THEOREMS ON UNLIKELY INTERSECTIONS BY COUNTING POINTS IN DEFINABLE SETS

THOMAS SCANLON

1. Comments on these notes

This document consists of my notes prepared in advance of a series of lectures at Luminy as part of the program on the Zilber-Pink conjecture. Since I did not meet the deadline imposed by the institute administrators, many parts of this paper are incomplete. Let me warn the reader that there are certainly errors in these notes, references are missing and they are not organized as I would wish. I plan to distribute a better version just before the lectures begin.

I have delivered a couple of general audience lectures about this topic and have prepared accompanying expository papers [21, 20]. The reader desiring a smoother presentation may prefer to read those notes and will likely notice that some of the present text was lifted from those sources. While I have made an effort to present some of the arguments in greater detail in these notes than in the earlier notes, the proofs are far from complete. The original papers on which this material is based are clearly written. I recommend that the reader study those papers directly.

2. An outline of the Pila-Zannier strategy

With this lecture we shall describe the key steps in the Pila-Zannier strategy leaving most of the key details to later lectures. This strategy was first successfully implemented by Pila and Zannier in [16] with their reproof of the Manin-Mumford conjecture (Raynaud's theorem [19]). Shortly thereafter, Masser and Zannier employed these ideas to give a proof of the first non-trivial instance of the Zilber-Pink conjecture for which the special varieties have positive dimension. Subsequently, various authors have extended these methods to prove other cases of the Zilber-Pink conjecture, most notably, Pila with his unconditional proof of the André-Oort conjecture for products of modular curves [15].

Let us consider two abstract forms of the Zilber-Pink conjecture. The first theorem template might be considered as an abstract version of the Manin-Mumford or André-Oort conjecture.

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

Theorem Template 2.1. *Let X be a* special variety *over a field K and $\Sigma \subseteq X(K)$ a set of* special points *on X, then the Zariski closure of Σ is a finite union of* special subvarieties*.*

Of course, Theorem Template 2.1 has no content unless the terms "special point" and "special variety" are given some geometric meaning. In the case that X is a semiabelian variety over $K = \mathbb{C}$, if one were to read "torsion point" for "special point" and "translate of a group subvariety by a torsion point" for "special variety," then this instance of Theorem Template 2.1 is precisely the Manin-Mumford conjecture. If we take X to be a Shimura variety over $K = \mathbb{C}$, then, as written, Theorem Template 2.1 is the André-Oort conjecture where the terms "special point" and "special variety" are defined relative to Deligne's theory of Shimura data.

An abstract form of the Zilber-Pink conjecture takes the following form.

Theorem Template 2.2. *Let X be a special variety over a field K and $Y \subseteq X$ an arbitrary subvariety. If Y is not a special subvariety of X, then the following set is* not *Zariski dense in Y.*

$$\bigcup_{\substack{Z \subseteq X \\ Z \text{ a special subvariety} \\ \dim(Z) + \dim(Y) < \dim(X)}} Z(K) \cap Y(K)$$

As with Theorem Template 2.1, Theorem Template 2.2 asserts something nontrivial only when the notions of special variety and special subvariety are made precise. Visibly, Theorem Template 2.2 (for a given X, Y and interpretation of the word "special") implies Theorem Template 2.1. While this formal implication is undeniable, there is a sense in which these theorem templates have qualitatively different characters. Theorem Template 2.1 describes the possible algebraic relations on the set of special points of X while Theorem Template 2.2 is more geometric in that it predicts the possible interactions between special algebraic relations and general algebraic relations. Paradoxically, methods from o-minimality are better suited to instances of Theorem Template 2.2 than they are to Theorem Template 2.1 in that in many cases of interest questions the class of special subvarieties of a given variety may be realized as the restriction of a definable family of definable sets. The issues implicated by Theorem Template 2.2 are thereby transformed into questions about definable sets. Admittedly, this comment is too vague to be meaningful and, from our earlier observation, facially false, but I hope that by the end of these lectures the sense in which I intend it has been made clear.

Let us sketch the Pila-Zannier method applied to Theorem Template 2.1. So that the reader may follow the proof of a true theorem we shall instantiate Theorem Template with a particularly simple case of the Manin-Mumford conjecture, namely, Mann's theorem about algebraic relation on roots of unity [9].

Theorem 2.3. *Let $g \geq 1$ be a positive integer and $Y \subseteq \mathbb{G}_m^g$ a subvariety of the g^{th} Cartesian power of the multiplicative group over the complex numbers. Then the set*

$$\{(\zeta_1, \ldots, \zeta_g) \in Y(\mathbb{C}) \ : \ \zeta_i \text{ is a root of unity for each } i \leq g\}$$

is a finite union of cosets of subgroups of $(\mathbb{C}^\times)^g$.

Remark 2.4. Theorem 2.3 is an instance of Theorem Template 2.1 in which we take the ambient special variety X to be $\mathbb{G}_m^g$, the special points to be g-tuples of roots of unity, Σ to be the displayed set in the statement of Theorem 2.3 and the special subvarieties to be the components of group subvarieties of $\mathbb{G}_m^g$.

Remark 2.5. There are several non-trivial difficulties with implementing the Pila-Zannier strategy in general which disappear in this special case we are considering now. In the later lectures we shall go into detail about these other theorems.

The first step of this method is to find some analytic covering map $\pi : \mathfrak{X} \to X(\mathbb{C})$. For this to be a useful move, we shall require some stronger properties of π and $\mathfrak{X}$ than merely that π be analytic. Returning to the case of Mann's theorem, $\mathfrak{X} = \mathbb{C}^g$ and the map $\pi : \mathbb{C}^g \to \mathbb{G}_m^g(\mathbb{C})$ is simply $(z_1, \ldots, z_g) \mapsto (e^{2\pi i z_1}, \ldots, e^{2\pi i z_n})$ where $\mathfrak{X} = \mathbb{C}^g$.

One requirement for the map $\pi : \mathfrak{X} \to X(\mathbb{C})$ to be useful is that the inverse image of the special points in $X(\mathbb{C})$ under π, what we shall call *pre-special points*, should also be "special points" in $\mathfrak{X}$. Again thinking of the case of Mann's theorem, the special points on $X(\mathbb{C})$ are g-tuples of roots of unity and the set of pre-special points is precisely $\mathbb{Q}^g$.

Assuming that the theorem is false, replacing Y by a component of some nonspecial component of the Zariski closed of the set of special points on Y, we may assume that Y is *not* a special subvariety but that the set of special points on Y is Zariski dense in Y. Let $\mathfrak{Y} := \pi^{-1}Y(\mathbb{C})$. Then $\mathfrak{Y}$ is a (not necessarily irreducible) analytic subvariety of $\mathfrak{X}$ which contains "many" pre-special points. Returning to Mann's theorem, $\mathfrak{Y}$ contains many rational points. That is, we have reduced the problem of describing the special points on $Y(\mathbb{C})$ to that of describing the pre-special points on the analytic variety $\mathfrak{Y}$.

At this point in the sketch, the reader would be justified in rejecting this strategy out of hand as there is almost nothing that one can say about rational points on analytic sets in general. On the one hand, for any set $S \subseteq \mathbb{Z}^g$ we can find an entire analytic function $f(x_1, \ldots, x_g)$ for which $f(z_1, \ldots, z_g) = 0$ just in case $(z_1, \ldots, z_g) \in S$. Thus, there is no restriction at all on a subset of $\mathbb{Z}^g$ to be the rational points on an analytic variety. On the other hand, even if we insist that $Z \subseteq \mathbb{C}^g$ is a particularly well behaved analytic set, then it may be very hard to describe $Z \cap \mathbb{Q}^g$. Indeed, if we ask that Z is the set of complex points of an algebraic variety, then problem of determining $Z \cap \mathbb{Q}^g$ is the central task of arithmetic geometry and is (conjecturally) an intractable problem. So, it would seem that we have converted the difficult problem of determining the set of special points on Y to the impossible task of classifying the possible sets of pre-special points on the analytic variety $\mathfrak{Y}$.

For our argument to succeed, we must steer clear of these extremes of a completely general analytic variety and algebraic varieties. We avoid the pathologies of general analytic varieties by requiring (a suitable restriction of) π to be *definable* in an o-minimal expansion of the real numbers. We skirt the problems in arithmetic geometry by showing that for purposes of this argument, we may simply ignore the rational points contained in algebraic subvarieties.

Let us delay a detailed discussion of definability and take the notion of a set definable in an o-minimal expansion of the real numbers as a black box for the moment, though at least to keep this sketch honest we should make the suitable restriction somewhat more explicit. Even in this special case of Mann's theorem where the function π is simply given by the complex exponential function, one must limit the domain of π in order to have any hope of its logical theory being reasonable. Indeed, using the complex exponential function itself, since one may define the integers as $\{z \in \mathbb{C} \ : \ e^{2\pi i z} = 1\}$, one may encode all of arithmetic in the theory of the complex field given together with the exponential function. As follows from Gödel's Incompleteness theorem, the resulting theory is wild. On the other hand, for the purposes of recognizing roots of unity as special values of the exponential function, since the exponential function is periodic, it is not necessary to consider the action of the exponential function everywhere. In the case of Mann's theorem, we could take $\mathfrak{X} := \{(z_1, \ldots, z_g) \in \mathbb{C}^g \ : \ 0 \leq \mathrm{Re}(z_i) < 1 \text{ for each } i \leq g\}$. With such a restriction the function $\pi : \mathfrak{X} \to X(\mathbb{C}) = (\mathbb{C}^\times)^g$ is definable in the o-minimal structure $\mathbb{R}_{an}$.

Remark 2.6. The astute reader has observed that in the case of Mann's theorem, one could work with an even smaller domain. That is, if e^a is root of unity, then a is a rational multiple of πi. Thus, if we were to take $\mathfrak{X} := [0,1]^g$ and $\nu : \mathfrak{X} \to X(\mathbb{C})$ defined by $(x_1, \ldots, x_g) \mapsto (\exp(2\pi i x_1), \ldots, \exp(2\pi i x_g))$, then it already the case that every special point in $(\mathbb{C}^\times)^g$ is in the image of ν. Thus, we could analyze the set of special points on the variety $Y \subseteq \mathbb{G}_m^g$ as the image under ν of the set of rational points on $\nu^{-1}Y(\mathbb{C})$. While this move would simplify the argument for Mann's theorem, it is somewhat spurious as one cannot perform such a reduction for most of the other cases (*eg* André-Oort or Manin-Mumford) to which this method applies.

For purposes of this argument, the most important consequence of $\mathfrak{Y}$ being definable in a o-minimal structure on the real numbers is a counting theorem of Pila and Wilkie asserting that if one avoids its *algebraic part*, there are few rational points on $\mathfrak{Y}$. Before we make this result precise, let us return to the problem at hand to see how such a theorem may be used. Our algebraic variety Y is not special, but it contains a Zariski dense set of special points. We shall show that the image under π of the algebraic part of $\mathfrak{Y}$ is contained in a proper subvariety of Y. We then conclude that there are many pre-special points in $\mathfrak{Y}$ outside of its algebraic part where "many" means that the image of this set under π is Zariski dense in Y. The Pila-Wilkie counting theorem is quantitative and on the face of it, there could be a set of points with Zariski dense image in Y but whose distribution is so sparse that one would not observe a contradiction. To reach a contradiction, we must employ a Galois theoretic argument to show that in this case, such sets must be fairly large.

Let us flesh out this argument beginning with a precise statement of the counting theorem. For this we need to say what we mean by the algebraic part of a set.

Definition 2.7. We say that a subset of $\mathbb{R}^n$ is *semi-algebraic* if it is a finite boolean combination of sets of the form $\{(x_1, \ldots, x_n) \in \mathbb{R}^n : f(x_1, \ldots, x_n) \geq 0\}$ where f is a polynomial over the real numbers. We say that a subset of $\mathbb{C}^n$ is semi-algebraic if its image in $\mathbb{R}^{2n}$ under the map $(z_1, \ldots, z_n) \mapsto (\mathrm{Re}(z_1), \mathrm{Im}(z_1), \ldots, \mathrm{Re}(z_n), \mathrm{Im}(z_n))$ is semi-algebraic. A function is semi-algebraic if its graph is. Given a set $Y \subseteq \mathbb{C}^n$ we define the *algebraic part* of Y, written Y^{alg}, to be the union of the ranges of all nonconstant semi-algebraic functions $\gamma : (0,1) \to Y$. We define the *transcendental part* of Y, written Y^{tr}, to be $Y \smallsetminus Y^{\mathrm{alg}}$.

Remark 2.8. We shall go into much more detail about the construction of the algebraic part of a set in the third lecture. For now, let us simply observe that it follows from a curve selection theorem in semi-algebraic geometry that if Y is itself semi-algebraic, connected and infinite, then it is equal to its own algebraic part.

Let us say now how we shall count the rational points in a definable set. Recall that the (multiplicative) height of a rational numbers is defined by $H(0) = 0$ and $H(\frac{a}{b}) := \max\{|a|, |b|\}$ provided that $\gcd(a,b) = 1$. We extend H to n-tuples by $H(x_1, \ldots, x_n) := \max\{H(x_i) : 1 \leq i \leq n\}$. For a set $X \subseteq \mathbb{R}^n$ and $t \in \mathbb{R}_+$, we define $X(t) := \{a \in X : H(a) \leq t\}$. In its simplest form, the Pila-Wilkie counting theorem asserts that if $X \subseteq \mathbb{R}^n$ is o-minimally definable, then there are subexponentially many rational points on the transcendental part of X.

Theorem 2.9 (Pila-Wilkie)**.** *Let $X \subseteq \mathbb{R}^n$ be o-minimally definable and $\epsilon > 0$. Then there is a constant $C = C(X, \epsilon)$ so that for all numbers $t \geq 1$ we have $\#X^{\mathrm{tr}}(t) \leq Ct^{\epsilon}$*

Returning to our sketch of the strategy to prove Theorem Template 2.1, to use Theorem 2.9 we should first determine $\mathfrak{Y}^{\mathrm{alg}}$, relating this set to the special subvarieties of Y. For Mann's theorem, this is achieved by applying a function field version of the Schanuel conjecture due to Ax.

Theorem 2.10. *Suppose that $\gamma_1(t), \ldots, \gamma_n(t) \in t\mathbb{C}[[t]]$ are complex power series with no constant terms which are linearly independent over $\mathbb{Q}$. Then the transcendence degree over $\mathbb{C}$ of the field $\mathbb{C}(\gamma_1, \ldots, \gamma_n, \exp(\gamma_1), \ldots, \exp(\gamma_n))$ is at least $n+1$.*

We use Theorem 2.10 to show that the image of $\mathfrak{Y}^{\mathrm{alg}}$ under π is the *Ueno locus* of Y:

$$\mathrm{Ueno}(Y) := \bigcup_{\substack{aH \subseteq X \\ H \leq \mathbb{G}_m^g \\ \dim(H) > 0}} aH$$

Remark 2.11. While it is not immediately obvious from the definition, it is not hard to show that the Ueno locus is actually a (not necessarily irreducible) subvariety of Y and that $Y = \mathrm{Ueno}(Y)$ if and only if the stabilizer of Y is infinite.

Of course, if $\pi(\xi) \in \mathrm{Ueno}(Y)$, then $\xi \in \mathfrak{Y}^{\mathrm{alg}}$ as there is a line with rational slope passing through ξ and contained in $\mathfrak{Y}$. Let us show now that every semialgebraic curve on $\mathfrak{Y}$ lies over the Ueno locus. Suppose now that $\gamma : (0,1) \to \mathfrak{Y}$ is a nonconstant semialgebraic function. Applying some minor reductions (decomposing $(0,1)$ into finitely many intervals and possibly precomposing with a monomial), we may assume that γ is real analytic on an interval properly containing $(0,1)$. Choosing coordinates on $\mathbb{C}^g$ around $\gamma(0)$, we may write γ in coordinates as $(\gamma_1(t), \ldots, \gamma_g(t))$ where $\gamma_i(0) = 0$ for each $i \leq g$. Since the image of γ is contained in $\mathfrak{Y}$, applying $\pi : \mathfrak{X} \to \mathbb{G}_m^g(\mathbb{C})$, we see that $(\exp(2\pi i \gamma_1(t)), \ldots, \exp(2\pi i \gamma_g(t)) \in Y(\mathbb{C})$ for all $t \in (0,1)$. Thus, the functions $\exp(2\pi i \gamma_i \gamma_1(t)), \ldots, \exp(2\pi i \gamma_g(t))$ are algebraically dependent over $\mathbb{C}$. As γ is semialgebraic, the transcendence degree over $\mathbb{C}$ of $\mathbb{C}(\gamma_1(t), \ldots, \gamma_g(t)) = 1$. Thus,

$$\operatorname{tr.deg}_{\mathbb{C}} \mathbb{C}(\gamma_1(t), \ldots, \gamma_g(t), \exp(2\pi i \gamma_1(t)), \ldots, \exp(2\pi i \gamma_g(t))) \leq g$$

so that by Theorem 2.10 the functions $\gamma_1, \ldots, \gamma_g$ must be linearly dependent over $\mathbb{Q}$, but the image of such a translate of a vector space defined by a the vanishing of a linear form over $\mathbb{Q}$ under π is simply a translate of an algebraic subgroup variety of $\mathbb{G}_m^g$. Let us call this subgroup H and $\pi(\gamma(0)) =: a$. Passing to a connected component of H if need be, we know that $H \cong \mathbb{G}_m^t$ for some $t < g$. Via this isomorphism composed with a translation, we may regard $Y \cap aH$ as a subvariety of $\mathbb{G}_m^t$ and the curve γ as a semialgebraic curve on the preimage of this variety. Hence, by induction, the image of γ is contained in the Ueno locus of $aH \cap Y$, which is contained in the Ueno locus of Y.

Remark 2.12. In the general case, various ideas may come to bear on this part of the problem of computing $\mathfrak{Y}^{\mathrm{alg}}$. As with Mann's theorem, analogues of Ax's theorem are known for the exponential maps associated to abelian varieties and an argument like the one sketched above may be used to show that $\mathfrak{Y}^{\text{alg}}$ is the preimage of the Ueno locus of Y. Ax's proof is differential algebraic in nature and one might expect that suitable analogues for the covering maps appearing in the theory of Shimura varieties, for example, could be proven through a consideration of the differential equations they satisfy. However, to date the differential algebraic techniques have not led to complete proofs. In Pila's proof of the André-Oort conjecture for products of modular curves, $\mathfrak{Y}^{\text{alg}}$ is characterized through another application of the counting theorem. With the Masser-Zannier theorem on simultaneous torsion, results on transcendence of theta functions are used.

Quotienting by the stabilizer of Y if need be, we reduce the problem to showing that there are only finitely many special points on the complement to $\text{Ueno}(Y)$ in Y. Equivalently, since we know that $\mathfrak{Y}^{\text{alg}}$ is the preimage of the Ueno locus of Y, we must show that there can be only finitely many rational points in $\mathfrak{Y}^{\text{tr}}$.

Let us observe that since Y contains a Zariski dense set of special points, each of which is algebraic, Y itself is defined over some number field, L. Moreover, it is clear that $\text{Ueno}(Y)$ is preserved by the action of the $\text{Gal}(\mathbb{Q}^{\text{alg}}/L)$. Thus, if $\xi \in \mathfrak{Y}^{\text{tr}} \cap \mathbb{Q}^g$ and $\sigma \in \text{Gal}(\mathbb{Q}^{\text{alg}}/L)$, there some point $\xi_\sigma \in \mathfrak{Y}^{\text{tr}} \cap \mathbb{Q}^g$ for which $\sigma(\pi(\xi)) = \pi(\xi_\sigma))$. In the case, it is very easy to describe the point ξ_σ. Indeed, we may write $\xi = (\frac{a_1}{b_1}, \ldots, \frac{a_g}{b_g})$ with $0 \leq a_i < b_i$, $a_i, b_i \in \mathbb{Z}$, $(a_i, b_i) = 1$ unless $a_i = 0$ and $b_i = 1$. Then $\sigma(\xi) = (\frac{a_1'}{b_1}, \ldots, \frac{a_g'}{b_g})$ with $a_i' < b_i$ and coprime to b_i unless $a_i = a_i' = 0$. As we know, if $n > 2$ and ζ_n is a primitive n^{th} root of unity, then $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$ where φ is Euler's totient function. Moreover, it is easy matter to show that for n big enough, $\varphi(n) > n/\log(n) > n^{\frac{1}{2}}$.

Let $C = C(\mathfrak{Y}, \frac{1}{3})$ be the constant of Theorem 2.9. It follows that $\mathfrak{Y}^{\text{tr}} \cap \mathbb{Q}^g \subseteq \mathfrak{Y}((C[L:\mathbb{Q}])^6)$ as if there were some point $\xi \in \mathfrak{Y}^{\text{tr}} \cap \mathbb{Q}^g$ with $H(\xi) = n > (C[K:\mathbb{Q}])^6$, then from the above calculation of the size of the Galois orbit of $\pi(\xi)$, we see that there would be at least $\frac{1}{[K:\mathbb{Q}]} n^{\frac{1}{2}}$ points in $\mathfrak{Y}^{\text{tr}}(n)$ while the counting theorem limits this set to $Ct^{\frac{1}{3}}$ points which contradicts the hypothesis that $n > (C[K:\mathbb{Q}])^6$.

This completes the sketch of the Pila-Zannier strategy implemented in the case of Mann's theorem. □

For the most part the instances of Theorem Template 2.2 proven via this strategy follow the general form of the argument we sketched for Mann's theorem. This is true even for the Masser-Zannier theorem on simultaneous torsion which is not an instance of the weaker Theorem Template 2.1. The recent work of Habegger and Pila on anomalous intersections in Shimura varieties is the first case where the natural extension of these methods to the more general Zilber-Pink conjecture has been applied successfully.

With our proof sketch, not only did we leave the notion of o-minimal definability undefined, but we did not even so much as indicate why the Pila-Wilkie counting theorem might be true. With the following lectures, we shall fill this gap. With the theory of o-minimality and its connection to arithmetic firmly established, we shall return to the more sophisticated applications to problems in diophantine geometry.

3. O-MINIMAL GEOMETRY

O-minimality is a logical condition isolated by van den Dries [27] from which the theory of semi-algebraic geometry may be developed axiomatically, and ultimately, generalized. In order to express the definition of o-minimality we require some terminology from mathematical logic and I would argue that to appreciate the strength of o-minimality one should approach the subject with a sensibility informed by logic. In earlier surveys of this topic, I made a point of suppressing the logical apparatus, but because I wish to explain in some detail how the Pila-Wilkie counting theorem really works, logic will take the foreground in these notes.

First, we need to make sense of definability. There are at least two standard ways to say what we mean by a structure in the sense of first-order logic. One can define a structure to be a nonempty set equipped with system of subsets of its Cartesian powers closed under natural operations. In so doing, it becomes unnecessary to discuss syntax.

Definition 3.1. A structure $\mathfrak{M}$ consists of a nonempty set M given together with a collection $\mathcal{D}_n$ of subsets of M^n for each $n \in \mathbb{Z}_+$. A set $X \in \mathcal{D}_n$ is said to be a *definable subset* of M^n. We require the following conditions.

- Each $\mathcal{D}_n$ is a Boolean algebra under the usual set theoretic operations,
- Each $\mathcal{D}_n$ is invariant under coordinate permutations. That is, if $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is a permutation and $X \in \mathcal{D}_n$, then $\sigma(X) := \{(x_1, \ldots, x_n) \in M^n : (x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)}) \in X\} \in \mathcal{D}_n$
- For each $a \in M$ the singleton set $\{a\}$ is definable.
- The diagonal $\Delta_M := \{(x, y) \in M^2 : x = y\}$ is definable.
- The class of definable sets is closed under products: if $A \in \mathcal{D}_n$ and $B \in \mathcal{D}_m$, then $A \times B$ is definable.
- The class of definable sets is closed under projections: if $A \in \mathcal{D}_{n+1}$ and $\pi : M^{n+1} \to M^n$ is given by $(x_1, \ldots, x_{n+1}) \mapsto (x_1, \ldots, x_n)$, then $\pi(A) \in \mathcal{D}_n$.

Remark 3.2. One might prefer to present a structure by giving a collection $\mathcal{B}_n$ of basic definable sets $X \subseteq M^n$ and then closing the conditions specified in Definition 3.1.

Example 3.3. If M is an algebraically closed field and $\mathcal{D}_n$ is the collection of all constructible subsets of M^n, then $\mathfrak{M}$ is a structure.

Example 3.4. If G is a group and we ask for the set $\Gamma := \{(x,y,z) \in G^3 : xy = z\}$ to be definable, then many other sets will be definable. For example, the center of G, $Z(G) := \{x \in G : (\forall y \in G) xy = yx\}$ Let see that this is in fact the case. By hypothesis $\Gamma \in \mathcal{D}_3$. Thus,

$$A_1 := \{(x_1, x_2, x_3, y_1, y_2, y_3) : x_1 x_2 = x_3 \& y_1 y_2 = y_3\} = \Gamma \times \Gamma \in \mathcal{D}_6$$

Applying the condition about closure under coordinate permutations, we see that

$$A_2 := \{(x_1, x_2, y_1, y_2, x_3, y_3) : x_1 x_2 = x_3 \& y_1 y_2 = y_3\} \in \mathcal{D}_6$$

Since $\mathcal{D}_4$ is a Boolean subalgebra of the power set of G^4, we see that $G^4 \in \mathcal{D}_4$. By our condition on the diagonal, $\Delta \in \mathcal{D}_2$. Hence, since the class of definable sets is closed under products, $G^4 \times \Delta \in \mathcal{D}_6$. By the closure under intersections, we see that

$$A_3 := A_2 \cap (M^4 \times \Delta) = \{(x_1, x_2, y_1, y_2, x_3, y_3) : x_1 x_2 = x_3 \& y_1 y_2 = y_3 \& x_3 = y_3\} \in \mathcal{D}_6$$

is definable. Projecting onto the first five coordinates and then onto the first four coordinates, we see that the following set is definable.

$$A_4 := \{(x_1, x_2, y_1, y_2) : x_1 x_2 = y_1 y_2\}$$

Applying coordinate permutation again, we see that the following set is definable.

$$A_5 := \{(x_1, y_2, x_2, y_1) : x_1 x_2 = y_1 y_2\}$$

Hence, the following set is definable.

$$A_6 := A_5 \cap (\Delta \times \Delta) = \{(x_1, y_2, x_2, y_1) : x_1 x_2 = y_1 y_2 \& x_1 = y_2 \& x_2 = y_1\}$$

Applying another permutation, we have that the following set is definable.

$$A_7 := \{(x_1, x_2, y_1, y_2) : x_1 x_2 = y_1 y_2 \& x_1 = y_2 \& x_2 = y_1\}$$

Projecting onto the first three coordinates and then onto the first two coordinates, we see that the following set is definable.

$$A_8 := \{(x_1, x_2) : x_1 x_2 = x_2 x_1\}$$

Finally, if $\pi : G^2 \to G$ is the projection onto the first coordinate, then we see that $Z(G) = G \smallsetminus \pi(G^2 \smallsetminus A_8)$. Thus, $Z(G) \in \mathcal{D}_1$.

I hope that the calculation in the above example convinces the reader that the presentation of the class of definable sets of Definition 3.1 is not well suited to mathematical arguments. Indeed, in checking that a particular set is definable in a given structure, the most effective method is to simply present the definition. This brings us back to the the definition of a structure as an interpretation of a formal language.

Definition 3.5. A (first-order) language $\mathcal{L}$ is given by the data of a set $\mathcal{C}$ of *constant symbols*, a sequence of sets $\langle \mathcal{F}_n \rangle_{n=1}^{\infty}$ of *function symbols*, and a sequence of sets $\langle \mathcal{R}_n \rangle_{n=1}^{\infty}$ of *relation symbols*. An $\mathcal{L}$-structure $\mathfrak{M}$ is a non-empty set M given together with interpretations of the symbols of $\mathcal{L}$. That is, for each $c \in \mathcal{C}$ are given an element $c^{\mathfrak{M}} \in M$, for each $f \in \mathcal{F}_n$ we are given a function $f^{\mathfrak{M}} : M^n \to M$, and for each $R \in \mathcal{R}_n$ we are given a subset $R^{\mathfrak{M}} \subseteq M^n$.

Remark 3.6. One may recover a structure in the sense of Definition 3.1 from an $\mathcal{L}$-structure in the sense of Definition 3.5 by taking the graph of each $f^{\mathfrak{M}}$ and each set $R^{\mathfrak{M}}$ to be definable and then closing off the class of definable sets under the requirements of Definition 3.1.

Definition 3.7. An *o-minimal structure* is a structure $(R, <, \ldots)$ in the sense of first-order logic for which $<$ is a linear order on R and the ellipses indicate that there may be other distinguished functions and relations for which every definable (with parameters) subset of R is a finite union of points and intervals.

To understand the conclusion of the Pila-Wilkie counting theorem, it suffices to consider only the case where the underlying ordered set is the set of real numbers with its usual ordering. However, the proof of that theorem uses more general o-minimal structures in an essential way.

While in Definition 3.5 we allowed for the formal language to have extra distinguished relations and constants, when studying o-minimal structures, the only extra relation required is the ordering. That is, it suffices to consider only those expansions specified by naming some distinguished functions.

Definition 3.8. Suppose that for each natural number n we are given a set $\mathcal{F}_n$ of functions $f : \mathbb{R}^n \to \mathbb{R}$. By $\mathbb{R}_{\mathcal{F}}$ we mean the structure whose underlying set is $\mathbb{R}$, whose order is the usual order on the real numbers, and which has the distinguished function $f : \mathbb{R}^n \to \mathbb{R}$ for each $f \in \mathcal{F}_n$.

Given such a structure $\mathbb{R}_{\mathcal{F}}$ by a *basic* or *atomic* definable set in $\mathbb{R}^n$ we mean a set of the form

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n \ : \ f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n)\}$$

or

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n \ : \ f(x_1, \ldots, x_n) < g(x_1, \ldots, x_n)\}$$

where f and g are functions of n variables built from the coordinate functions, constant functions and the distinguished functions in $\mathcal{F}$ via appropriate compositions. It follows that the class of definable sets in $\mathbb{R}_{\mathcal{F}}$ is generated from these basic sets by the rules outlined in Definition 3.1.

Perhaps, some examples are in order.

Example 3.9. Taking $\mathcal{F}$ to consist of all polynomials in any number of variables over $\mathbb{R}$, it follows from work of Tarski on the decidability of Euclidean geometry [24] that every definable set is semi-algebraic, that is, a finite Boolean combination of sets defined by by conditions of the form $f(x_1, \ldots, x_n) > 0$ where $f \in \mathbb{R}[x_1, \ldots, x_n]$. Since a polynomial in one variable changes sign only finitely many times, it follows that $\mathbb{R}_{\mathcal{F}}$ is o-minimal.

Example 3.10. Note that $\{x \in \mathbb{R} : \sin(x) = 0\} = \mathbb{Z}\pi$ is an infinite, discrete set and as such cannot be expressed as a finite union of points and intervals. Hence, $\mathbb{R}_{\{\sin\}}$ is not o-minimal.

Example 3.11. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is a *restricted analytic function* if there is a neighborhood $U \supseteq [-1,1]^n$ of the n-cube $[-1,1]^n$ and a real analytic function $\widetilde{f} : U \to \mathbb{R}$ for which $f(x) = \widetilde{f}(x)$ for $x \in [-1,1]^n$ and $f(x) = 0$ for $x \in \mathbb{R}^n \smallsetminus [-1,1]^n$. If we let $\mathcal{F}$ consist of all polynomials over $\mathbb{R}$ and all restricted analytic functions, then van den Dries observed [28] that the o-minimality of $\mathbb{R}_{\mathcal{F}}$ (usually denoted as $\mathbb{R}_{an}$) follows as a consequence of results of Gabrielov [6] on semi-analytic geometry. Thereafter, Denef and van den Dries [3] presented a more direct proof of the o-minimality of $\mathbb{R}_{an}$. The key technical observation required for their proof is that the Weierstrass Preparation and Division Theorems permit one to replace conditions on the sign of an analytic function of a single variable over a closed interval with the same conditions on an associated polynomial.

Example 3.12. Extending work of Khovanski on so-called fewnomials [7], Wilkie [31] showed that if $\mathcal{F}$ contains all the polynomials over $\mathbb{R}$ together with the real exponential function, then $\mathbb{R}_{\exp} := \mathbb{R}_{\mathcal{F}}$ is o-minimal. Wilkie extended this result to obtain the stronger theorem that the expansion of the real field by all functions which satisfy iterated Pfaffian differential equations is o-minimal [32].

Example 3.13. Amalgamating the last two examples so that $\mathcal{F}$ consists of all restricted analytic functions, all polynomials, and the real exponential function we obtain $\mathbb{R}_{an,\exp}$ which van den Dries and Miller proved to be o-minimal [30]. In subsequent work, van den Dries, Macintyre, and Marker analyzed the definable sets in $\mathbb{R}_{an,\exp}$ through the study of generalized power series models [29]. Thereafter, Speissegger showed that if $\mathbb{R}_{\mathcal{F}}$ is an o-minimal structure and $f : \mathbb{R} \to \mathbb{R}$ is a function which satisfies a Pfaffian differential equation over $\mathbb{R}_{\mathcal{F}}$, that is, there is some $G(x,y) \in \mathcal{F}_2$ for which f satisfies the differential equation $Y' = G(x,Y)$, then the structure obtained by adjoining f to $\mathcal{F}_1$ is still o-minimal [23].

The great virtue of the notion of o-minimality is that from the hypothesis about the simplicity of the definable subsets of the line one may deduce strong regularity results about the definable sets in higher dimensions. The fundamental theorem of o-minimality is the cell decomposition theorem which was first proven by van den Dries under the hypothesis that the underlying ordered set is $(\mathbb{R}, <)$ [27] and in full generality by Knight, Pillay and Steinhorn [8, 17, 18].

Definition 3.14. Given an o-minimal structure $(R, <, \ldots)$ we define the class of cells in R^n and their dimension by recursion on n. When $n = 1$, singleton sets $\{a\}$ and intervals (a, b) where we allow the possibility that $a = -\infty$ and that $b = \infty$ are cells. Their dimensions are 0 and 1, respectively. If $X \subseteq R^n$ is a cell and $f : X \to R$ is a continuous (with respect to the order topology), definable function (in the sense that its graph is a definable set), then the graph of f is a cell in R^{n+1} with the same dimension as that of X. If $g : X \to R$ is another continuous, definable function on X for which $f(x) < g(x)$ for every $x \in X$, then the parametrized interval

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

is a cell of dimension one more than that of X. Likewise, infinite intervals $(-\infty, f)_X$ and $(f, \infty)_X$ (with the obvious definitions) are cells also of dimension one more than that of X.

Theorem 3.15. *If $(R, <, \ldots)$ is an o-minimal structure and $X \subseteq R^n$ is definable, then there is a partition of R^n into finitely many cells so that X may be expressed as a union of some of these cells.*

Proof. One argues by induction on n noting that the case of $n = 1$ is exactly the definition of o-minimality and proving along the way two auxiliary results. First, if $f : R \to R$ is any definable function then R may be decomposed into finitely many points and open intervals so that on each such open interval f is strictly monotone or constant. Secondly, every definable function $f : R^n \to R$ is piecewise continuous in the sense that the domain admits a decomposition into finitely many cells for which the restriction of f is continuous. For the inductive argument one shows first that for every definable set X in R^{n+1} there is a cell decomposition of R^{n+1} compatible with X and then establishes the piecewise continuity of definable functions $f : R^{n+1} \to R$.

The key to proving piecewise monotonicity of functions of a single variable is the observation that the sets where f is locally increasing (or decreasing or constant) are definable. Using o-minimality, one shows that if the lemma failed, then there would be an open interval I on which f is never locally constant, locally increasing, or locally decreasing and then derives a contradiction to o-minimality by considering sets of the form $\{x \in I \ : \ f(x) < f(c)\}$ for some fixed $c \in I$.

Piecewise continuity of definable functions $f : R^{n+1} \to R$ is shown by observing that the set of points at which f is continuous is definable and then invoking cell decomposition to conclude that if the result were false there would be an open cell on which f is everywhere discontinuous. One then reaches a contradiction by considering the family of functions $(g_a : R^n \to R)_{a \in R}$ given by $g_a(x_1, \ldots, x_n) := f(x_1, \ldots, x_n, a)$ which we know to be piecewise continuous and the functions $(h^b : R \to R)_{b \in R^n}$ given by $h(x) := f(b, x)$ which we know to be piecewise monotone.

Finally, for $X \subseteq R^{n+1}$ a definable set, we define a sequence of (possibly partial) functions $f_m : R^n \to R$ by sending a to the m^{th} point in the boundary of $X_a := \{y \in R \ : \ (a, y) \in X\}$. Via a nontrivial argument one shows that the cardinality of the boundary of X_a is bounded. By induction, we may decompose R^n into cells on which all of the functions f_i are continuous and the truth value of conditions of the form $f_i(x) \in X_x$ or $y \in X_x$ for some $y \in (f_j(x), f_{j+1}(x))$ is constant. The cell decomposition statement for X follows. □

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

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

Proposition 3.16. *If $\{X_b\}_{b \in B}$ is a definable family of definable sets in some o-minimal structure on the real numbers (with the usual ordering) then there are only finitely many homeomorphism types represented in the family.*

As a corollary of Proposition 3.16 we obtain a theorem of Khovanski [7] on fewnomials. To be fair, while the theorem on fewnomials which we shall discuss is logically a consequence of Proposition 3.16 both temporally and intellectually it is prior. Khovanski's work on fewnomials inspired much of the development of theory of o-minimality and many of his specific results underly Wilkie's proof of the o-minimality of $\mathbb{R}_{\exp}$. Moreover, the argument we outline below is patterned on Khovanski's own proof through the passage from polynomials of indeterminate degree to exponential polynomials.

Theorem 3.17. *For fixed integers k and n there are only finitely many homeomorphism types amongst the following sets*

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

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

To prove Theorem 3.17 we observe that it suffices show that there are only finitely many homeomorphism types even if we allow m to range through $M_{k \times n}(\mathbb{R})$ rather than merely $M_{k \times n}(\mathbb{N})$. The above family of semi-algebraic sets may be embedded into the following $\mathbb{R}_{\exp}$-definable family.

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

The finiteness of the number of homeomorphism types is now a special case of Proposition 3.16.

Another crucial property of o-minimal structures is the existence of definable choice functions.

Theorem 3.18. *Let $(R, <, +, 0, 1, \cdots)$ be an o-minimal expansion of an ordered group (where $1 > 0$) and $\{X_b\}_{b \in B}$ a definable family of nonempty definable subsets of R^m for some m. Then there is a definable function $f : B \to R^m$ so that $f(b) \in X_b$ for each $b \in B$.*

Proof. We work by induction on m. If $m = 1$, then define

$$f(b) := \begin{cases} 0 & \text{if } X_b = R \\ a & \text{if } a = \min X_b \\ a - 1 & \text{if } a \text{ is the least boundary point and } (-\infty, a) \subseteq X_b \\ a + 1 & \text{if } a = \inf X_b \text{ and } (a, \infty) = X_b \\ \frac{a+c}{2} & \text{if } a \notin X_b,\ a = \inf X_b, \\ & \text{and } c \text{ is the least boundary point of } X_b \text{ greater than } a \end{cases}$$

Let $\pi : R^{m+1} \to R^m$ be the projection to the first m coordinates. That by our inductive hypothesis, there is a choice function $g : B \to R^m$ for the family of definable sets $\{\pi(X_b)\}_{b \in B}$. For $b \in B$ define $Y_b := \{x \in R : (g(b), x) \in X_b\}$. Then $\{Y_b\}_{b \in B}$ is a definable family of nonempty subsets of R. Hence there is a choice function $h : B \to R$ for this family. Our desired f is given by $b \mapsto (g(b), h(b)$. □

4. Pila-Wilkie counting theorem

The proof of the Pila-Wilkie counting theorem consists of two major parts. First, a general theorem about parametrization of definable sets in o-minimal structures is proven. With this result, it is shown that every bounded set may be expressed as the image of a finite number of smooth functions on unit boxes having small derivatives. The crucial part of this theorem is that the finite number and the bounds on the derivatives may be obtained uniformly in families. Once this result is established, standard methods from the theory of diophantine approximation may be used to bound the number of rational points in definable sets.

The Pila-Wilkie counting theorem only really makes sense for o-minimal structures on the real numbers. As such, it may seem that we should (or, at least, could) restrict attention to such structures in its proof. However, for our applications we require a uniform version of the parametrization theorem which will follow from a case by case version proven in an arbitrary o-minimal structure. This alone might not be enough to justify the move to general o-minimal structures as one might imagine that proving the uniform version over the real numbers might be simply a matter of keeping track of variation in families. However, as the reader will see soon enough, our inductive argument will be complicated enough, but if we were to try to implement it uniformly, it would degenerate into an incomprehensible mess. Of course, there may be a natural way to retool the argument so as to avoid nonstandard models, but I believe that the conceptual simplification afforded by this move amply rewards the decision to work model theoretically.

Notation 4.1. Throughout this section, M is an o-minimal structure expanding an ordered field.

Definition 4.2. Let $n \in \mathbb{Z}_+$ be a positive integer. We say that a set $X \subseteq M^n$ 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 4.3. If the underlying ordered field of M is simply the field of real numbers, then there is no distinction between bounded and strongly bounded. The difference becomes apparent only for nonstandard models.

Notation 4.4. Let $\ell \in \mathbb{Z}_+$ be a positive integer and $\phi : (0,1)^\ell \to M$ a definable function. For $\alpha \in \mathbb{N}^\ell$ an multi-index, we denote by $\phi^{(\alpha)}$ the function $\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_\ell^{\alpha_\ell}} \phi$. Note that while $\phi^{(\alpha)}$ might not be defined at every point in $(0,1)^\ell$, by the cell decomposition theorem it is defined on a dense open definable subset. More generally, if $n \in \mathbb{Z}_+$ is any positive integer and $\psi : (0,1)^\ell \to M^n$ is a definable function expressed as $\psi(x_1, \ldots, x_\ell) = (\phi_1(x_1, \ldots, x_\ell), \ldots, \phi_n(x_1, \ldots, x_\ell))$, then we write $\phi^{(\alpha)}$ for $(\phi_1^{(\alpha)}, \ldots, \phi_n^{(\alpha)})$.

Definition 4.5. Let $X \subseteq M^n$ be a definable set with $\dim(X) = \ell$. A *partial parametrization* of X is a definable function $\phi : (0,1)^\ell \to X$. A *parametrization* of X is a finite set S of partial parameterizations for which $X = \bigcup_{\phi \in S} \phi((0,1)^\ell)$. 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 then a parametrization of X by partial k-parameterizations.

Remark 4.6. In the definition of a k-parametrization, at the cost of increasing the finite number of functions in S we could strengthen the conclusion from $\phi^{(\alpha)}$ being strongly bounded to $|\phi^{(\alpha)}| \leq 1$. Indeed, suppose that $N \in \mathbb{Z}_+$ were an upper bound for $|\phi^{(\alpha)}|$. For $a = (a_1, \ldots, a_\ell) \in \{0, 1, \ldots, N-1\}^\ell$ define $\phi_a(x) := \phi(\frac{x_1 + a_1}{N}, \ldots, \frac{x_\ell + a_\ell}{N})$. A simple calculation using the chain rule shows that $|\phi_a^{(\alpha)}| \leq 1$ and clearly the image of ϕ is the union over the N^ℓ such choices of a of the the images of ϕ_a.

Definition 4.7. We call a k-parametrization S of X a *strong k-parametrization* if for each $\phi \in S$ and each multi-index α with $|\alpha| \leq k$ we have $|\phi^{(\alpha)}(x)| \leq 1$.

The main structural theorem of this section is the following parametrization theorem.

Theorem 4.8. *For any strongly bounded definable set $X \subseteq M^n$ and any positive integer $k \in \mathbb{Z}_+$ there is a k-parametrization of X.*

Remark 4.9. By Remark 4.6 we may strengthen Theorem 4.8 to the conclusion that X admits a strong k-parametrization. It will be convenient throughout the course of our inductive argument to allow for the relaxed notion of a mere k-parametrization.

I regard Theorem 4.8 as a dual version of the cell decomposition theorem. With the cell decomposition theorem we take an arbitrary definable set X in M^n and cut up the ambient space into finitely many simple definable pieces, namely the cells, so that X and its complement may be decomposed into these parts. By bootstrapping the cell decomposition procedure we may ask that the boundaries of the cells be given by the graphs of arbitrarily smooth functions. With Theorem 4.8 instead of breaking X into simple pieces, each of which is definably homeomorphic to some open box, we cover X by the images of such boxes under functions with small derivatives. As we shall see, not only are the statements of these theorems analogous, but their proofs share a common structure (though to be honest there is a substantial difference in the organization of the induction to improve the order of differentiability which we shall address in due course). 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. In the case of the cell decomposition theorem, this regularity theorem takes the form of the theorems that in dimension one every unary definable function is piecewise monotone and in general that every definable function is piecewise continuous where the pieces in question are cells. For Theorem 4.8, the regularity theorem concerns *reparameterizations* of definable functions.

Definition 4.10. Let $X \subseteq M^m$ be a definable set, $\Phi : X \to M^n$ 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 4.11. 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 4.12. *For every positive integer k every strongly bounded definable function admits a k-reparametrization.*

Using the compactness theorem one shows that a uniform version of Theorem 4.8 follows as a corollary. Let us recall the notions of definable families of sets and of functions.

Definition 4.13. By a definable family $\{X_b\}_{b \in B}$ of definable subsets of M^n we mean a definable set $X \subseteq M^n \times B \subseteq M^n \times M^m$ (where $B \subseteq M^m$ is itself a definable set) where for $b \in B$ the set X_b is the fibre $X_b := \{x \in M^n : \langle x, b \rangle \in X\}$. By a definable family of definable functions we mean simply a definable family of definable sets each of which is a definable function, or if your prefer, the graph of a definable function.

Theorem 4.14. *For any triple of positive integers $m, n, k \in \mathbb{Z}_+$ and definable family $\{X_b\}_{b \in B}$ of definable subsets of $(0,1)^m$ there exists a number $N \in \mathbb{Z}_+$ and N families of definable functions $\{\phi_{i,b}\}_{b \in B}$ for $i \leq N$ so that for each $b \in B$ the set $\{\phi_{i,b} : i \leq m\}$ is a strong k-parametrization of X_b.*

Proof. To deduce Theorem 4.14 from Theorem 4.8 we use a standard application of the compactness theorem of first-order logic.

We first prove an apparently weaker result. Given a definable family $\{X_b\}_{b \in B}$ of definable subsets of $(0,1)^m$ there is some $N \in \mathbb{Z}_+$ and N families $\{\phi_{i,c}\}_{c \in C}$ of definable functions so that for any $b \in B$ there is some $I \subseteq \{1, \ldots, N\}$ and some $c \in C$ so that $\{\phi_{i,c}\}_{i \in I}$ is a strong k-parametrization of X_b.

The full uniform theorem follows from this ostensibly weaker result by the existence of definable choice functions, Theorem 3.18.

If the uniform theorem were false, there would be some definable family of definable subsets $\{X_b\}_{b \in B}$ of $(0,1)^m$ so that for any finite sequence of definable families of definable functions $\{\phi_{i,c}\}_{c \in C}$ for $1 \leq i \leq N$, there would be some $b \in B$ for which no collection $\{\phi_{i,c} : i \in I\}$ is a strong k-parametrization of X_b. Applying the compactness theorem to the theory of R together with the set of sentences (using the new constant symbol b) asserting that for each such family of definable functions there is no parameter c for which $\{\phi_{i,c} : 1 \leq i \leq n\}$ is a strong k-parametrization of X_b, we find an o-minimal structure M^* and a point $b \in B(M^*)$ for which the strongly bounded definable set X_b does not admit a strong k-parametrization, but this contradicts Theorem 4.8. □

As we indicated above, the proof of Theorem 4.8 is organized as an induction with which we interweave a proof of Theorem 4.12 starting with a particularly strong form of Theorem 4.12 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 4.12 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 thereby improving 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 core of the inductive argument by showing $(\forall k)P(n,k) \to (\forall k)R(n+1,1,k)$ and $(\forall k,n)R(m,n,k) \to (\forall k)P(n+1,k)$.

With Theorem 4.14 established, we deduce the counting theorem. We begin with a proposition about algebraic relations on the rational points in the image of a function with small derivatives.

Proposition 4.15. *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 C^k function with image X satisfying $|\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_m^{\alpha_m}} \phi_i(x)| \leq 1$ for all multi-indices α with $|\alpha| \leq k$, then for any $t \geq 1$ the set $X(t)$ 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 $d \to \infty$.*

Proof. We begin by changing the problem slightly observing that it would be enough to show that the set of rational points with a fixed denominator t lie in $O(t^\epsilon)$ hypersurfaces of degree d.

Take $Q_1, \ldots, Q_\ell \in (0,1)^m$ so that the points $\phi(Q_i)$ are the distinct rational points in X with denominator t. Let $D := \binom{n+d}{d}$. Consider the $\ell \times D$-matrix $(\phi(Q_i)^\mu)$ where μ ranges through the multi-indices in $\mathbb{N}^\ell$ with $|\mu| \leq d$. Using a Taylor series expansion and our presumed bound on the size of the derivatives of ϕ, we see that there is a constant $b = b(m,n,d) \in \mathbb{Z}_+$ and $C' > 0$ so that for any $0 < r < 1$ if $z_1, \ldots, z_D$ are points in $(0,1)^m$ all within a given ball of radius r, then $|\det(\phi(z_i))^\mu| \leq C' r^b$. On the other hand, if the z_i's are chosen from the Q_j's then $\det(\phi(z_i)^\mu) \in \frac{1}{t^D}\mathbb{Z}$. Thus, if they are all taken from the same small enough ball, this determinant must be zero. It follows from the implied linear dependencies on the columns of the matrix $(\phi(z_i)^\mu)$ that when the points Q_j are all taken from the same small ball, then their images lie on a common hypersurface of degree d. Covering $(0,1)^m$ with small enough balls, we see that the set of rational points on X with fixed denominator t is contained in $(C')^{\frac{n}{b}} t^{Dn} b$ hypersurfaces of degree d. To complete the argument one needs to be somewhat more careful with the calculation of the combinatorial quantities D and b than I have been, but upon doing so one sees that the exponent $\frac{Dn}{b}$ tends to zero as $d \to \infty$. □

Combining Proposition 4.15 with Theorem 4.14 we deduce a qualitative version of the counting theorem.

Proposition 4.16. *Let $\{X_b\}_{b \in B}$ be a definable family of subsets of $(0,1)^n$ and $\epsilon > 0$. Then there is a constant $C > 0$ and a number $d \in \mathbb{Z}_+$ depending only on* *the family and ϵ so that for any $b \in B$ and $t \geq 1$, the set $X_b(t)$ is contained in no more than Ct^ϵ hypersurfaces of degree d.*

Proof. Let k and d be large enough so that the quantity $\epsilon(m, n, d)$ of Proposition 4.15 is less than our prescribed ϵ. By Theorem 4.14, there is a number M so that every X_b may be covered by the image of at most M strong k-parametrizations. The constant C of this proposition is M times the constant of Proposition 4.15. □

The counting theorem itself now follows from Proposition 4.16 by induction on the fibre dimension of X_b.

5. Applications

In this section we sketch some of the more sophisticated diophantine theorems proved using the Pila-Wilkie counting theorem.

We begin with the first nontrivial instance of the Zilber-Pink conjecture in which one considers intersections with positive dimensional special varieties rather than merely with special points: a theorem of Masser and Zannier [11] about simultaneous torsion on elliptic curves. They consider the family of elliptic curves presented in their affine Legendre form where E_λ is defined by the affine planar equation $y^2 = x(x-1)(x-\lambda)$ for $\lambda \in \mathbb{C} \smallsetminus \{0, 1\}$. From theory of elliptic curves, E_λ considered together with the point at infinity has a unique structure of an algebraic group with that point at infinity as the identity. For a fixed complex number a we might consider the set of λ for which the point $(a, \sqrt{a(a-1)(a-\lambda)})$ is torsion in the group $E_\lambda(\mathbb{C})$. It is not hard to see that for $a = 0$ or $a = 1$, then these points are always torsion. On the other hand, for every other a there are only countably many λ for which this point is torsion in $E_\lambda(\mathbb{C})$. Nevertheless, computing the rational functions which define the multiplication by n map on E_λ it is fairly easy to show that for any such a there will be infinitely many λ for which $(a, \sqrt{a(a-1)(a-\lambda)})$ is torsion. Masser and Zannier address the question: if we consider two number a and b, for how many λ are $(a, \sqrt{a(a-1)(a-\lambda)})$ and $(b, \sqrt{b(b-1)(b-\lambda)})$ both torsion in $E_\lambda(\mathbb{C})$? In the special case of $a = 2$ and $b = 3$ they given an answer.

Theorem 5.1. *There are only finitely many complex numbers λ for which*

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

and

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

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

Remark 5.2. The proof of Theorem 5.1 applies perfectly well to any two numbers a and b for which the points

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

and

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

are linearly independent over $\mathbb{Z}$ in the group $E_\lambda(\mathbb{Q}(\lambda))$. Indeed, in [10] the same authors prove such a result, even allowing for the sections to be nonconstant.

The proof of Theorem 5.1 follows the pattern of the proof of Theorem 2.3 we have outlined above. For each elliptic curve E_λ, the the theory of analytic uniformizations gives a complex analytic covering map $\pi_\lambda : \mathbb{C} \to E_\lambda(\mathbb{C})$. As with the usual exponential function, this covering is not definable in any o-minimal expansion of the real numbers. However, if we restrict π_λ to a fundamental domain, it is. Moreover, at the cost of treating π_λ as simply a real analytic function, we may normalize the fundamental domain so that the domain of π_λ is the square $[0,1) \times [0,1)$ and the map π_λ is a group homomorphism when $[0,1)$ is given the usual wrap around additive structure. With some work, one can show that the two variable (or, really, four real variable) function $(\lambda, z) \mapsto \pi_\lambda(z)$ is definable in $\mathbb{R}_{an,\exp}$ relative to the usual interpretation of $\mathbb{C}$ in $\mathbb{R}$ and when z is restricted to $[0,1) \times [0,1)$. Masser and Zannier then study the set

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

Visibly, $\widetilde{X}$ is definable in $\mathbb{R}_{an,\exp}$ and it is not hard to see that the rational points on $\widetilde{X}$ all come from λ for which P_λ and Q_λ are simultaneously torsion. Transcendence results about the Weierstraß $\wp$-function are used in place of Ax's theorem to show that $\widetilde{X}^{\mathrm{alg}}$ is empty and a theorem of David [2] about the degree of the field extension required to define elliptic curves with elements of specified order plays the rôle of the calculation of the degree of a cylcotomic extensions.

It bears noting that the published sketch of Theorem 5.1 avoids an explicit reference to definability in o-minimal structures as Pila had proved a provisional version of Theorem 2.9 for subanalytic surfaces without invoking the theory of o-minimality [13]. On the other hand, due to the work of Peterzil and Starchenko [12] on the uniform definability of theta functions in $\mathbb{R}_{an,\exp}$, it follows that the question of simultaneous torsion in families of higher dimensional abelian varieties may be analyzed via these methods.

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

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

There is an analytic function $j : \mathfrak{h} \to \mathbb{C}$ having the property that $E_\tau(\mathbb{C})$ and $E_\sigma(\mathbb{C})$ are isomorphic as elliptic curves if and only if $j(\tau) = j(\sigma)$. We refer to the value $j(\tau)$ as the *j-invariant* of the elliptic curve E_τ. Let us say that a complex number ζ is a *special point* if it is the j-invariant of an elliptic curve with complex multiplication. By the above discussion, we see that a number is special if and only if it is the value of the analytic j-function on a quadratic imaginary number. The André-Oort conjecture in this case predicts the form of the algebraic subvarieties $X \subseteq \mathbb{A}^n_{\mathbb{C}}$ of affine n-space which contain a Zariski dense set of n-tuples of special points. Specializing to the case of $n = 2$, it proposes a solution to the question of for which polynomials $g(x, y) \in \mathbb{C}[x, y]$ are there infinitely many pairs (ξ, ζ) of special points for which $g(\xi, \zeta) = 0$? This case was solved early in the investigations around the André-Oort conjecture, first assuming the Riemann Hypothesis by Edixhoven [4] and then unconditionally by André [1].

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

Theorem 5.3 (Pila). *Let $X \subseteq \mathbb{A}^n_{\mathbb{C}}$ be an irreducible algebraic subvariety of affine n-space over the complex numbers. Suppose that the set*

$$\{(\xi_1, \ldots, \xi_n) \in X(\mathbb{C}) \ : \ \textit{each } \xi_i \textit{ is the j-invariant of a CM-elliptic curve}\}$$

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

The proof of Theorem 5.3 follows our by now familiar pattern. First, Pila observes that j restricted to a fundamental domain is definable in $\mathbb{R}_{an,\exp}$. Indeed, again relative to the usual interpretation of $\mathbb{C}$ as $\mathbb{R}^2$, the complex exponential function is definable in $\mathbb{R}_{\exp}$ when restricted to a fundamental domain for j and maps the fundamental domain into a disk properly contained in the unit disk. It is well known that the q expansion of the j-function is meromorphic on the unit disk with a simple pole at the origin. Hence, the restriction of this function to any proper subdisk is definable in $\mathbb{R}_{an}$. Thus, the restriction of the j-function to a fundamental domain is definable in $\mathbb{R}_{an,\exp}$. One then moves from a study of X to that of $\widetilde{X}$, the inverse image of $X(\mathbb{C})$ via j (or really, the map $(z_1, \ldots, z_n) \mapsto (j(z_1), \ldots, j(z_n))$) restricted to its fundamental domain, which is a definable set in $\mathbb{R}_{an,\exp}$. One must determine $\widetilde{X}^{\mathrm{alg}}$. In this case, one considers of the action of the modular group showing that the algebraic part comes from the pre-images of finitely many varieties of the desired form. At this point, the goal is to show that if X does not already have the desired form, then there are only finitely many quadratic imaginary points in $\widetilde{X}^{\mathrm{tr}}$. The counting theorem, Theorem 2.9, applied to rational points, but Pila proves similar bounds for algebraic points of bounded degree [14]. Thus, for any $\epsilon > 0$ there is some constant C for which the number of quadratic imaginary points of height at most t in $\widetilde{X}^{\mathrm{tr}}$ is at most Ct^{ϵ}. As in the proof of Theorem 2.3, he reduces to the case that X is defined over a number field and observes that if there are special points coming from $\widetilde{X}^{\mathrm{tr}}$, then all of their Galois conjugates are also in this set. At this point, he estimates the size of these orbits from below using Siegel's theorem on the growth of the class number [22] to find that for $\epsilon < 2$ one has a lower bound of Ct^{ϵ} thus contradicting the upper bound from the counting theorem.

Remark 5.4. Theorem 5.3 had been proven previously by Edixhoven and Yafaev [5] under the assumption of the Generalized Riemann Hypothesis for quadratic imaginary fields. Their proof shares the same kind strategy at the end: find upper bounds geometrically and lower bounds via Galois theory and analytic number theory.

Remark 5.5. The paper in which the proof of Theorem 5.3 appears [15] includes proofs of theorems in the direction of the Pink-Zilber conjectures. On the other hand, while many parts of this argument succeed when applied to other Shimura varieties, some steps are incomplete. For example, it is known that the analytic covering maps for the moduli spaces of principally polarized abelian varieties are definable in $\mathbb{R}_{an,\exp}$ (again, after suitable restriction) [12] and it seems plausible that the arguments employed to determine the algebraic parts of inverse images of algebraic varieties by Cartesian powers of j should work for these maps, too, but to date no one has carried out the details. Recent work of Tsimerman [25] and Ullmo and Yafaev [26] has established exponential lower bounds for the size of Galois orbits of special points on some moduli spaces of abelian varieties. It is not unreasonable to hope that these methods might lead an unconditional proof of the André-Oort conjecture.

References

[1] Yves André. Finitude des couples d'invariants modulaires singuliers sur une courbe algébrique plane non modulaire. *J. Reine Angew. Math.*, 505:203–208, 1998.

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

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

[4] Bas Edixhoven. Special points on the product of two modular curves. *Compositio Math.*, 114(3):315–328, 1998.

[5] Bas Edixhoven and Andrei Yafaev. Subvarieties of Shimura varieties. *Ann. of Math. (2)*, 157(2):621–645, 2003.

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

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

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

[9] Henry B. Mann. On linear relations between roots of unity. *Mathematika*, 12:107–117, 1965.

[10] D. Masser and U. Zannier. Torsion points on families of squares of elliptic curves. *Mathematische Annalen*, 16 February 2011. Online First.

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

[12] Ya'acov Peterzil and Sergei Starchenko. Definability of restricted theta functions and families of abelian varieties. *preprint*, 2010.

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

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

[15] Jonathan Pila. O-minimality and the André-Oort conjecture for $\mathbb{C}^n$. *Ann. of Math. (2)*, (to appear).

[16] 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.

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

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

[19] 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.

[20] Thomas Scanlon. The André-Oort conjecture via o-minimality (after J. Pila). 2011. Séminaire Bourbaki. Vol. 2010/2011.

[21] Thomas Scanlon. Counting special points: logic, diophantine geometry and transcendence theory. 8 January 2011. Current Events Bulletin.

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

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

[24] Alfred Tarski. *A Decision Method for Elementary Algebra and Geometry*. RAND Corporation, Santa Monica, Calif., 1948.

[25] J. Tsimermann. Brauer-siegel for arithmetic tori and lower bounds for galois orbits of special points. preprint, 2011.

[26] E. Ullmo and A. Yafev. Nombre de classes des tores de multiplication complexe et bornes inférieures pour orbites Galoisiennes de points spéciaux. preprint, 2011.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] A. J. Wilkie. A theorem of the complement and some new o-minimal structures. *Selecta Math. (N.S.)*, 5(4):397–421, 1999.

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

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