertion that for every n, m and k the condition R(n,m,k) holds.

One begins with the observation that P(1,k) is a trivial consequence of the definitions. One then proves B(k) by induction on k (thereby establishing R(1,1,k) for all k). For the base case of B(1) one considers a simple linear change of variables. The inductive case uses a trick to improve the number of derivatives which are strongly bounded through a quadratic change of variables. Through an induction on n one shows that to prove R(m,n,k) it suffices to prove R(1,n,k). One then completes the core of the inductive argument by showing $(\forall k)P(n,k) \to (\forall k)R(n+1,1,k)$ and $(\forall k,n)R(m,n,k) \to (\forall k)P(n+1,k)$.

4.3. The parametrization theorem in more detail. — Let us now spell out some of the details of the proof of Theorem 4.8.

We wish to establish B(k) for all k and we do so in steps. The key to bootstrapping the order of smoothness is a very elementary lemma about precomposing with the squaring function. While I shall omit many of the technical details from the proofs of various lemmata required for the proof of Theorem 4.8, since the proof of this lemma is so clean I cannot resist repeating it.

In the sequel, we shall say that a function f of one variable is weakly increasing if $x \leq y \Rightarrow f(x) \leq f(y)$ while f is weakly decreasing if $x \leq y \Rightarrow f(x) \geq f(y)$.

Lemma 4.20. — *Let $f : (0,1) \to M$ be a definable function with $f \in \mathcal{C}^k$. Suppose that for each $j < k$ the function $f^{(j)}$ is strongly bounded and that the function $x \mapsto |f^{(k)}(x)|$ is weakly decreasing. Define $g : (0,1) \to M$ by $g(x) := f(x^2)$. Then $g^{(j)}$ is strongly bounded for all $j \leq k$.*

Proof. — Using the chain rule, by induction on j we compute

$$g^{(j)} = \sum_{i=0}^{j} \rho_{i,j}(x) f^{(i)}(x^2)$$

where each $\rho_{i,j}(x) \in \mathbb{Z}[x]$ is a polynomial with integer coefficients and $\rho_{j,j} = 2^j x^j$. Clearly, each $\rho_{i,j}$ is strongly bounded on $(0,1)$ and by hypothesis each $f^{(i)}$ is strongly bounded with the possible exception of $f^{(k)}$. The class of strongly bounded functions forms a ring. Hence, to finish the proof of this lemma, we need only show that $2^k x^k f^{(k)}(x^2)$ is strongly bounded.

Take $c \in \mathbb{Z}_+$ so that $|f^{(k-1)}(x)| < c$ for all $x \in (0,1)$. I claim that $|f^{(k)}(x)| \leq 4c/x$ for all $x \in (0,1)$. Assume for the moment that this claim is false. Let $x^* \in (0,1)$ satisfy $|f^{(k)}(x^*)| > 4c/x^*$. By the mean value theorem there would be a point $\xi \in [\frac{x^*}{2}, x^*]$ so that $f^{(k-1)}(x^*) - f^{(k-1)}(x^*/2) = f^{(k)}(\xi) \cdot (x^* - x^*/2)$. We have assumed that $|f^{(k)}|$ is weakly decreasing. Thus, $|f^{(k)}(\xi)| \geq |f^{(k)}(x^*)| > 4c/x^*$. Combining these observations we see:

$$\begin{aligned} 2c &\geq |f^{(k-1)}(x^*) - f^{(k-1)}(x^*/2)| \\ &> (4c/x^*)|x^* - x^*/2| \\ &= 2c \end{aligned}$$

This contradiction establishes the claim. Hence,

$$|2^k x^k f^{(k)}(x^2)| \leq 2^k x^k 4c/x \leq 2^{k+2} c x^{k-1} \leq 2^{k+1} c$$

for $x \in (0,1)$. □

Using Lemma 4.20 one proves B(k), that every strongly bounded one variable function admits a k-reparametrization either by polynomials or for which the reparametrized function is a polynomial.

Lemma 4.21. — *Let $k \in \mathbb{Z}_+$ and $f : (0,1) \to M$ be a definable strongly bounded function. Then f has a k-reparametrization by functions $\phi : (0,1) \to (0,1)$ for which either ϕ or $f \circ \phi$ is a polynomial with strongly bounded coefficients.*

The proof of Lemma 4.21 is also elementary, but requires more bookkeeping than does Lemma 4.20. Allow me simply to indicate the main ideas. One works by induction on k with the case of $k = 1$ following easily from the cell decomposition and monotonicity theorems: we decompose $(0,1)$ into finitely many intervals (a,b) on which f is strictly monotone (or constant) and $|f'(x)| \leq 1$ or $|f'(x)| > 1$. If $|f'(x)| \leq 1$ on (a,b), then $\phi(x) := a+(b-a)x$ parametrizes (a,b) and $|(f \circ \phi)'(x)| < 1$. If $|f'(x)| > 1$, then we see that o-minimality together with boundedness shows that $c := \lim_{x \to a^+} f(x)$ and $d := \lim_{x \to b^-} f(x)$ exist. If we define $\phi(x) := f^{-1}(c+(d-c)x)$, then $|(f \circ \phi)'| < 1$. In the inductive case, we use the inductive hypothesis together with the cell decomposition and monotonicity theorems to arrange that the composites of f with the various k-reparametrization functions have monotone $(k+1)^{\text{st}}$ derivatives on each interval in the decomposition. Reversing the parametrization of the intervals if need be, we may assume that Lemma 4.20 applies and thereby obtain a $(k+1)$-parametrization by precomposing with the squaring function.

With B(k) established for all k, namely that if $F : (0,1) \to M$ is a definable function, then there is a k-reparametrization of F so that for each ϕ in the reparametrization either ϕ or $F \circ \phi$ is a polynomial, we have the base case for Theorem 4.19. On the other hand, the base case for Theorem 4.8, namely $(\forall k)P(1,k)$: Every strongly bounded definable set $X \subseteq M$ admits a k-parametrization, is an immediate consequence of o-minimality.

To complete the proof, one argues through strong induction showing that $(\forall k)(\forall i \leq n)R(m,i,k) \to (\forall k)P(n+1,k)$ and that $(\forall k)(\forall i \leq n)P(i,k) \to (\forall k)R(m,n,k)$. That is, one shows that the reparametrization theorem in dimension n implies the parametrization theorem in dimension $n+1$ which thereby implies the reparametrization theorem in dimension $n+1$.

Deducing the parametrization theorem in dimension $n+1$ from the reparametrization and parametrization theorems in dimension n is straightforward, once one has established the technical, though elementary, fact that if one knows how to find k-reparametrizations of functions into R, then one can find k-reparametrizations for functions into R^m for any m.

Using the cell decomposition theorem, one reduces to the case of finding a k-parametrization for a cell, X. The case of $X = (f, g)_Y$, that is, X is an interval over the ℓ-cell Y in n-space, is slightly more difficult than the case of X being the graph of a strongly bounded function on some cell. By the inductive hypothesis, we can find a k-parametrization of Y and then for each function ϕ in that parametrization we can find reparametrizations of $\langle f \circ \phi, g \circ \phi \rangle$. Taking the functions

$$(x_1, \dots, x_\ell, x_{\ell+1}) \mapsto \langle \phi(x_1, \dots, x_{\ell+1}), f \circ \phi \circ \psi(x_1, \dots, x_\ell) + x_{\ell+1} g \circ \phi \circ \psi(x_1, \dots, x_\ell) \rangle$$

as ϕ ranges through the functions in a k-parametrization on Y and ψ ranges through the functions in a k-reparametrization of $\langle f \circ \phi, g \circ \phi \rangle$ gives the desired k-parametrization of X.

The more difficult step is to prove the reparametrization theorem in dimension $n + 1$. Using cell decomposition and induction, one reduces to the problem of finding a k-reparametrization of some strongly bounded definable function $f : (0, 1)^{n+1} \to R$. Treating the $(n + 1)^{\mathrm{st}}$ variable as a parameter, we may regard f as a definable family $\{f_u : (0, 1)^n \to R\}_{u \in R}$ of strongly bounded functions of n variables defined by $f_u(x_1, \dots, x_n) := f(x_1, \dots, x_n, u)$. By induction and a compactness argument like that for Proposition 4.11, one shows that there is a uniform family of k-reparametrizations $\phi_u : (0, 1)^n \to (0, 1)^n$ of f_u. Returning to the original function f, we have a something like a reparametrization of f given by $\widetilde{\phi} : (x_1, \dots, x_n, x_{n+1}) \mapsto (\phi_{x_{n+1}}(x_1, \dots, x_n), x_{n+1})$, in the sense that the partial derivatives of $f \circ \widetilde{\phi}$ of order up to k with respect to $x_1, \dots, x_n$ are strongly bounded. Our work is to increase the order of strong boundedness with respect to the last variable. Ideas from o-minimality are used throughout, including some we have already seen (existence of Skolem functions, cell decomposition, compactness to deduce uniformities) and one of a different character: o-minimality implies strong regularity properties of limits. The specific result used is the following lemma.

Lemma 4.22. — *Let $\{F_t\}_{t \in (0,1)}$ be a definable family of strongly bounded functions $F_t : X \to R$. Suppose that each F_t is $\mathcal{C}^k$ and that the derivatives of F_t of order up to and including k are strongly bounded. Then the rule $F_0(x) := \lim_{t \to 0^+} F_t(x)$ defines a function of class $\mathcal{C}^{k-1}$ all of whose derivatives up to order $k - 1$ are strongly bounded.*

The key observation for the proof Lemma 4.22 is that for any $x \in X$, the function $t \mapsto F_t(x)$ is a strongly bounded one-variable function so that by the monotonicity theorem, $\lim_{t \to 0^+} F_t(x) \in R$ and therefore $F_0 : X \to R$ is a well-defined, definable function. The rest of the proof is elementary analysis.

Remark 4.23. — Lemma 4.22 may be regarded as part of the theory of *definably Banach* spaces as developed by Thomas [**73**].

To finish the proof of the reparametrization theorem, it suffices to show that possibly at the cost of restricting to a dense, open, definable subset of the domain, one may reparametrize, changing only the last coordinate, so as to guarantee that the reparametrized function has strongly bounded first derivatives. The general result we require follows by applying this result to the derivatives.

Lemma 4.24. — *Suppose that $U \subseteq (0,1)^{m+1}$ is a dense, open definable subset of the open box $(0,1)^{m+1}$ and that $f : U \to R^N$ is a definable strongly bounded function having the property that for each $i \leq m$ (NB: not necessarily $i = m+1$) we know that $\frac{\partial f}{\partial x_i}$ is continuous and strongly bounded on U. Then for each $k \geq 2$ there is a $k-1$-parametrization S_k of $(0,1)$ and a dense, open definable set $V \subseteq U$ such that for any $\phi \in S_k$ if $I_\phi(x_1, \ldots, x_{m+1}) := (x_1, \ldots, x_m, \phi(x_{m+1}))$, we know that $I_\phi(V) \subseteq U$, $f_\phi := f \circ I_\phi \in \mathcal{C}^1(V)$, and all of the first derivatives of f_ϕ are strongly bounded.*

Proof. — To ease notation, we shall work with $N = 1$.

By o-minimality, we may find a dense, open definable $W \subseteq U$ for which the restriction of f to W is $\mathcal{C}^1$.

For t and y from $(0,1)$ let

$$W_t(y) := \{x \in (0,1)^m \ : \ \mathrm{dist}(x, [0,1]^m \times \{y\} \smallsetminus W) \geq t\}$$

Note that $W_t(y)$ is a relatively closed subset of a closed box. Thus for fixed y and t, the function $x \mapsto |\frac{\partial f}{\partial x_{m+1}}(x,y)|$ is defined, continuous, and takes a maximum value (provided that $W_t(y) \neq \varnothing$). Using the definability of Skolem functions, there is a definable function $s_t(y)$ (of the two variables t and y) so that if $W_t(y) \neq \varnothing$, then $x \mapsto |\frac{\partial f}{\partial x_{m+1}}(x,y)|$ attains its maximal value at $x = s_t(y)$.

Note that for every $t \in (0,1)$, $y \in (0,1)$ and $x \in W_t(y)$ we have

$$|\frac{\partial f}{\partial x_{m+1}}(s_t(y), y)| \geq |\frac{\partial f}{\partial x_{m+1}}(x,y)|$$

Consider now the definable family of definable functions

$$\{g_t : (0,1) \to (0,1)^m \times R\}_{t \in (0,1)}$$

given by

$$g_t(y) := \langle s_t(y), f(s_t(y), y) \rangle$$

(When s_t is undefined, output $\langle \frac{1}{2}, \ldots, \frac{1}{2} \rangle$.)

As each g_t is a function of a single variable and strongly bounded at that (as $s_t(y)$ takes values in $(0,1)^m$ and f is itself strongly bounded), for each t we have a k-reparameterization S_t of g_t. Arguing again by compactness, we may assume that the functions making up S_t vary in a definable family.

Let now S_0 be the limit of S_t. Then by Lemma 4.22, S_0 is a $(k-1)$-parametrization of a cofinite subset of $(0,1)$. Lemma 4.22 gives that all of the derivatives of (the components of) S_0 up to order $k-1$ are strongly bounded. That the images cover a cofinite subset of $(0,1)$ requires an additional argument. Each function in S_t is piecewise continuous as a two variable function. It follows that S_0 omits at most finitely many points. $\square$

It is now a routine matter using cell decomposition and induction to deduce the reparametrization theorem from Lemma 4.24.

4.4. Diophantine approximation. — As we noted above, there are two fundamentally distinct parts of the counting theorem: the parametrization theorem in o-minimal theories and then some general results about the distribution of rational points in sets which admit parametrizations. In this section, we shall discuss these diophantine approximation arguments.

Let us introduce some basic notation and some relevant combinatorial quantities.

Notation 4.25. — If $\mu = \langle \mu_1, \ldots, \mu_n \rangle \in \mathbb{N}^n$ is a multi-index and $x \in \mathbb{R}^n$, then we write x^μ for $\prod_{i=1}^n x_i^{\mu_i}$. We define $|\mu| := \sum_{i=1}^n \mu_i$.

Note that a polynomial in n-variables of total degree at most d may be expressed as a sum $\sum_{|\alpha| \leq d} c_\alpha x^\alpha$.

Notation 4.26. — Given $n, d \in \mathbb{N}$ we define $D(n,d)$ to be the dimension of the space of polynomials of total degree at most d in n variables. Equivalently,

$$D(n,d) = \#\{\alpha \in \mathbb{N}^n \ : \ |\alpha| \leq d\}$$

We define $L(n,d)$ to be the dimension of the space of polynomials of total degree exactly d in n variables. We define

$$V(n,d) := \sum_{i=0}^{d} L(n,i)i$$

The quantities $L(n,d)$ and $D(n,d)$ admit other descriptions.

$$\begin{aligned} D(n,d) &= \binom{n+d}{d} = \sum_{i=0}^{d} L(n,i) \\ L(n,d) &= \binom{n+d-1}{n-1} \end{aligned}$$

Definition 4.27. — Given $m, n, d \in \mathbb{Z}_+$ there is a unique $b := b(m,n,d)$ for which $D(m,b) \leq D(n,d) < D(m,b+1)$. We define

$$\begin{aligned} B(m,n,d) &:= \sum_{i=0}^{b} L(m,i)i + (D(n,d) - \sum_{i=0}^{b} L(m,i))(b+1) \\ &= V(m,b) + (D(n,d) - D(m,b))(b+1) \end{aligned}$$

For us, it the limiting behaviors of these quantities which are relevant.

With m and n fixed, as $d \to \infty$ we have

$$\begin{aligned} b(m,n,d) &= (\frac{m!d^n}{n!})^{\frac{1}{m}}(1+o(1)) \\ B(m,n,d) &= \frac{1}{(m+1)(m-1)!}(\frac{m!}{n!})^{(m+1)/m} d^{n(m+1)/m}(1+o(1)) \\ V(n,d) &= \frac{1}{(n+1)(n-1)!} d^{n+1}(1+o(1)) \end{aligned}$$

In particular that if we define

$$\epsilon(m,n,d) := \frac{dD(n,d)}{B(m,n,d)}$$

then provided $m < n$, we have $\lim_{d\to\infty} \epsilon(m,n,d) = 0$.

Let us note a very easy but useful linear algebraic proposition.

Proposition 4.28. — *Let R be a field, n and d two positive integers and $S \subseteq R^n$. We suppose that for each subset $S_0 \subseteq S$ of cardinality $D(n,d)$ the determinant of the matrix (Q^μ) whose columns are indexed by multi-indices μ with $|\mu| \leq d$ and whose rows are indexed by $Q \in S_0$ vanishes. Then there is a nonzero polynomial $f \in R[x_1,\dots,x_n]$ of total degree at most d which vanishes on every $Q \in S$.*

Proof. — Without loss of generality, we may assume that $\#S \geq D(n,d)$.

Let V be the vector space $R^{\{\mu\in\mathbb{N}^n \,:\, |\mu|\leq d\}} \cong R^{D(n,d)}$ and let W be the subspace of V generated by $(Q^\mu)_{\{\mu\in\mathbb{N}^n \,:\, |\mu|\leq d\}}$ as Q ranges over S. Our hypothesis is that $\dim W < D(n,d)$. Let $S_0 \subseteq S$ so that $\{(Q^\mu) \,:\, Q \in S_0\}$ is a basis for W. Let $S_1 \supsetneqq S_0$ be a subset of S containing S_0 and having $\#S_1 = D(n,d)$. Since $\det(Q^\mu) = 0$ the columns of this matrix must be dependent. That is, we can find $(c_\mu) \in V \smallsetminus \{(0,\dots,0)\}$ so that for each $Q \in S_1$ we have $\sum c_\mu Q^\mu = 0$. That is, if we let $f(x_1,\dots,x_n) := \sum c_\mu x^\mu \in R[x_1,\dots,x_n]$, then f vanishes on each $Q \in S_1$. Let now P be an arbitrary element of S. Then as $\{(Q^\mu) \,:\, Q \in S_0\}$ is a basis for W, then there are scalars $\lambda_Q \in R$ for $Q \in S_0$ so that $(P^\mu) = \sum \lambda_Q (Q^\mu)$. Thus, $f(P) = \sum c_\mu P^\mu = \sum_{|\mu|\leq d} c_\mu \sum_{Q\in S_0} \lambda_Q Q^\mu = \sum_{Q\in S_0} \lambda_Q \sum_{|\mu|\leq d} c_\mu Q^\mu = \sum_{Q\in S_0} \lambda_Q f(Q) = 0$. □

Remark 4.29. — The converse to Proposition 4.28 holds as well: If there is a nonzero polynomial $f \in R[x_1,\dots,x_n]$ of total degree at most d which vanishes on every element of the set S, then for every $S_0 \subseteq S$ of cardinality $D(n,d)$ the determinant of the matrix (Q^μ) as Q ranges through S_0 and μ ranges through the multi-indices with $|\mu| \leq d$ vanishes. Indeed, writing $f = \sum c_\alpha x^\alpha$ we see that the vector (c_α) witnesses a linear dependence amongst the columns of this matrix.

With the next lemma we show that points on the image of a set parametrized by a function with small derivatives must be very close to lying on an algebraic hypersurface.

In the midst of the proof we use an elementary fact that while the determinant of a sum of linear maps cannot be computed in terms of the determinants of the individual maps, it can be expressed using exterior products. (I thank Emmanuel Breuillard for drawing my attention to the papers **[1]** and **[63]** in which formulae for determinants of sums are worked out explicitly.) Recall that if $L : V \to V$ is a linear map from an n-dimensional vector space back to itself, then $\Lambda^n(L) = (\det(L))\,\mathrm{id}_{\Lambda^n V}$. Suppose now that $\psi_i : V \to V$ (for $i \leq r$) is a finite sequence of linear maps on this vector space, then

$$\Lambda^n\Big(\sum_{i=1}^r \psi_i\Big) = \sum_{\tau\in\{1,\dots,r\}^n} \psi_{\tau_1} \wedge \cdots \wedge \psi_{\tau_n} \tag{1}$$

Taking into account the ranks of each of the maps ψ_i, we see that the term corresponding to τ is nonzero only if $\#\{i \ : \ \tau_i = j\} \leq \mathrm{rank}(\psi_j)$ for each $j \leq r$.

To read Equation 1 in terms of determinants of matrices one should fix a basis $e_1, \ldots, e_n$ of V, and write $\psi_i(e_j) = \sum a_{j,k}^{(i)} e_k$ to obtain the following expression.

$$\det(\sum_{i=1}^{r} (a_{j,k}^{(i)})) = \sum_{\tau \in \{1,\ldots,r\}^n} \det((a_{j,k}^{(\tau_j)})) \tag{2}$$

Combining Equation 2 with our observation about ranks, we obtain a simplified expression for the determinant of a sum.

$$\det(\sum_{i=1}^{r} (a_{j,k}^{(i)})) = \sum_{\substack{\tau \in \{1,\ldots,r\}^n \\ \#\{\ell \ : \ \tau_\ell = i\} \leq \mathrm{rank}((a_{j,k}^{(i)})) \\ \text{for all } i \leq r}} \det((a_{j,k}^{(\tau_j)})) \tag{3}$$

With these observations in place we may prove our key estimate.

Proposition 4.30. — *Let $m, n, d \in \mathbb{Z}_+$ and set $k := b(m,n,d) + 1$. Then there is a constant $K = K(m,n,d)$ so that if $f : (0,1)^m \to \mathbb{R}^n$ is a strong k-parametrization of its image, $r \in (0,1)$ is a positive real number less than one, and $Q_0, Q_1, \ldots, Q_{D(n,d)} \in (0,1)^m$ are points for which the distance between Q_i and Q_0 is at most r, then the absolute value of the determinant of the matrix $(f(Q_i)^\mu)$ whose rows are indexed by $1 \leq i \leq D(n,d)$ and whose columns are indexed by $\mu \in \mathbb{N}^n$ with $|\mu| \leq d$ is at most $Kr^{B(m,n,d)}$.*

Proof. — Expanding each $f(Q_i)$ as a Taylor polynomial around Q_0 up to order b and then with an order $b+1$ remainder term, we see that $f(Q_i)$ may be expressed as a polynomial in $(Q_i - Q_0)$ of total degree $b+1$ with coefficients bounded by one. It then follows that for any $\mu \in \mathbb{N}^n$ with $|\mu| \leq d$ we may express $f(Q_i)^\mu$ as a polynomial in $(Q_i - Q_0)$ with coefficients bounded by a constant depending only on n and d (d^n would do). Let us write the matrix as $(f(Q_i)^\mu) = (\sum_{j=0}^{b+1} R_{i,\mu}^{(j)})$ where for $j < b+1$ each entry of $R^{(j)}$ is a homogeneous polynomial of degree j in $(Q_i - Q_0)$ and each entry of $R^{(b+1)}$ is a sum of homogenous polynomials of degree greater than b. Observe that for each $j \leq b$ the matrix $R^{(j)}$ has rank at most $L(m,j)$ as the space of homogeneous polynomials of degree j in m variables has dimension $L(m,j)$.

From Equation 3 we see that

$$\det(f(Q_i^\mu)) = \sum_{\substack{\tau \in \{0,\ldots,b+1\}^D \\ \#\{\ell \ : \ \tau_\ell = j\} \leq L(m,j) \\ \text{for all } j \leq b}} \det((R_{i,\mu}^{(\tau_i)}))$$

In general, the contribution from τ is bounded by a constant (depending just on m, n, and d) times $r^{|\tau|}$. The smallest that this exponent could be is when $\#\tau^{-1}\{j\} = L(m,j)$ for all $j \leq b$. That is, when the exponent is

$$\sum_{j=0}^{b} jL(m,j) + (b+1)(D(n,d) - D(m,b)) = B(m,n,d)$$

Taking the constants we have considered thusfar and multiplying by $(b+1)^n$, for instance, we obtain the desired result. $\square$

Combining Propositions 4.28 and 4.30 we prove that the rational points in the image of a strong k-parametrization are constrained to a small number of algebraic hypersurfaces.

Theorem 4.31. — *Given $m < n \in \mathbb{Z}_+$ and $d \in \mathbb{Z}_+$. Then there exists a number $k = k(m,n,d) \in \mathbb{Z}_+$ and constants $\epsilon(m,n,d)$ (as defined above) and $C = C(m,n,d)$ so that whenever $f : (0,1)^m \to \mathbb{R}^n$ is a strong k-parametrization of its image, called X, and $t \in \mathbb{Z}_+$, then $X(\mathbb{Q},t)$ is contained in the union of Ct^ϵ (possibly reducible) hypersurfaces of degree d.*

Proof. — Let K be the constant of Lemma 4.30. Given t, set $r := t^{\frac{-dD(n,d)}{B(m,n,d)}} K^{\frac{-1}{B(m,n,d)}}$. Let us observe that if $Q_1, \ldots, Q_{D(n,d)} \in (0,1)^m$ are points for which $f(Q_i) \in X(\mathbb{Q},t)$, then there is an integer s with $|s| \leq t^{dD(n,d)}$ so that $\det(f(Q_i)^\mu) \in \frac{1}{s}\mathbb{Z}$. In particular, if we further require the Q_i's to belong to a ball of radius r, then this determinant would have size bounded by $Kr^B < (\frac{1}{t})^{dD(n,d)}$ by Lemma 4.30. Hence, this determinant vanishes. We may cover $(0,1)^m$ by a constant multiple of $\frac{1}{r^m}$, or if you prefer, a constant multiple of $t^{\frac{dD(n,d)}{m}}$ balls of radius r. For each such ball, we know by Proposition 4.28 there is a single hypersurface of degree d containing all rational points of height at most t which are in the image of that ball. $\square$

4.5. Completing the proof of the counting theorem and its refinements. — We conclude the discussion of the Pila-Wilkie counting theorem. In this section, we finish its proof by combining Theorems 4.8 and 4.31.

To carry out the induction to prove Theorem 4.6, it is better to prove a uniform version.

Theorem 4.32. — *Let $\{X_b\}_{b\in B}$ be a definable family of definable subsets of $\mathbb{R}^n$ in some o-minimal expansion of the real field and let $\epsilon > 0$ be any real number. Then there is a definable family of sets $\{Y_b\}_{b\in B}$ and a constant C depending just on ϵ and the family so that for every $b \in B$ we have $Y_b \subseteq X_b^{alg}$ and $N(X_b \setminus Y_b, t) \leq Ct^\epsilon$ for $t \geq 1$*

Let us note what we may conclude by combining Proposition 4.11 with Theorem 4.31.

Proposition 4.33. — *Given a definable family $\{X_b\}_{b\in B}$ of definable subsets of $(0,1)^n$ each of dimension strictly less than n and $\epsilon > 0$ there is a number $d \in \mathbb{Z}_+$*

and a constant C depending only on ϵ and the family so that for each $b \in B$ the set $X_b(\mathbb{Q}, t)$ is contained in the union of Ct^ϵ hypersurfaces of degree d.

Proof. — Let us prove Theorem 4.32 by induction n and then on the fibre dimension of $\{X_b\}_{b \in B}$.

Through some elementary considerations of the invariance of the height function under additive and multiplicative inverses, one sees that it suffices to consider the case that $X_b \subseteq (0,1)^n$.

Using uniform cell decomposition, one sees that it suffices to consider the case that each X_b is a cell. Considering projections and cell decompositions again, one sees that it suffices to consider the case that $X_b = \Gamma(f_b)_{Z_b}$, the graph of a definable continuous function $f_b : Z_b \to (0,1)$ where $Z_b \subseteq R^{n-1}$ is an open cell.

By Theorem 4.31 there is a number $d \in \mathbb{Z}_+$ and a constant C depending just on the family and on ϵ so that for each $b \in B$ and $t \geq 1$ the set $X_b(\mathbb{Q}, t)$ is contained in the union of $Ct^{\frac{\epsilon}{2}}$ hypersurfaces of degree d.

The family of hypersurfaces of degree d is naturally a definable family itself $\{H_c\}_{c \in B'}$ (parametrized by $\mathbb{P}^{D(n,d)-1}(\mathbb{R})$, for example).

Let us note that if X_b is contained in some hypersurface, then $X_b = X_b^{\mathrm{alg}}$ as necessarily f_b would be the restriction of an algebraic function to an open set. Thus, we may assume that for each hypersurface H_c the intersection $H_c \cap X_b$ has dimension strictly smaller than $\dim(X_b)$. Thus, by induction there is some constant C' and a definable family of definable sets

$$\{Y_{b,c} \subseteq (H_c \cap X_b)^{\mathrm{alg}}\}_{b,c}$$

for which $N((H_c \cap X_b) \smallsetminus Y_{b,c}, t) \leq C' t^{\frac{\epsilon}{2}}$. Define $\widetilde{Y_b} := \bigcup Y_{b,c}$, then $\widetilde{Y_b} \subseteq X_b^{\mathrm{alg}}$, is definable, and $N(X_b \smallsetminus Y_b, t) \leq (Ct^{\frac{\epsilon}{2}})(C't^{\frac{\epsilon}{2}}) = (CC')t^\epsilon$, as required. □

In our applications of Theorem 4.6, refinements in two different directions will be required. First, we will need to count algebraic points, not merely rational points. Secondly, sometimes we need to bound the rational points in a definable set even when it has a large algebraic part. In fact, these two refinements are not unrelated as the most natural way to study algebraic points is to use a restriction of scalars construction which tends to produce sets having no transcendental part.

Definition 4.34. — Let $H : \mathbb{Q}^{\mathrm{alg}} \to \mathbb{R}$ be the multiplicative height function on the algebraic numbers. We extend H to n-tuples by $H(x_1, \ldots, x_n) := \max\{H(x_i) : i \leq n\}$. For a positive integer k and set $X \subseteq \mathbb{R}^n$, we define

$$X(k,t) \;:=\; \{(x_1, \ldots, x_n) \in X \;:\; [\mathbb{Q}(x_i) : \mathbb{Q}] \leq k \;\&\; H(x_i) \leq t \text{ for each } i \leq n\}$$

and $N(X,k,t) := \#X(k,t)$.

Remark 4.35. — In line with our earlier notation, $X(1,t) = X(\mathbb{Q}, t)$.

The direct generalization of Theorem 4.6 to bounding $N(X,k,t)$ holds. That is, Pila has shown [**52**] that for any o-minimally definable set $X \subseteq \mathbb{R}^n$, positive number $\epsilon > 0$, and positive integer k, there is a constant $C = C(X,k,\epsilon)$ for which the inequality $N(X^{\mathrm{tr}}, k, t) \leq Ct^{\epsilon}$ holds. The proof of this version of the counting theorem passes through an analysis of algebraic blocks.

Definition 4.36. — A *block* $X \subseteq \mathbb{R}^n$ in some o-minimal expansion of the real field is an infinite, connected, definable set for which there is a connected semialgebraic set $Y \subseteq \mathbb{R}^n$ for which X is contained in the nonsingular locus of Y and X agrees with Y in some neighborhood of each point of X.

As we have seen in the proof of Theorem 4.6, the rational points on definable sets may be confined to a small number of algebraic hypersurfaces. In the course of proving the extension of the counting theorem to algebraic points, one shows that the algebraic points of fixed degree may be confined to few blocks by mixing the earlier result on rational points with considerations of how algebraic points may be understood in terms of rational points on some associated definable set.

There are examples of sets X definable in $\mathbb{R}_{\mathrm{an}}$ for which $N(X^{\mathrm{tr}}, t)$ grows faster than every polynomial in $\log(t)$. Indeed, it is not hard to achieve this with restricted analytic functions $f : [0, 1] \to \mathbb{R}$ by constructing highly lacunary power series so that f takes rational values at many rational arguments. On the other hand, no such sets definable in $\mathbb{R}_{\exp}$ are known to exist, and Wilkie conjectures that, in fact, for any set $X \subseteq \mathbb{R}^n$ definable in $\mathbb{R}_{\exp}$ (or more generally, in $\mathbb{R}_{\mathrm{Pfaff}}$, the expansion of $\mathbb{R}$ by all Pfaffian functions) there are constants C and K so that for $t \gg 1$ one has $N(X^{\mathrm{tr}}, t) \leq C(\log(t))^K$. This conjecture has been proven for curves in the plane by Pila [**51**], under some additional hypotheses for surfaces by Jones and Thomas [**27**], and for algebraic points in surfaces definable in $\mathbb{R}_{\exp}$ by Butler [**9**], but remains open in higher dimension. The definable sets which arise in the diophantine geometric applications of the counting theorem are for the most part not definable in $\mathbb{R}_{\mathrm{Pfaff}}$, but they are defined using functions which do satisfy differential equations. As the known restricted analytic functions giving many rational points are differentially transcendental, one might speculate that a generalization of Wilkie's conjecture might hold for sets definable in o-minimal expansions of the real field by functions satisfying a more general class of differential equations than merely Pfaffian equations, but I am loth to formulate this speculation as a conjecture. If one knew such bounds, then one would need only prove weaker Galois theoretic lower bounds in order to apply the counting theorem to diophantine problems.

5. Applications of the counting theorem to diophantine geometry

In this section we shall explain several theorems in diophantine geometry proven using the Pila-Wilkie counting theorem. We shall begin with the Pila-Zannier reproof of the Manin-Mumford conjecture followed by Pila's unconditional proof of the André-Oort conjecture (and some generalizations) for products of modular curves. The generalizations Pila considers, namely for which the ambient varieties are products of modular curves and certain split semiabelian varieties, are in the direction of Pink's conjectures on mixed Shimura varieties. With the theorem of Masser and Zannier on simultaneous torsion in the Legendre family of elliptic curves, we see an instance of the Pink-Zilber conjectures on anomalous intersections where unlike the cases of the

Manin-Mumford and André-Oort conjectures rather than considering merely special points, that is special subvarieties of dimension zero, we study intersections with collections of higher dimensional special subvarieties. The theorem of Habegger and Pila on some unlikely intersections in powers of the j-line is an explicit attempt at the Pink conjecture in the case of Shimura varieties. The theorems we have chosen to expose do not exhaust the known results in diophantine geometry which have been proven using the Pila-Zannier strategy but they do indicate the scope of the method.

5.1. Manin-Mumford conjecture. — The Manin-Mumford conjecture, first proven by Raynaud using p-adic methods **[62]**, asserts that if a closed subvariety of an abelian variety over $\mathbb{C}$ contains a Zariski dense set of torsion points, then that subvariety must be a translate by a torsion point of an abelian subvariety (where for this statement we consider the trivial subgroup as an abelian subvariety). Pila and Zannier offered a new proof of Raynaud's theorem under the additional hypothesis that the abelian variety in question is defined over a number field. Presumably, either a specialization argument or a refinement of Masser's theorems on the Galois action on torsion points to a transcendental base could be used to extend their argument to the general case. However, in keeping with my promise not to disfigure the theorems I am exposing by introducing my own "improvements", I will restrict myself to the case of abelian varieties over number fields.

The main theorem of **[58]** reads as follows.

Theorem 5.1. — *Let k be a number field, A be an abelian variety over k, and $X \subseteq A$ be a closed, irreducible subvariety. We suppose that the set*

$$\{\xi \in X(\mathbb{C}) \ : \ (\exists n \in \mathbb{Z}_+)[n]_A(\xi) = 0\}$$

of torsion points of $A(\mathbb{C})$ lying on X is Zariski dense in X. Then X is a translate by a torsion point of an abelian subvariety of A.

To prove Theorem 5.1 we shall employ some standard reductions; some of which will appear in other arguments, but some of which lack natural analogues outside the context of abelian varieties.

First, we have already assumed A to be defined over a number field. It follows from Lagrange interpolation that since all of the torsion points are algebraic, the variety X must be defined over a number field as well. Thus, we may assume that both A and X are defined over k. Secondly, we may assume that the stabilizer of X in A is trivial. Indeed, if $S = \mathrm{Stab}_A(X)$ is the stabilizer of X in A, $\pi : A \to A/S$ is the quotient map, and $\overline{X} = \pi(X)$ is the image of X under π, then A/S is an abelian variety, $\overline{X}(\mathbb{C})$ meets $(A/S)(\mathbb{C})_{\mathrm{tor}}$ in a Zariski dense set, the stabilizer of $\overline{X}$ is trivial, and if $\overline{X}$ were a translate by a torsion point of an abelian subvariety, then $X = \pi^{-1}\overline{X}$ would also be a translate by a torsion point of an abelian subvariety of A. For the last reduction we recall the notion of the special locus, or what is sometimes called the Ueno locus.

Let us define the special locus at the level of its $\mathbb{C}$-valued points

$$\mathrm{SpL}(X)(\mathbb{C}) \quad : \quad \{a \in X(\mathbb{C}) \ : \ a + B \subseteq X \\ \text{for some positive dimensional algebraic subgroup } B \leq A\}$$

While we have defined $\mathrm{SpL}(X)$ as a set of points, it is, in fact, a closed subvariety of X. The proof of this fact follows by a fairly easy induction on $\dim(X)$ provided that one proves a stronger uniform version to the effect that if X is allowed to vary in a constructible family of subvarieties of A, then $\mathrm{SpL}(X)$ also varies in a constructible family of subvarieties and that, moreover, for X irreducible and positive dimensional, $\mathrm{SpL}(X) = X$ if and only if the stabilizer of X is positive dimensional. Given such a constructible family $\{X_b\}_{b \in B}$ of subvarieties of A, via a base change if necessary, one reduces to the case that each X_b is irreducible and then considers a further family $\{(a + X_b) \cap X_b\}_{(a,b) \in A \times B}$. Apply induction and the observation that if $c \in \mathrm{SpL}(X_b)$ and X_b has a finite stabilizer, then there is some $a \in A$ with $c \in \mathrm{SpL}(X_b \cap (a + X_b))$ to complete the argument.

With the introduction of the special locus, we have our third reduction: it suffices to show that there are only finitely many torsion points in $X \smallsetminus \mathrm{SpL}(X)$. We shall now convert this to an analytic problem. Let $\pi : \mathbb{C}^g \to A(\mathbb{C})$ express the compact complex Lie group $A(\mathbb{C})$ as a quotient of $\mathbb{C}^g$ by some lattice $\Lambda \leq \mathbb{C}^g$. As a *real* vector space, we have $\mathbb{C}^g = \mathbb{R} \otimes \Lambda$. Using a basis of Λ as an $\mathbb{R}$-basis of $\mathbb{C}^g$, we regard $\mathbb{C}^g$ as $\mathbb{R}^{2g}$ so that $\pi : \mathbb{R}^{2g} \to A(\mathbb{C})$ is a *real* analytic function. While this choice of real coordinates differs from the usual presentation of $\mathbb{C}$ in terms of real and imaginary parts, since we can move from one to the other by a real linear change of variables, the question of whether some subset of $\mathbb{C}^g$ is semialgebraic does not depend upon our choice of coordinates. Of course, the map $\pi : \mathbb{R}^{2g} \to A(\mathbb{C})$ is not o-minimally definable, but if we restrict it to the fundamental domain $\mathfrak{D} := [0,1)^{2g}$, which is semialgebraic, then because $\pi \restriction [0,1]^{2g}$ is the restriction of a globally analytic function to a compact box, the resulting function $\widetilde{\pi} := \pi \restriction [0,1)^{2g}$ is definable in $\mathbb{R}_{\mathrm{an}}$.

With respect to our choice of real coordinates, the torsion points in $A(\mathbb{C})$ are in bijection with $\mathbb{Q}^{2g} \cap \mathfrak{D}$ via $\widetilde{\pi}$. Let $\mathfrak{X} := \widetilde{\pi}^{-1}(X(\mathbb{C}))$ and let $\mathfrak{S} := \widetilde{\pi}^{-1}(\mathrm{SpL}(X))$. Our charge is to show that $(\mathfrak{X} \smallsetminus \mathfrak{S}) \cap \mathbb{Q}^{2g}$ is finite. To apply Theorem 4.6 to $\mathfrak{X}$ we must first determine its algebraic part.

Proposition 5.2. — $\mathfrak{X}^{alg} = \mathfrak{S}$

More than one proof of Proposition 5.2 is available to us: one based on differential algebra, as explicated by Marker **[37]**, and a second proof using o-minimality, including a second application of Theorem 4.6. One expects the differential algebraic arguments to generalize, but to date, it is only the o-minimal approach which has been applied successfully to compute the algebraic parts of the definable sets arising in the other applications of this method. For all such arguments, one begins by arguing that the algebraic part may be computed by considering the *complex* algebraic part rather than the *real* algebraic part. More precisely, we have the following lemma.

Lemma 5.3. — *Let $Y \subseteq \mathbb{C}^n$ be an irreducible complex analytic subset of $\mathbb{C}^n$ for some $n \in \mathbb{Z}_+$. We write $\Delta := \{z \in \mathbb{C} \ : \ |z| < 1\}$ for the open unit disc. Define*

$$Y^{ca} := \bigcup_{\substack{\gamma : \Delta \to Y(\mathbb{C}) \\ \text{nonconstant complex analytic} \\ \text{and complex algebraic}}} \gamma(\Delta)$$

$$Y^{ra} := \bigcup_{\substack{\gamma : (-1,1) \to Y(\mathbb{C}) \\ \text{nonconstant real analytic} \\ \text{and semialgebraic}}} \gamma(-1,1)$$

Then $Y^{ca} = Y^{ra}$

Remark 5.4. — In saying that $\gamma : (-1,1) \to Y(\mathbb{C})$ is real analytic and semialgebraic, we are regarding $\mathbb{C}^n$ with its real structure, thus, as $\mathbb{R}^{2n}$, whereas when we say that $\gamma : \Delta \to Y(\mathbb{C})$ is complex analytic and complex algebraic, we treat $\mathbb{C}^n$ as n-dimensional.

Proof. — The containment $Y^{ca} \subseteq Y^{ra}$ follows easily from the observation that if $\gamma : \Delta \to Y(\mathbb{C})$ is nonconstant complex analytic and algebraic, then its image is is equal to the union of the images of the restriction of γ to each line through the origin and each such function is a nonconstant real analytic, semialgebraic function.

For the other inclusion, consider some nonconstant real analytic, semialgebraic function $\gamma : (-1,1) \to Y(\mathbb{C})$. Express this function in coordinates as

$$t \mapsto (\gamma_{1,R}(t) + i\gamma_{1,I}(t), \ldots, \gamma_{n,R}(t) + i\gamma_{n,I}(t))$$

Set

$$z_j(t) := \gamma_{j,R}(t) + i\gamma_{j,I}(t)$$

Note that I am *not* asserting that z_j is complex analytic. Since $\operatorname{tr.deg}_{\mathbb{R}}(\mathbb{R}(\gamma(t))) = 1$ and $\mathbb{C}(z_1(t), \ldots, z_n(t)) \subseteq \mathbb{R}(\gamma(t))^{\mathrm{alg}}$, we conclude that $\operatorname{tr.deg}_{\mathbb{C}}(\mathbb{C}(z_1(t), \ldots, z_n(t)) = 1$. Hence, there is an irreducible affine algebraic curve $X \subseteq \mathbb{A}^n_{\mathbb{C}}$ for which $X(\mathbb{C}) \cap Y(\mathbb{C})$ is infinite. As Y is an analytic variety, $X \subseteq Y$. To complete the argument, use the fact that a complex algebraic curve may be covered by complex analytic charts. $\square$

We shall now employ an abelian analogue of a transcendence theorem of Ax **[2]** to complete the proof of Proposition 5.2. Schanuel has conjectured that if $\alpha_1, \ldots, \alpha_n$ are $\mathbb{Q}$-linearly independent complex numbers, then $\operatorname{tr.deg}_{\mathbb{Q}}(\mathbb{Q}(\alpha_1, \ldots, \alpha_n, e^{\alpha_1}, \ldots, e^{\alpha_n}) \geq n$. While some notable special cases of Schanuel's conjecture are known, for example the Lindemann-Weierstraß theorem **[35, 84]** for which the numbers $\alpha_1, \ldots, \alpha_n$ are all assumed to be algebraic and Baker's theorem on linear forms in logarithms in which the numbers $e^{\alpha_1}, \ldots, e^{\alpha_n}$ are all algebraic **[3]**, as a whole, the conjecture remains wide open. However, Ax has proven a functional version of the conjecture best stated in terms of differential fields.

Theorem 5.5 **(Ax)**. — *Let K be a field of characteristic zero with a derivation $\partial : K \to K$ having field of constants $C := \{a \in K : \partial(a) = 0\}$. Suppose that $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in K$ be elements of K satisfying the differential equations*

$$\partial \alpha_i = \frac{\partial \beta_i}{\beta_i}$$

Suppose moreover that $1, \alpha_1, \ldots, \alpha_n$ are C-linearly independent. Then

$$\operatorname{tr.deg}_C(C(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n)) \geq n + 1$$

Specializing to the case of $C = \mathbb{C}$, $K = \mathbb{C}((t))$, $\partial = \frac{\partial}{\partial t}$, and $\alpha_1(t), \ldots, \alpha_n(t) \in t\mathbb{C}[[t]]$ which are $\mathbb{Q}$-linearly independent, and $\beta_j(t) := \exp(\alpha_j(t)) = \sum_{n=0}^{\infty} \frac{1}{n!} \alpha_j(t)^n$, one recovers a formal version of the Schanuel conjecture. For the problem under consideration, one requires a generalization of Ax's theorem to the Weierstraß $\wp$-functions, due to Brownawell and Kubota **[8]**.

Remark 5.6. — Further generalizations of Ax's theorem to semiabelian varieties which are not necessarily defined over the constants have been proven by Bertrand and Pillay **[5]**. The elliptic case has been studied by several other authors. See, for example, **[12, 30, 29]**.

Let us now complete the characterization of $\mathfrak{X}^{\text{alg}}$. First, we observe that $\mathfrak{S} \subseteq \mathfrak{X}^{alg}$ as the preimage of a translate of an abelian variety contained in X is the intersection with $\mathfrak{D}$ of an affine space entirely contained in $\pi^{-1}X(\mathbb{C})$.

For the other inclusion, we observe that if $a \in \mathfrak{X}^{\text{alg}}$, then $\pi(a) \in \operatorname{SpL}(X)$. Indeed, by Lemma 5.3 we may find $\gamma : \Delta \to \mathfrak{X}$ a nonconstant complex analytic and algebraic map with $\gamma(0) = a$. Let $\beta := \pi \circ \gamma : \Delta \to X(\mathbb{C})$. Taking K to be the field of meromorphic functions on Δ, we may regard γ as a K-valued point of $T_e A$ and β as a K-valued point of X. Differentiating the equation $\beta = \pi(\gamma)$, we see that $\partial \log_A(\beta) = \partial(\gamma)$. By algebraicity of γ we have $\operatorname{tr.deg}_{\mathbb{C}}(\mathbb{C}(\gamma)) = 1$ while $\operatorname{tr.deg}_{\mathbb{C}}(\mathbb{C}(\beta)) \leq \dim(X) < \dim(A)$. Hence, the elliptic transcendence theorem of Brownawell and Kubota implies that there is some proper algebraic group $B < A$ with $\beta \in \pi(a) + B(K)$. As β is nonconstant, $\dim(B) > 0$. Minimizing $\dim(B)$, translating and applying the transcendence theorem again with A replaced by B, we see that β is generic in $\pi(a) + B$. Since $\beta \in X(K)$, we conclude that $\pi(a) + B \subseteq X$ so that $\pi(a) \in \operatorname{SpL}(X)$ as claimed.

To finish the proof of Theorem 5.1 we shall invoke a theorem of Masser **[40]** on the Galois action on torsion points.

Theorem 5.7 **(Masser)**. — *Let A be an abelian variety over a number field k. Then there are constants $C = C(A, k)$ and $\eta = \eta(A, k)$ so that for any torsion point $\xi \in A(k^{alg})$ of exact order n one has*

$$[k(\xi) : k] \geq C n^{\eta}$$

Remark 5.8. — Masser's theorem gives much more precise information on the constants C and η and holds even for families of abelian varieties over some fixed number field. The coarse bounds suffice for our applications.

Let us make one final observation before finishing the proof. If $x \in \mathbb{Q}^{2g} \cap \mathfrak{D}$, then $H(x)$ and the order of x in the group $(\mathbb{R}/\mathbb{Z})^{2g}$, or, what is the same thing, the order of $\pi(x)$ in $A(\mathbb{C})$, which we shall write as $o(\pi(x))$, are related by

$$H(x) \leq o(\pi(x)) \leq H(x)^{2g}$$

Indeed, if we write

$$x = (\frac{p_1}{q_1}, \ldots, \frac{p_{2g}}{q_{2g}})$$

where $0 \leq p_i \leq q_i$ are integers with $\frac{p_i}{q_i}$ in lowest terms, then (as long as $x \neq 0$)

$$H(x) = \max\{q_i : 1 \leq i \leq 2g\}$$

while

$$o(\pi(x)) = \operatorname{lcm}(\{q_i : 1 \leq i \leq 2g\}$$

We finish the proof now. Let C and η be the constants from Theorem 5.7. Let C' be the constant from Theorem 4.6 with $\epsilon = \frac{\eta}{2}$ so that $N(\mathfrak{X}^{\mathrm{tr}}, t) \leq C' t^{\epsilon}$. If $X(\mathbb{C}) \smallsetminus \mathrm{SpL}(X)$ contained infinitely many torsion points, then one could find such a torsion point ξ of exact order $n > (C'/C)^{\frac{2}{\eta}}$. As X and $\mathrm{SpL}(X)$ are defined over k, if $\sigma \in \mathrm{Gal}(k^{\mathrm{alg}}/k)$, then $\sigma(\xi) \in X(\mathbb{C}) \smallsetminus \mathrm{SpL}(X)$ as well. Taking into account the inequality $H(x) \leq o(\pi(x))$ for $x \in \mathfrak{D}$ and the fact that π induces a bijection between $\mathbb{Q}^{2g} \cap \mathfrak{X}^{\mathrm{tr}}$ and the set of torsion points in $X(\mathbb{C}) \smallsetminus \mathrm{SpL}(X)$, we see from Theorem 5.7 that $N(\mathfrak{X}^{\mathrm{tr}}, n) \geq C n^{\eta}$ while Theorem 4.6 gives $N(\mathfrak{X}^{\mathrm{tr}}, n) \leq C' n^{\frac{\eta}{2}}$ From our choice of n, these two inequalities are inconsistent.

5.2. Pila's André-Oort theorem. — Overall, Pila's proof of the André-Oort conjecture for products of curves follows the strategy of the Pila-Zannier reproof of the Manin-Mumford conjecture with some notable changes. First, to obtain the definability of the relevant analytic covering maps, we must work in $\mathbb{R}_{\mathrm{an,exp}}$; restricted analytic functions do not suffice. Secondly, the lower bounds on the sizes of Galois orbits are obtained from Siegel's theorem on the class number of quadratic imaginary fields. The requisite bounds for general Shimura varieties are not yet known unconditionally, though Tsimerman has produced such bounds for the coarse moduli spaces of principally polarized abelian varieties of any dimension [**74**]. Finally, a differential algebraic transcendence theorem analogous to the Ax-Lindemann-Weierstraß theorem is not known for the j-function. Consequently, the characterization of the algebraic part of definable sets arising as pullbacks of algebraic varieties requires a different argument based on o-minimality. It bears noting that the proof of Proposition 5.2 in [**58**] employs similar considerations rather than the differential algebraic methods we used.

Pila's theorem is often cited as an unconditional version of the André-Oort conjecture in the case where the ambient Shimura variety is $\mathbb{A}^n_{\mathbb{C}}$ regarded as the n^{th} Cartesian power of the j-line. However, he proves a stronger theorem, namely that some cases of Pink's generalized André-Oort + Manin-Mumford conjecture [**61**] hold for varieties expressible as products of modular curves, elliptic curves defined over a number field, and powers of the multiplicative group. Let us give a precise formulation before we begin its proof. We begin by defining the special points.

Definition 5.9. — We regard $\mathbb{A}^1$ as the (coarse) moduli space of elliptic curves. If $Y = \mathbb{A}^n \times A$ is a product of some Cartesian power of affine space with a commutative algebraic group A, then by a special point on Y we mean a $\mathbb{C}$-valued point of the form

$$(\xi_1, \dots, \xi_n, \zeta) \in (\mathbb{A}^1 \times \cdots \times \mathbb{A}^1 \times A)(\mathbb{C})$$

where each ξ_i is a moduli point of an elliptic curve with complex multiplication and $\zeta \in A(\mathbb{C})$ is a torsion point.

Let us say now what we would mean by a special subvariety of a product of affine lines and algebraic groups.

Definition 5.10. — By a *weakly special subvariety* of $\mathbb{A}^n$ we mean a component of a variety defined by systems of equations of the form $\Phi_N(x_i, x_j) = 0$ and $x_k = \xi$ where Φ_N is the N^{th} modular polynomial (the case of $N = 1$ is allowed) and ξ is some fixed point. If we require that ξ be a special point, then the variety is a *special subvariety*. A subvariety of a product of $\mathbb{A}^n$ with a commutative algebraic a group is (weakly) special if it is a product of a (weakly) special subvariety of $\mathbb{A}^n$ with a translate by a torsion point (by a general point) of an algebraic subgroup of A.

With these definitions in place we may state Pila's theorem.

***Theorem 5.11* (Pila)**. — *Let $N \in \mathbb{N}$ be a natural number and A an algebraic group over a number field which is expressible as a product of elliptic curves and a Cartesian power of the multiplicative group. If $Y \subseteq X := \mathbb{A}^N \times A$ is an irreducible subvariety for which the set*

$$\{a \in Y(\mathbb{C}) \ : \ a \in (\mathbb{A}^N \times A)(\mathbb{C}) \textit{ is a special point } \}$$

is Zariski dense in Y, then Y is a special subvariety.

As with the proof of the Manin-Mumford conjecture, we shall prove Theorem 5.11 by fixing an o-minimally definable analytic covering of $\pi : \mathfrak{X} \to \mathbb{A}^N(\mathbb{C}) \times A(\mathbb{C})$ for which the special points come from arithmetically simple points in $\mathfrak{D}$ and then consider the set of such points on $\mathfrak{X} := \pi^{-1}(X(\mathbb{C}))$.

Of course, the j-function is not o-minimally definable as a function on all of the upper half-space

$$\mathfrak{h} := \{z \in \mathbb{C} \ : \ \mathrm{Im}(z) > 0\}$$

but if one restricts j to a fundamental domain, say, the standard fundamental domain

$$\mathfrak{D} := \{z \in \mathbb{C} \ : \ \frac{-1}{2} \leq \mathrm{Re}(z) < \frac{1}{2} \text{ and } |z| \geq 1\} \ ,$$

then the resulting function is definable in $\mathbb{R}_{\mathrm{an,exp}}$ (relative to the usual presentation of $\mathbb{C}$ via real and imaginary parts). Indeed, it is well known that $j(z) = J(\exp(2\pi i z))$ where J is a meromorphic function on the open unit disc with a simple pole at the origin whose first few terms are given by

$$J(q) = \frac{1}{q} + 744 + 196884q + 21493760q^2 + 864299970q^3 + \cdots$$

The map $z \mapsto \exp(2\pi i z)$ restricted to $\mathfrak{D}$ is o-minimally definable as

$$z \mapsto \exp(-2\pi \operatorname{Im}(z))(\cos(2\pi \operatorname{Re}(z)) + i \sin(2\pi \operatorname{Re}(z))$$

While the trigonometric functions over the full real line are not definable in $\mathbb{R}_{\text{an,exp}}$, since the real part of z is bounded between $\frac{-1}{2}$ and $\frac{1}{2}$ on $\mathfrak{D}$, only the restrictions of sine and cosine to a compact interval are required, and these are clearly restricted analytic. The image of $\mathfrak{D}$ under $z \mapsto \exp(2\pi i z)$ is contained in $B_{\exp(-\sqrt{3}\pi)}(0)$, the disc of radius $\exp(-\sqrt{3}\pi)$ about the origin, thus, we may express $j \restriction \mathfrak{D}$ as

$$j(z) = (J \restriction B_{\exp(-\sqrt{3}\pi)}(0))(\exp(2\pi i z))$$

The restriction of J to $B_{\exp(-\sqrt{3}\pi}(0)$ is the quotient of a restricted analytic function by a polynomial, and is thus definable in $\mathbb{R}_{\text{an,exp}}$, showing that the restriction of j to $\mathfrak{D}$ is definable.

From the analytic theory of elliptic curves, we know that the elliptic curve

$$E_{j(\tau)}(\mathbb{C}) = \mathbb{C}/(\mathbb{Z} + \mathbb{Z}\tau)$$

has complex multiplication if and only if $[\mathbb{Q}(\tau) : \mathbb{Q}] = 2$. Thus the *prespecial* points in $\mathfrak{D}$, by which we mean the points in $\mathfrak{D}$ which map to special points by our given analytic covering map, namely the j-function, are the quadratic imaginary numbers in $\mathfrak{D}$.

As we have already discussed when considering the Manin-Mumford conjecture, if B is a connected commutative algebraic group over $\mathbb{C}$, then its universal covering space is $\mathbb{C}^g$, where $g = \dim(G)$. More specifically, if B is an abelian variety, then by restricting the covering map to a fundamental domain and taking a basis of the kernel as a basis for $\mathbb{C}^g$ as an $\mathbb{R}$-vector space, the we may think of this covering map as $\pi : [0,1)^{2g} \to B(\mathbb{C})$ where π is the restriction to a semialgebraic set of a restricted analytic function and is therefore definable in $\mathbb{R}_{\text{an}}$. In this case, the prespecial points, namely the points mapping to torsion points, are the rational points in the domain of π. When $B = \mathbb{G}_m$ is the multiplicative group, then our covering map $\pi : \mathbb{C} \to \mathbb{G}_m(\mathbb{C})$ is given by $z \mapsto \exp(2\pi i z)$. As noted above, by restricting this map to the fundamental domain

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

the complex exponential map becomes definable in $\mathbb{R}_{\text{an,exp}}$.

Using these various analytic covering maps as the coordinates of a single covering map we obtain an analytic covering map $\widetilde{\pi} : \widetilde{\mathfrak{X}} \to X(\mathbb{C})$ where $\widetilde{\mathfrak{X}} = \mathfrak{h}^N \times \mathbb{C}^g$ with $g = \dim A$ and then by restricting to a fundamental domain $\mathfrak{X}$, we obtain an o-minimally definable covering map $\pi : \mathfrak{X} \to X(\mathbb{C})$ for which the prespecial points in the coordinates corresponding to $\mathbb{A}^1$ are the quadratic imaginary numbers and in the other coordinates are (some of) the rational points relative to the choice of real coordinates noted above. We have thus converted the problem of describing the special points in $Y(\mathbb{C})$ to that of describing the prespecial points in $\mathfrak{Y} := \pi^{-1}Y(\mathbb{C})$.

As with the Manin-Mumford conjecture, we define a special locus for Y and then reduce to showing that there are only finitely many special points in Y not contained in its special locus. However, unlike the case of abelian varieties, our proof that the special locus is constructible uses the analytic covering in an essential way. The key observation is that the (weakly) special subvarieties come from *geodesic* analytic subvarieties of $\mathfrak{X}$.

The map $\widetilde{\pi} : \widetilde{\mathfrak{X}} \to X(\mathbb{C})$ expresses $X(\mathbb{C})$ as a quotient of $\mathfrak{X}$ by an arithmetic group. The domain $\widetilde{\mathfrak{X}}$ is a homogenous space for the group $G(\mathbb{R}) := \mathrm{PSL}_2(\mathbb{R})^N \times \mathbb{G}_a^{2g}(\mathbb{R})$ and, again relative to our choice of real coordinates, we may regard $X(\mathbb{C})$ as $\Gamma\backslash\widetilde{\mathfrak{X}}$ where $\Gamma = H(\mathbb{Z})$ for a certain algebraic subgroup of G. Here the group H is a product of the algebraic groups PSL_2 for each $\mathbb{A}^1$ factor of X, $\mathbb{G}_a^2$ for each elliptic factor, and $\mathbb{G}_a \times \{0\}$ for each multiplicative group factor.

If $Y \subseteq X$ were a special variety and $\widetilde{\mathfrak{Y}}$ were a component of $\widetilde{\pi}^{-1}Y(\mathbb{C})$, then $\widetilde{\mathfrak{Y}}$ would be a product of an analytic subvariety of $\mathfrak{h}^N$ defined by equations of the form $z_i = \gamma z_j$ or $z_k = \xi$ for some $\gamma \in \mathrm{PSL}_2(\mathbb{Q})$ or prespecial point ξ and an affine subspace of $\mathbb{C}^g$ (which relative to the real coordinates comes from one defined over $\mathbb{Q}$).

Definition 5.12. — In general, we say that an analytic subvariety of $\widetilde{\mathfrak{X}}$ is weakly geodesic if it is of the form $U \times V$ where $U \subseteq \mathfrak{h}^N$ is defined by equations of the form $x_i = \gamma x_j$ or $x_k = \xi$ for an arbitrary $\gamma \in \mathrm{PSL}_2(\mathbb{R})$ or point $\xi \in \mathfrak{h}$ and $V \subseteq \mathbb{C}^g$ is a translate of some subspace. If the elements γ come from $\mathrm{PSL}_2(\mathbb{Q})$ and $\pi(V)$ is a translate of an algebraic subgroup of A, then we say that $U \times V$ is *geodesic*.

Let us note that if $\mathfrak{Y}$ is a subset of a proper geodesic subvariety of $\mathfrak{h}^N \times \mathbb{C}^g$, then we may pass to an appropriate quotient without changing the hypothesis that $Y(\mathbb{C})$ contains a Zariski dense set of special points but is not a special variety. Thus, it suffices to consider the case that no proper geodesic variety contains $\mathfrak{Y}$. We define the special locus of Y, $\mathrm{SpL}(Y)$ to be the union of all positive dimensional irreducible weakly special subvarieties of Y. On the face of it, this is a countable union of algebraic varieties. However, after reducing to the case that $\mathfrak{Y}$ is not contained in a proper geodesic subvariety of $\mathfrak{h}^N \times \mathbb{C}^g$ and Y is not special, we show that $\mathfrak{Y}^{\mathrm{alg}} = \pi^{-1}\,\mathrm{SpL}(Y)$ and that $\mathrm{SpL}(Y)$ is a proper subvariety of Y. This fact may be regarded as an analogue of the Ax-Lindemann-Weierstraß theorem, but its proof has resisted differential algebraic treatments and occupies three sections of [**53**].

One obtains this characterization of the algebraic part of $\mathfrak{Y}$ in stages.

First, let $\widetilde{\mathfrak{Y}}$ be a component of $\pi^{-1}Y(\mathbb{C})$ which meets $\mathfrak{X}$ in a set of full dimension. We argue that if $Z \subseteq \widetilde{\mathfrak{Y}}$ is a (positive dimensional) maximal algebraic subvariety of $\widetilde{\mathfrak{Y}}$, then Z is contained in a geodesic subvariety $\widetilde{\mathfrak{X}}$ which is contained in $\widetilde{\mathfrak{Y}}$. Secondly, invoking Lemma 5.3 we conclude that $\mathfrak{Y}^{\mathrm{alg}}$ is covered by geodesic subvarieties. Finally, using the compactness theorem and the fact that there are only countably many shapes of geodesic varieties, we show that $\pi(\mathfrak{Y}^{\mathrm{alg}}) = \mathrm{SpL}(Y)$ is a proper closed subvariety of Y.

Let us explain how the first step works. Translating by an element of $H(\mathbb{Z})$ if need be, we may assume that $Z(\mathbb{C}) \cap \mathfrak{X}$ is Zariski dense in Z. We then consider two definable sets.

$$B_Z := \{\gamma \in H(\mathbb{R}) \ : \ \dim(\gamma Z \cap \mathfrak{X}) = \dim Z\}$$

and

$$C_Z := \{\gamma \in H(\mathbb{R}) \ : \ \dim(\gamma Z \cap \mathfrak{X} \cap \widetilde{\mathfrak{Y}}) = \dim(Z)\}$$

The set C_Z is definable in $\mathbb{R}_{\mathrm{an,exp}}$ while B_Z is semialgebraic, but they share many integral points in common. Indeed, if $\Gamma := \{\gamma \in H(\mathbb{Z}) \ : \ \gamma\widetilde{\mathfrak{Y}} = \widetilde{\mathfrak{Y}}\}$, then $\Gamma \cap B_Z = \Gamma \cap C_Z$. Using the fact that $\widetilde{\mathfrak{Y}}$ is a component of a positive dimensional $H(\mathbb{Z})$-periodic set and some easy estimates on the number of translates of $\mathfrak{X}$ which meet Z in a dense set, one shows that number of points in $B_Z \cap \Gamma$ of height up to t outstrips the Pila-Wilkie bounds. Thus, C_Z must contain an infinite, connected semialgebraic set S. It must be the case that S stabilizes Z for otherwise, for any $s_0 \in S$ $Z \subsetneq s_0^{-1} \cdot S \cdot Z \subseteq \widetilde{\mathfrak{Y}}$ violating the maximality of Z as we have observed that the semialgebraic part of a complex analytic variety is equal to its complex algebraic part. Thus, Z is a homogeneous space for a positive dimensional semialgebraic group. Using a description of the algebraic subgroups of G, one shows that some weakly geodesic relation holds on Z. Through further considerations of the structure of H, it follows that, in fact, a geodesic relation holds on Z. The analysis continues in much the same way as in the proof of the Manin-Mumford conjecture. It is shown that either a geodesic relation holds on all of $\widetilde{\mathfrak{Y}}$ or that the algebraic part of $\widetilde{\mathfrak{Y}}$ is contained in an analytic set of strictly smaller dimension. In the case that a geodesic relation holds on $\widetilde{\mathfrak{Y}}$, one passes to a quotient while in the latter case, we have reached the desired conclusion.

Remark 5.13. — This basic strategy for determining the algebraic part of the pullback of a subvariety $Y(\mathbb{C})$ under an analytic covering map $\pi : \mathfrak{X} \to X(\mathbb{C})$ expressing a variety $X(\mathbb{C})$ as a quotient $\Gamma\backslash\mathfrak{X}$ of a homogeneous space $\mathfrak{X}$ works more generally. By comparing the word metric in Γ to the Euclidean metric in $\mathfrak{X}$, Ullmo and Yafaev [**78**] have shown that if $X(\mathbb{C}) = \Gamma\backslash\mathfrak{X}$ is a Shimura variety for which Γ is co-compact, then for any algebraic variety $Y \subseteq X$, the maximal algebraic subvarieties of $\pi^{-1}(Y(\mathbb{C}))$ are geodesic. Pila and Tsimerman have proven the Ax-Lindemann-Weierstraß theorem for $\mathcal{A}_g$ [**56**]. A general treatment of this problem is given by Klingler, Ullmo and Yafaev [**31**].

Remark 5.14. — In later work, Pila and Habegger [**25**] showed that if one were to define the special locus of a subvariety $Y \subseteq \mathbb{A}^n$ to be the union of all of the positive dimensional weakly special subvarieties of Y, then as in the case of abelian varieties, the special locus is actually a closed subvariety.

One concludes the proof of Theorem 5.11 almost exactly as with the proof of the Manin-Mumford conjecture, with the minor change that the lower bound on the Galois orbits of special points is obtained from Siegel's theorem on the class number of quadratic imaginary fields rather than Masser's theorem on torsion points on abelian varieties.

Lemma 5.15. — *There is a constant C so that for any prespecial $\xi \in \mathfrak{X}$ one has $[k(\pi(\xi)) : k] \geq CH(\xi)^{\frac{1}{2}}$.*

Proof. — For an elliptic curve E_τ with complex multiplication, $[\mathbb{Q}(\tau) : \mathbb{Q}]$ is equal to the class number of $\mathbb{Z}[\tau]$ which is bounded below by a constant multiple of $\sqrt{H(\tau)}$ by Siegel's class number formula [**68**]. For an elliptic curve E defined over a number field k, we know from Theorem 5.7 that the degree of the field generated by a torsion point grows faster than some constant multiple of a power of its order. Finally, from the theory of cyclotomic field extensions, we know that if ζ is a root of unity of order N, then $[\mathbb{Q}(\zeta) : \mathbb{Q}] = \varphi(N)$ where φ is Euler's totient function. It is elementary to see that for any positive $\epsilon < 1$, one has $\varphi(N) > \frac{1}{2}N^\epsilon$. In particular, we could take $\epsilon = \frac{1}{2}$. □

Arguing exactly as in the case of the Manin-Mumford conjecture, if Theorem 5.11 were false, then we could reduce to considering $Y \subseteq X$ an irreducible subvariety defined over a number field k having infinitely many special points not contained in its special locus. We then observe that if $\xi \in \mathfrak{Y}$ were a prespecial point not living in the preimage of the special locus, then each of the $\mathrm{Gal}(k^{\mathrm{alg}}/k)$ conjugates of $\pi(\xi)$ would also correspond to a distinct prespecial point of $\mathfrak{Y}$ outside of its algebraic part but of bounded height. By choosing the order of the group component, that is the projection of $\pi(\xi)$ to $A(\mathbb{C})$, or the discriminant of the modular component high enough, we would contradict Theorem 4.6.

5.3. Simultaneous torsion. — The diophantine theorems we have discussed up to this point have taken the form that a subvariety of a special variety containing a Zariski dense set of special points must be a special variety. With the theorem of Masser and Zannier on simultaneous torsion in the Legendre family of elliptic curves we encounter an instance of the more general Pink-Zilber conjectures on unlikely intersections in higher dimension. Let me state the Masser-Zannier theorem and then explain how it fits into the Pink-Zilber framework.

Theorem 5.16 **(Masser-Zannier).** — *There are only finitely many complex numbers λ for which the points $[2 : \sqrt{2(2-\lambda)} : 1]$ and $[3 : \sqrt{6(3-\lambda)} : 1]$ are both torsion points on the elliptic curve defined as a subvariety of $\mathbb{P}^2$ by the equation $zy^2 = x(x-z)(x-\lambda z)$.*

Remark 5.17. — We have stated Theorem 5.16 in the striking form (which appears in [**41**]) that the abscissas are restricted to be 2 and 3. The proof works equally well for any two constant abscissas provided that the points are linearly independent on the generic fibre. (We leave it to the reader to verify that the points $[2 : \sqrt{2(2-\lambda)} : 1]$ and $[3 : \sqrt{6(3-\lambda)} : 1]$ are indeed linearly independent generically.) With additional work (as shown in [**38**]) the restriction to constant abscissas may be dropped.

Remark 5.18. — Stoll has given a direct proof of Theorem 5.16, showing the stronger result that there are no exceptional λ [**71**].

Let us reformulate Theorem 5.16 as a result about intersections of curves with special varieties. Let us take our ambient special variety X to be the square of the Legendre family over the λ-line. That is, we present X as a closed subvariety of $(\mathbb{A}^1 \smallsetminus \{0,1\}) \times \mathbb{P}^2 \times \mathbb{P}^2$ defined by the equations $z_1 y_1^2 = x_1(x_1 - z_1)(x_1 - \lambda z_1)$ and $z_2 y_2^2 =$

$x_2(x_2-z_2)(x_2-\lambda z_2)$ where λ ranges over $\mathbb{A}^1 \smallsetminus \{0,1\}$, $[x_1 : y_1 : z_1]$ are the homogenous coordinates on the first $\mathbb{P}^2$ factor and $[x_2 : y_2 : z_2]$ are the homogenous coordinates on the second $\mathbb{P}^2$ factor. In so doing, via the projection onto the first coordinate, we may regard $\pi : X \to (\mathbb{A}^1 \smallsetminus \{0,1\})$ as an abelian scheme over $\mathbb{A}^1 \smallsetminus \{0,1\}$, each fibre being the square of an elliptic curve. The points $[2, \sqrt{2(2-\lambda)} : 1]$ and $[3 : \sqrt{6(3-\lambda)} : 1]$ determine a curve C in X defined by the equations $x_1 - 2z_1 = 0$ and $x_2 - 3z_2 = 0$. The numbers λ for which both $[2, \sqrt{2(2-\lambda)} : 1]$ and $[3 : \sqrt{6(3-\lambda)} : 1]$ are torsion, are exactly the numbers λ for which $([2, \sqrt{2(2-\lambda)} : 1], [3 : \sqrt{6(3-\lambda)} : 1]) \in X_\lambda(\mathbb{C})_{\mathrm{tor}}$. To put it another way, they are the points in $(\mathbb{A}^1 \smallsetminus \{0,1\})(\mathbb{C})$ which lie in the image under π of $(C \cap X[N])(\mathbb{C})$ for some $N \in \mathbb{Z}_+$ where $X[N]$ is the kernel of multiplication by N in the abelian scheme X. As it is clearly the case that C cannot contain infinitely many torsion points on a fixed fibre (as, for instance, there are at most four points in C on any fibre of π), Theorem 5.16 may be re-expressed as the assertion that the set

$$C(\mathbb{C}) \cap \bigcup_{N \in \mathbb{Z}_+} X[N](\mathbb{C})$$

is finite.

The varieties $X[N]$ are special in the sense that they are subgroup schemes of X and because $\dim(C) + \dim(X[N]) = 1 + 1 < 3 = \dim(X)$, one expects $C \cap X[N]$ to be empty. The conjectures of Pink-Zilber type (note that I did not say "*the* Pink-Zilber conjecture" as the experts have yet to settle on a final formulation) assert that if X is a special variety and $Y \subseteq X$ is a subvariety of dimension d not contained in a proper special subvariery of X, then the set

$$Y \cap \bigcup_{\substack{Z \subseteq X \text{ special} \\ \operatorname{codim}_X(Z) < d}} Z$$

is not Zariski dense in Y. There is some ambiguity as to which subvarieties of our X, obtained as the square of the Legendre family over the λ-line, should be considered as special. One natural choice would be to regard any component of a subgroup scheme to be special as well as any component of an algebraic subgroup of a fibre of X over a CM-point. In any case, it is easy to see that Theorem 5.16 is a theorem of Pink-Zilber type.

Let us now sketch a proof of Theorem 5.16. Letting $\widetilde{\mathfrak{X}} := \mathfrak{h} \times \mathbb{C}^2$, we have a natural analytic covering map $\widetilde{\pi} : \widetilde{\mathfrak{X}} \to X(\mathbb{C})$. Here the map π takes the form

$$(\tau, z_1, z_2) \mapsto (j_2(\tau), [\wp(z_1, \tau) : \wp'(z_1, \tau) : 1], [\wp(z_2, \tau) : \wp'(z_2, \tau) : 1])$$

where $j_2 : \mathfrak{h} \to \mathbb{A}^1(\mathbb{C}) \smallsetminus \{0,1\}$ expresses $\mathbb{A}^1(\mathbb{C}) \smallsetminus \{0,1\}$ as $\Gamma_0(2)\backslash\mathfrak{h}$ and $\wp$ is the Weierstraß $\wp$-function. Of course, $\pi(\tau, z_1, z_2)$ is a torsion point on $X_{j_2(\tau)}$ if and only if $z_1 \in \mathbb{Q} + \mathbb{Q}\tau$ and $z_2 \in \mathbb{Q} + \mathbb{Q}\tau$. Identifying $\mathfrak{h} \times \mathbb{R}^4$ with $\mathfrak{h} \times \mathbb{C}^2$ via $(\tau, x_1, x_2, x_3, x_4) \mapsto (\tau, x_1 + \tau x_2, x_3 + \tau x_4)$, we see that a point $(\tau, x_1, x_2, x_3, x_4) \in \mathfrak{h} \times \mathbb{R}^4$ is sent to a torsion point in $X_{j_2(\tau)}$ just in case $x_i \in \mathbb{Q}$ for each $i \leq 4$.

The definability in $\mathbb{R}_{\mathrm{an,exp}}$ of the restriction of j_2 to a fundamental domain follows on similar grounds to the definability of j (restricted to a fundamental domain). For fixed τ, we have already used the definability of (the restriction to a fundamental domain) of the function $z \mapsto \wp(z,\tau)$. That the two variable function is definable in $\mathbb{R}_{\mathrm{an,exp}}$ was established by Peterzil and Starchenko [**44**]. Again, the reasoning is similar to the case of the j-function. It is a classical result that, writing Δ for the open unit disc, there is a meromorphic function expressing the universal elliptic curve over $Y_0(2)$ as a quotient of $\Delta \times \Delta$. The function $E : (\tau, z) \mapsto (\exp(2\pi i\tau), \exp(2\pi iz))$ takes a fundamental domain $\wp$ into a compact subset of $\Delta \times \Delta$ and the function E restricted to this fundamental domain is definable. Hence, the composite is definable. Thus, upon restricting $\widetilde{\pi}$ to some fundamental domain $\mathfrak{X}$ we have that $\pi := \widetilde{\pi} \restriction \mathfrak{X}$ is definable and that relative to an appropriate choice of coordinates, the points mapping to torsion points are those whose last coordinates are rational.

Let $p : \mathfrak{h} \times \mathbb{C}^2 \to \mathbb{C}^2$ be the natural projection map and consider $\mathfrak{C} := p\pi^{-1}C(\mathbb{C})$. The set $\mathfrak{C}$ is definable, being the projection of a definable set. Masser and Zannier show that $\mathfrak{C}^{\mathrm{alg}} = \varnothing$ by invoking known results on the algebraic independence of theta functions. Thus, one is left with the problem of showing that there are only finitely many torsion points on C or equivalently that there are only finitely many rational points on $\mathfrak{C}$.

One obtains the lower bounds on the size of Galois orbits of torsion points through two theorems. First, as a consequence of work of David [**13**] and of Masser [**40**] on the Galois action on the torsion points on abelian varieties we may bound the order of the points $[2 : \sqrt{2(2-\lambda)} : 1]$ and $[3 : \sqrt{6(3-\lambda)} : 1]$ by a function of the height of λ and the degree of λ over $\mathbb{Q}$.

Lemma 5.19. — *There is a constant C so that for any algebraic number λ if $[2 : \sqrt{2(2-\lambda)} : 1]$ or $[3 : \sqrt{6(3-\lambda)} : 1]$ is torsion point of exact order n on the elliptic curve E_λ defined as a subvariety of $\mathbb{P}^2$ by the equation $zy^2 = x(x-z)(x-\lambda z)$, then $n \le c[\mathbb{Q}(\lambda) : \mathbb{Q}]^2(1 + h(\lambda))$.*

The second result is a consequence of Silverman's specialization theorem [**69**].

Lemma 5.20. — *There is a constant C' so that if either $[2 : \sqrt{2(2-\lambda)} : 1]$ or $[3 : \sqrt{6(3-\lambda)} : 1]$ is torsion on E_λ, then $h(\lambda) \le C'$.*

Combining Lemmata 5.19 and 5.20, we see that there is a constant D so that if either $[2 : \sqrt{2(2-\lambda)} : 1]$ or $[3 : \sqrt{6(3-\lambda)} : 1]$ is torsion of exact order n on E_λ, then $[\mathbb{Q}(\lambda) : \mathbb{Q}] \ge Dn^{\frac{1}{2}}$.

We use this conclusion to complete the argument. If there were infinitely many torsion points on C, then we could find such torsion points ξ of arbitrarily high order n. Any Galois conjugate of ξ would also have order n and lie on C, but from our above observation we know that ξ has at least $Dn^{\frac{1}{2}}$ conjugates. The height of $p\pi^{-1}(\xi)$ (and of each conjugate) is at most n. By uniform finiteness in the o-minimal structure $\mathbb{R}_{\mathrm{an,exp}}$ we see that there is a uniform finite bound B on the size of the sets $\pi^{-1}C(\mathbb{C}) \cap (\mathfrak{D} \times \{a\})$ as a ranges through $p(\mathfrak{X})$. Thus, there are at least $(D/B)n^{\frac{1}{2}}$ points in $\mathfrak{C}$ of height at most n, contradicting our hypothesis that there are infinitely many torsion points on C and thereby completing our sketch of the proof of Theorem 5.16.

One might ask whether a variant of the Masser-Zannier theorem holds for higher dimensional families of abelian varieties. Habegger took up precisely that challenge for the Weierstraß family of elliptic curves. For a pair of numbers (a, b), the equation

$$zy^2 = x^3 + az^2x + bz^3$$

cuts out an elliptic curve $E_{(a,b)}$ in $\mathbb{P}^2$ provided that $4a^3 + 27b^2 \neq 0$. From dimension considerations, one sees that one must specify at least three abscissa values in order to expect there to be only finitely many specializations for which all the the points are torsion. Habegger proves that this is in fact the case for the points with abscissa 1, 2 and 3 [**26**].

Theorem 5.21 **(Habegger)**. — *There are only finitely many pairs of complex numbers (a, b) with $8a^3 + 27b^2 \neq 0$ for which the following three points are all torsion in $E_{(a,b)}$:*

$$[1 : \sqrt{1 + a + b} : 1], [2 : \sqrt{8 + 2a + b} : 1], \text{ and } [3 : \sqrt{27 + 3a + b} : 1]$$

While the proof of Theorem 5.21 follows a similar strategy to that of Theorem 5.16, there are some notable novelties. First, while Masser and Zannier could make do with Pila's theorem on counting rational points in subanalytic surfaces [**49**], the definable sets implicated by Theorem 5.21 are complex surfaces, and, hence, four dimensional as definable sets. Thus, the counting theorem for more general o-minimally definable sets makes an essential appearance. Secondly, the transcendence results on theta functions used to show that the definable sets appearing in the proof of Theorem 5.16 have trivial algebraic parts are not strong enough to prove the corresponding theorem in this case. Indeed, some related transcendence theorems of Bertrand [**4**] are used to prove that certain curves have trivial algebraic parts. The counting theorem is then applied in the refined form about the algebraic points being confined to blocks.

5.4. Unlikely intersections in Shimura varieties. — Several theorems along the lines of the Pink-Zilber conjectures have been proven using the methods we have outlined. Of these, the one having the most distinctive character is a theorem of Habegger and Pila [**25**] on unlikely intersections in $\mathbb{A}^n$.

Theorem 5.22 **(Habegger, Pila)**. — *Suppose that $C \subsetneq \mathbb{A}^n$ is an algebraic curve which is asymmetric in the sense that there is at most one pair $i < j \leq n$ of integers for which $\deg(\pi_i \upharpoonright C) \neq \deg(\pi_j \upharpoonright C)$ where $\pi_i : \mathbb{A}^n \to \mathbb{A}^1$ (respectively, π_j) is the projection onto the i^{th} (respectively, j^{th}) coordinate. If C is not contained in a proper special subvariety, then the following set is finite.*

$$C(\mathbb{C}) \cap \bigcup_{\substack{S \subseteq \mathbb{A}^n \\ S \text{ special} \\ \operatorname{codim}(S) \geq 2}} S(\mathbb{C})$$

Remark 5.23. — The hypothesis that the curve is asymmetric is used in the proof, but does not appear to be necessary.

The special subvarieties of codimension two in $\mathbb{A}^n$ are components of varieties defined by two equations of the form $\Phi_N(x_i, x_j) = 0$ where Φ_N is the N^{th} modular polynomial and $x_k = \xi$ where ξ is a special point. Thus, one may convert the problem of describing the intersection C with the codimension two special varieties to that of the intersection of the preimage under the j-function (or, really, the function $\pi : \mathfrak{h} \to \mathbb{A}^n(\mathbb{C})$ which is given by the j-function coordinatewise) of $C(\mathbb{C})$ with the geodesic varieties in $\mathfrak{h}^n$ defined by two equations of the form $\gamma\tau_i = \tau_j$ or $\tau_k = \xi$ where $\gamma \in \mathrm{PSL}_2(\mathbb{Q})$ and ξ is quadratic imaginary. We do not repeat the work from [**25**], but will rather simply note a couple of interesting points from the proof.

The key transcendence result they prove is dual to the Lindemann-Weierstraß theorems appearing earlier, being a functional modular analogue of Baker's theorem on linear forms in logarithms [**3**].

Theorem 5.24 **(Habegger, Pila)**. — *If $\gamma : \Delta \to \mathfrak{h}^n$ is a nonconstant, analytic function for which $(j, \ldots, j) \circ \gamma : \Delta \to \mathbb{A}^n(\mathbb{C})$ is an algebraic function whose image is not contained in any proper special subvariety, then the image of γ is Zariski dense.*

The proof of Theorem 5.24 passes through an analysis of the monodromy of the j-function. This theorem is used to show that it cannot happen that the preimage of $C(\mathbb{C})$ is contained in a *weakly* geodesic subvariety of $\mathfrak{h}^n$. Thus the intersection of the preimage of $C(\mathbb{C})$ with any weakly geodesic variety is necessarily finite.

A second remarkable feature of the proof of Theorem 5.22 is that the refined counting theorem on the number of blocks containing the algebraic points is required since the definable sets in question naturally have the structure of homogeneous spaces for algebraic groups and therefore are equal to their own algebraic parts.

Using similar methods, Pila and Tsimerman have generalized both the Ax-Lindemann-Weierstraß theorem and Theorem 5.24 with an Ax-Schanuel-style theorem for the j-function [**57**]. In fact, generalizing an earlier result of Pila [**54**] strengthening the hyperbolic Ax-Lindemann-Weierstraß theorem to include a transcendence statement for the derivatives of j, the hyperbolic Ax-Schanuel theorem also takes into account the derivatives of the j-function.

5.5. Further results and prospects. — Other theorems in the vein of the Pink-Zilber conjectures have been proven using the Pila-Zannier strategy. Shortly after the strategy became widely known, several results were obtained. Peterzil and Starchenko reproved the Manin-Mumford conjecture for semiabelian varieties using essentially the methods we have described [**45**]. Daw and Yafaev proved the André-Oort conjecture for Hilbert modular surfaces [**15**] and Pila and Tsimerman proved the André-Oort conjecture for the moduli space of principally polarized abelian surfaces [**55**].

It is not unreasonable to hope that these methods may be extended to give an unconditional proof of the André-Oort conjecture for all Shimura varieties, and, in fact, the strategy has been carried out for, arguably, the most important class of Shimura varieties: the coarse moduli spaces $\mathcal{A}_g$ of principally polarized abelian varieties of dimension g for $g \in \mathbb{Z}_+$.

Peterzil and Starchenko established definability in $\mathbb{R}_{\mathrm{an,exp}}$ of the relevant covering maps. Pila and Tsimerman proved the Ax-Lindemann-Weierstraß theorem for $\mathcal{A}_g$

in **[56]**, noting at the time that the André-Oort conjecture for $\mathcal{A}_g$ had thereby been reduced to the problem of proving suitable lower bounds on the size of Galois orbits of special points. Tsimerman completed the proof **[74]** by making use of the average Colmez conjecture proven by two separate groups of mathematicians, Andreatta, Goren, Howard and Madapusi-Pera **[17]**, and Yuan and Zhang **[91]**.

There are no obvious obstructions to extending this method to prove the André-Oort conjecture for general Shimura varieties. Ullmo presented a detailed proof schema for the André-Oort conjecture in **[75]**. He shows that once one has established the o-minimal definability of the covering maps and the Ax-Lindemann-Weierstraß theorem, then a weak form of the André-Oort conjecture follows. That is, if S is a Shimura variety, $Y \subsetneq S$ is an irreducible subvariety which is not expressible as $S' \times Y'$ where S is realizable as $S' \times S''$ with S' and S'' Shimura subvarieties and $Y' \subseteq S''$, then the union of all positive dimensional special subvarieties of Y is not Zariski dense in Y.

Klingler, Ullmo and Yafaev have established the definability in $\mathbb{R}_{\mathrm{an,exp}}$ of all of the relevant covering maps when restricted to suitable fundamental domains as well as the appropriate analogue of the Ax-Lindemann-Weierstraß theorem for all arithmetic varieties, that is, varieties whose complex points may be expressed in the form $\Gamma\backslash\mathbf{X}$ where $\mathbf{X}$ is Hermitian domain and Γ is an arithmetic group, and in particular, for all Shimura varieties **[31]**. The proofs in **[31]** follow the general form of Pila's more concrete version of the hyperbolic Ax-Lindemann-Weierstraß theorem, but the more general form requires more intrinsic, geometric arguments. Whereas the definability of the j-function was deduced from an explicit computation (as is the definability of the covering maps for $\mathcal{A}_g$ in **[46]**) as are the relations between heights and hyperbolic distances, such explicit formulae are unavailable in general.

In related work, Ullmo and Yafaev have characterized the weakly special subvarieties of Shimura varieties as exactly those varieties for which each complex analytic irreducible component of its preimage under the natural covering map is algebraic **[76]**. Yafaev has given an alternative characterization of weakly special varieties **[88]**: If $\pi : \mathbf{X} \to S(\mathbb{C})$ is the map expressing the Shimura variety S as a quotient of the Hermitian domain $\mathbf{X}$ by the action of an arithmetic group, then a subvariety $Y \subseteq S$ is weakly special if and only if each component of $\pi^{-1}Y(\mathbb{C})$ is definable in some o-minimal expansion of $\overline{\mathbb{R}}$.

The problem of producing the requisite lower bounds on the sizes of Galois orbits in general is still open, but under the assumption of the Generalized Riemann Hypothesis, Ullmo and Yafaev have shown that polynomial lower bounds hold for all Shimura varieties **[77]**. This result on the Galois orbits of special points plays an important rôle in the proof by Klingler and Yafaev of the André-Oort conjecture conditionally upon GRH **[32]**. On the other hand, the necessary upper bounds on the heights of pre-special points have been established by Daw and Orr **[14]**. Thus, to complete the proof of the André-Oort conjecture using this strategy, it remains to prove polynomial lower bounds on the size of Galois orbits of special points.

As we have seen with the test cases of simultaneous torsion and curves in products of the j-line, the Pila-Zannier method is well-suited to studying Pink-Zilber-style problems on anomalous intersections with special varieties. More generally, the the use of Theorem 5.24 in the proof of an Ax-Schanuel theorem for the j-function, suggests that a an Ax-Schanuel theorem for all arithmetic varieties will be useful in attacking Pink-Zilber-type conjectures.

Orr has employed this method to prove some instances of the André-Pink conjecture, a restriction of the Zilber-Pink conjectures [**42, 43**]. The André-Pink conjecture predicts that if S is a Shimura variety, $\Sigma \subseteq S(\mathbb{C})$ is the (generalized) Hecke orbit of some point, and $Y \subseteq S$ is an irreducible subvariety for which $\Sigma \cap Y(\mathbb{C})$ is Zariski dense in Y, then Y must be a weakly special variety. Orr has verified this conjecture in the case that S is a Shimura variety of abelian type and Y is a curve. More generally, he has shown that unless Y is a point, it must take the form $S' \times Y'$ where S' and S'' are Shimura varieties, $\dim S' > 0$, $S' \times S'' \hookrightarrow S$ is a weakly special subvariety of S, and $Y' \subseteq S''$ is a subvariety of S''. One might expect that by passing to a quotient or to a smaller ambient Shimura variety, the argument could be completed by induction. However, there are some complications in implementing this strategy. First, the presentation of $S' \times S''$ as a product may be incompatible with the structure of S as a Shimura variety. Secondly, one cannot simply replace S with S_0, the smallest Shimura subvariety of S containing Y, because it may happen that $S_0 \cap \Sigma$ decomposes into infinitely many generalized Hecke orbits in S_0. Orr deals successfully with the first difficulty, but the second obstruction remains.

Generalizing both the André-Oort and André-Pink conjectures, Gao has implemented the Pila-Zannier strategy for mixed Shimura varieties [**19**]. Specifically, he has established the Ax-Lindemann-Weierstraß theorem for mixed Shimura varieties [**22**] and has shown that the generalization of the André-Oort conjecture for mixed Shimura varieties follows from the existence of suitable lower bounds on the size of Galois orbits of special points on the associated pure Shimura varieties [**20**]. Specializing to the case of universal abelian schemes over fine moduli spaces, Gao proves a generalization of Orr's theorem describing those subvarieties which meet a generalized Hecke orbit in a Zariski dense set [**21**].

Even after this flurry of activity, there remain many cases of the Zilber-Pink conjecture to which the Pila-Zannier strategy should apply. Methods from o-minimal geometry have become important tools in diophantine geometry. I expect that in time several further theorems towards the Zilber-Pink conjectures and related problems will be proven following these ideas.

References

[1] S. A. Amitsur. On the characteristic polynomial of a sum of matrices. *Linear and Multilinear Algebra*, 8(3):177–182, 1979/80.

[2] James Ax. On Schanuel's conjectures. *Ann. of Math. (2)*, 93:252–268, 1971.

[3] A. Baker. Linear forms in the logarithms of algebraic numbers. I, II, III. *Mathematika 13 (1966), 204-216; ibid. 14 (1967), 102-107; ibid.*, 14:220–228, 1967.

[4] D. Bertrand. Extensions de D-modules et groupes de Galois différentiels. In *p-adic analysis (Trento, 1989)*, volume 1454 of *Lecture Notes in Math.*, pages 125–141. Springer, Berlin, 1990.

[5] Daniel Bertrand and Anand Pillay. A Lindemann-Weierstrass theorem for semi-abelian varieties over function fields. *J. Amer. Math. Soc.*, 23(2):491–533, 2010.

[6] Edward Bierstone and Pierre D. Milman. Semianalytic and subanalytic sets. *Inst. Hautes Études Sci. Publ. Math.*, (67):5–42, 1988.

[7] E. Bombieri and J. Pila. The number of integral points on arcs and ovals. *Duke Math. J.*, 59(2):337–357, 1989.

[8] W. Dale Brownawell and K. K. Kubota. The algebraic independence of Weierstrass functions and some related numbers. *Acta Arith.*, 33(2):111–149, 1977.

[9] Lee Butler. *The density of algebraic points in sets definable in o-minimal structures.* PhD thesis, University of Bristol, 2012.

[10] Antoine Chambert-Loir and François Loeser. A non-archimedean Ax-Lindemann theorem. `arXiv:1511.03259`, 2015.

[11] Raf Cluckers, Georges Comte, and François Loeser. Non-archimedian Yomdin-Gromov parametrizations and points of bounded height. *Forum of Mathematics, Pi*, 3:60, 2015.

[12] Robert F. Coleman. On a stronger version of the Schanuel-Ax theorem. *Amer. J. Math.*, 102(4):595–624, 1980.

[13] Sinnou David. Points de petite hauteur sur les courbes elliptiques. *J. Number Theory*, 64(1):104–129, 1997.

[14] Christopher Daw and Martin Orr. Heights of pre-special points of shimura varieties. `arXiv:1502.00822`, 2015.

[15] Christopher Daw and Andrei Yafaev. An unconditional proof of the André-Oort conjecture for Hilbert modular surfaces. *Manuscripta Math.*, 135(1-2):263–271, 2011.

[16] J. Denef and L. van den Dries. p-adic and real subanalytic sets. *Ann. of Math. (2)*, 128(1):79–138, 1988.

[17] Fabrizio Fabrizio Andreatta, Eyal Z. Goren, Benjamin Howard, and Keerthi Madapusi-Pera. Faltings heights of abelian varieties with complex multiplication. `arXiv:1508.00178`, 2015.

[18] A. M. Gabrièlov. Projections of semianalytic sets. *Funkcional. Anal. i Priložen.*, 2(4):18–30, 1968.

[19] Ziyang Gao. *The mixed Ax-Lindemann theorem and its applications to the Zilber-Pink conjecture.* PhD thesis, Université Paris-Sud and Universiteit Leiden, 2014.

[20] Ziyang Gao. About the mixed André-Oort conjecture: reduction to a lower bound for the pure case. available at `http://webusers2.imj-prg.fr/~ziyang.gao/ziyang_gao_en.html`, 2015.

[21] Ziyang Gao. A special point problem of André-Pink-Zannier in the universal family of abelian varieties. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze*, 2015. (to appear).

[22] Ziyang Gao. Towards the André-Oort conjecture for mixed Shimura varieties: the Ax-Lindemann theorem and lower bounds for Galois orbits of special points. *J. Reine Angew. Math.*, 2015.

[23] Kurt Gödel. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatsh. Math. Phys.*, 37(1):349–360, 1930.

[24] M. Gromov. Entropy, homology and semialgebraic geometry. *Astérisque*, (145-146):5, 225–240, 1987. Séminaire Bourbaki, Vol. 1985/86.

[25] P. Habegger and J. Pila. Some unlikely intersections beyond André-Oort. *Compos. Math.*, 148(1):1–27, 2012.

[26] Philipp Habegger. Torsion points on elliptic curves in Weierstrass form. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 12(3):687–715, 2013.

[27] G. O. Jones and Margaret M. E. Thomas. The density of algebraic points on certain Pfaffian surfaces. *Quart. J. Math.*, pages 1–15, 2011.

[28] A. G. Khovanskiĭ. *Fewnomials*, volume 88 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1991. Translated from the Russian by Smilka Zdravkovska.

[29] Jonathan Kirby. Corrigendum to: "A Schanuel condition for Weierstrass equations" [J. Symbolic Logic **70** (2005), no. 2, 631–638]. *J. Symbolic Logic*, 70(3):1023, 2005.

[30] Jonathan Kirby. A Schanuel condition for Weierstrass equations. *J. Symbolic Logic*, 70(2):631–638, 2005.

[31] B. Klingler, E. Ullmo, and A. Yafaev. The hyperbolic Ax-Lindemann-Weierstraß conjecture. *Publications mathématiques de l'IHÉS*, pages 1–28, 2015.

[32] Bruno Klingler and Andrei Yafaev. The André-Oort conjecture. *Ann. of Math. (2)*, 180(3):867–925, 2014.

[33] Julia F. Knight, Anand Pillay, and Charles Steinhorn. Definable sets in ordered structures. II. *Trans. Amer. Math. Soc.*, 295(2):593–605, 1986.

[34] Serge Lang. *Number theory. III*, volume 60 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 1991. Diophantine geometry.

[35] F. Lindemann. Über die Zahl π. *Math. Ann.*, 20(2):213–225, 1882.

[36] Jerzy Łoś. The algebraic treatment of the methodology of elementary deductive systems. *Studia Logica*, 2:151–212, 1955.

[37] David Marker. Model theory for algebra and algebraic geometry. Lectures at Orsay, available at `http://homepages.math.uic.edu/~marker/orsay/`, 2010.

[38] D. Masser and U. Zannier. Torsion anomalous points and families of elliptic curves. *Amer. J. Math.*, 132(6):1677–1691, 2010.

[39] D. Masser and U. Zannier. Torsion points on families of squares of elliptic curves. *Math. Ann.*, 352(2):453–484, 2012.

[40] D. W. Masser. Small values of the quadratic part of the Néron-Tate height on an abelian variety. *Compositio Math.*, 53(2):153–170, 1984.

[41] 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.

[42] Martin Orr. *La conjecture d'André–Pink : Orbites de Hecke et sous-variétés faiblement spéciales*. PhD thesis, Université Paris-Sud, 2013.

[43] Martin Orr. Families of abelian varieties with many isogenous fibres. *J. Reine Angew. Math.*, 705:211–231, 2015.

[44] 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.

[45] Ya'acov Peterzil and Sergei Starchenko. Around Pila-Zannier: the semiabelian case. *preprint*, 2009.

[46] Ya'acov Peterzil and Sergei Starchenko. Definability of restricted theta functions and families of abelian varieties. arXiv:1103.3110, 2011.

[47] J. Pila. Density of integral and rational points on varieties. *Astérisque*, (228):4, 183–187, 1995. Columbia University Number Theory Seminar (New York, 1992).

[48] J. Pila and A. J. Wilkie. The rational points of a definable set. *Duke Math. J.*, 133(3):591–616, 2006.

[49] Jonathan Pila. Integer points on the dilation of a subanalytic surface. *Q. J. Math.*, 55(2):207–223, 2004.

[50] Jonathan Pila. Rational points on a subanalytic surface. *Ann. Inst. Fourier (Grenoble)*, 55(5):1501–1516, 2005.

[51] Jonathan Pila. The density of rational points on a Pfaff curve. *Ann. Fac. Sci. Toulouse Math. (6)*, 16(3):635–645, 2007.

[52] Jonathan Pila. On the algebraic points of a definable set. *Selecta Math. (N.S.)*, 15(1):151–170, 2009.

[53] Jonathan Pila. O-minimality and the André-Oort conjecture for $\mathbb{C}^n$. *Ann. of Math. (2)*, 173(3):1779–1840, 2011.

[54] Jonathan Pila. Modular Ax-Lindemann-Weierstrass with derivatives. *Notre Dame J. Form. Log.*, 54(3-4):553–565, 2013.

[55] Jonathan Pila and Jacob Tsimerman. The André-Oort conjecture for the moduli space of abelian surfaces. *Compos. Math.*, 149(2):204–216, 2013.

[56] Jonathan Pila and Jacob Tsimerman. Ax-Lindemann for $\mathcal{A}_g$. *Ann. of Math. (2)*, 179(2):659–681, 2014.

[57] Jonathan Pila and Jacob Tsimerman. Ax-Schanuel for the j-function. *Duke Math. J.*, (to appear). `arXiv:1412.8255`.

[58] 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.

[59] Anand Pillay and Charles Steinhorn. Definable sets in ordered structures. I. *Trans. Amer. Math. Soc.*, 295(2):565–592, 1986.

[60] Anand Pillay and Charles Steinhorn. Definable sets in ordered structures. III. *Trans. Amer. Math. Soc.*, 309(2):469–476, 1988.

[61] Richard Pink. A combination of the conjectures of Mordell-Lang and André-Oort. In *Geometric methods in algebra and number theory*, volume 235 of *Progr. Math.*, pages 251–282. Birkhäuser Boston, Boston, MA, 2005.

[62] 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.

[63] Christophe Reutenauer and Marcel-Paul Schützenberger. A formula for the determinant of a sum of matrices. *Lett. Math. Phys.*, 13(4):299–302, 1987.

[64] 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.

[65] Thomas Scanlon. The André-Oort conjecture via o-minimality (after J. Pila, A. Wilkie, and U. Zannier). 2011. Séminaire Bourbaki. Vol. 2010/2011.

[66] Thomas Scanlon. A Euclidean Skolem-Mahler-Lech-Chabauty method. *Math. Research Letters*, 18(5):833–842, 2011.

[67] Thomas Scanlon. Counting special points: logic, diophantine geometry and transcendence theory. *Bull. Amer. Math. Soc.*, 49:51–71, 2012.

[68] C. L. Siegel. Über die Classenzahl quadratischer Zahlkörper. *Acta Arith.*, (1):83–86, 1935.

[69] Joseph H. Silverman. Heights and the specialization map for families of abelian varieties. *J. Reine Angew. Math.*, 342:197–211, 1983.

[70] Patrick Speissegger. The Pfaffian closure of an o-minimal structure. *J. Reine Angew. Math.*, 508:189–211, 1999.

[71] Michael Stoll. Simultaneous torsion in the Legendre family. `arXiv:1410.7070`, 2014.
[72] Alfred Tarski. *A Decision Method for Elementary Algebra and Geometry*. RAND Corporation, Santa Monica, Calif., 1948.
[73] Margaret E. M. Thomas. Convergence results for function spaces over o-minimal structures. *J. Logic & Analysis*, 4(1):1–14, 2012.
[74] Jacob Tsimerman. A proof of the André-Oort conjecture for $\mathscr{A}_g$. `arXiv:1506.01466`.
[75] Emmanuel Ullmo. Applications du théorème d'Ax-Lindemann hyperbolique. *Compos. Math.*, 150(2):175–190, 2014.
[76] Emmanuel Ullmo and Andrei Yafaev. A characterization of special subvarieties. *Mathematika*, 57(2):263–273, 2011.
[77] Emmanuel Ullmo and Andrei Yafaev. Galois orbits and equidistribution of special subvarieties: towards the André-Oort conjecture. *Ann. of Math. (2)*, 180(3):823–865, 2014.
[78] Emmanuel Ullmo and Andrei Yafaev. Hyperbolic Ax-Lindemann theorem in the cocompact case. *Duke Math. J.*, 163(2):433–463, 2014.
[79] Lou van den Dries. Remarks on Tarski's problem concerning $(\mathbf{R}, +, \cdot, \exp)$. In *Logic colloquium '82 (Florence, 1982)*, volume 112 of *Stud. Logic Found. Math.*, pages 97–121. North-Holland, Amsterdam, 1984.
[80] Lou van den Dries. A generalization of the Tarski-Seidenberg theorem, and some nondefinability results. *Bull. Amer. Math. Soc. (N.S.)*, 15(2):189–193, 1986.
[81] Lou van den Dries. *Tame topology and o-minimal structures*, volume 248 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1998.
[82] 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.
[83] 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.
[84] C. Weierstraß. On Lindemann's paper: "On Ludolph's number". (Zu Lindemanns Abhandlung: "Über die Ludolph'sche Zahl".). *Berl. Ber.*, pages 1067–1086, 1885.
[85] 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.
[86] A. J. Wilkie. Diophantine properties of sets definable in o-minimal structures. *J. Symbolic Logic*, 69(3):851–861, 2004.
[87] Alex J. Wilkie. O-minimal structures. *Astérisque*, (326):Exp. No. 985, vii, 131–142 (2010), 2009. Séminaire Bourbaki. Vol. 2007/2008.
[88] Andrei Yafaev. An o-minimal characterization of special subvarieties. *Bulletin of the London Mathematical Society*, 46(4):732–736, 2014.
[89] Y. Yomdin. C^k-resolution of semialgebraic mappings. Addendum to: "Volume growth and entropy". *Israel J. Math.*, 57(3):301–317, 1987.
[90] Y. Yomdin. Volume growth and entropy. *Israel J. Math.*, 57(3):285–300, 1987.
[91] Xinyi Yuan and Shou-Wu Zhang. On the averaged Colmez conjecture. `arXiv:1507.06903`, 2015.

THOMAS SCANLON • *E-mail :* `scanlon@math.berkeley.edu`, Department of Mathematics, University of California, Berkeley, Evans Hall, Berkeley, CA 94720-3840, USA
Url : `http://www.math.berkeley.edu/~scanlon`