WILKIE’S CONJECTURE FOR PFAFFIAN STRUCTURES

GAL BINYAMINI, DMITRY NOVIKOV, AND BENNY ZACK

Abstract. We prove an effective form of Wilkie’s conjecture in the structure generated by restricted sub-Pfaffian functions: the number of rational points of height $H$ lying in the transcendental part of such a set grows no faster than some power of $\log H$. Our bounds depend only on the Pfaffian complexity of the sets involved. As a corollary we deduce Wilkie’s original conjecture for $\mathbb{R}^{\exp}$ in full generality.

1. Introduction

1.1. Main results. Let $(S, \Omega)$ be a $\aleph_0$-minimal structure admitting sharp cell decomposition and sharp derivatives (for the definition of these notions see §2). The structure $\mathbb{R}_{\text{Pfaff}}$ of restricted sub-Pfaffian sets (see §3) is an example of this setup, and the unfamiliar reader may keep this example in mind in place of the general setting.

If $X \subset \mathbb{R}^n$ then following [15] we denote by $X^{\text{alg}}$ the union of all connected, positive-dimensional semialgebraic sets contains in $X$, and denote $X^{\text{trans}} := X \setminus X^{\text{alg}}$. For $g, H \in \mathbb{N}$ we denote

$$X(g, H) := \{x \in X : [Q(x) : \mathbb{Q}] \leq g, H(x) \leq H\}, \quad X_Q(H) := X(1, H)$$

where $H(\cdot)$ denotes the multiplicative Weil height on $\overline{\mathbb{Q}}$, extended to $\overline{\mathbb{Q}}^n$ as the maximum of the heights of the coordinates. If unfamiliar see §1.4 for the asymptotic notation used below.

Theorem 1. Let $X \in \Omega_{F,D}$. Then

$$\#X^{\text{trans}}(g, H) \leq \text{poly}_F(D, g, \log H).$$

This establishes, in the restricted Pfaffian setting, a conjecture by Pila [18, Conjecture 1.5]. As an immediate corollary we obtain the following.

Corollary 1 (Wilkie’s conjecture). Let $X$ be definable in $\mathbb{R}^{\exp}$. Then

$$\#X^{\text{trans}}(g, H) \leq \text{poly}_X(g, \log H).$$

Proof. By Wilkie’s theorem of the complement [24] we have $X = \pi_n(Y)$ where $Y \subset \mathbb{R}^N$ is quantifier-free in $\mathbb{R}^{\exp}$, and $\pi_n : \mathbb{R}^N \to \mathbb{R}^n$ is the projection to the first

2010 Mathematics Subject Classification. Primary 03C64,11U09; Secondary 11G50, 14P10.

Key words and phrases. Yomdin-Gromov lemma, point-counting, Pila-Wilkie theorem, $o$-minimality.

This research was supported by the ISRAEL SCIENCE FOUNDATION (grant No. 1167/17) and by funding received from the MINERVA Stiftung with the funds from the BMBF of the Federal Republic of Germany. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 802107).
n coordinates. Let \( g, H \in \mathbb{N} \) and choose \( M \gg 1 \) such that \( X(g, H) = X_M(g, H) \)
where
\[
X_M := \pi_n(Y_M), \quad Y_M := Y \cap [-M, M]^N.
\]
Now \( Y_M \) is restricted semi-Pfaffian, as it is defined by Pfaffian functions (exponential polynomials) restricted to \([-M, M]^N\). Crucially, \( Y_M \in \Omega_{\mathcal{T}, D} \) where \( \mathcal{T}, D \) depend on \( Y \) but not on \( M \). Then the same is true for \( X_M \), and we conclude
\[
\#X^{\text{trans}}(g, H) \leq \#X_M^{\text{trans}}(g, H) = \text{poly}_X(g, \log H) \tag{5}
\]
by Theorem 1.

We also have a “blocks” generalization of Theorem 1. Recall from [19] that a definable set \( B \subset \mathbb{R}^n \) is called a basic block if it is connected and regular, and contained in a connected regular semialgebraic set of the same dimension (which we call a semialgebraic closure of \( B \), though not this is not uniquely defined). We denote by \( \Omega^{\text{alg}} \) a filtration making \( (\mathbb{R}_{\text{alg}}, \Omega^{\text{alg}}) \) into a \( \# \)-o-minimal structure (one can take, e.g., the filtration from [21] for the empty Pfaffian chain).

**Theorem 2.** Let \( X \subset \mathbb{R}^n \) with \( X \in \Omega_{\mathcal{T}, D} \). Then there exists a collection \( \{B_\eta \subset X\} \) of basic blocks with semialgebraic closures \( S_\eta \) such that \( X(g, H) \subset \cup_\eta B_\eta \) and
\[
\#\{B_\eta\} = \text{poly}_\mathcal{T}(D, g, \log H), \quad \forall \eta : S_\eta \in \Omega^{\text{alg}}_{\mathcal{O}_{\mathcal{T}, (1)}, \text{poly}_X(g, \log H)}. \tag{6}
\]

Theorem 2 clearly implies Theorem 1 since the positive-dimensional basic blocks \( B_\eta \) are subsets of \( X^{\text{alg}} \) by definition.

1.2. A \( C^r \)-parametrization lemma. For a \( C^r \)-smooth function \( f : U \rightarrow \mathbb{R} \) on a domain \( U \subset \mathbb{R}^m \) we denote
\[
\|f\| := \sup_{x \in U} |f(x)|, \quad \|f\|_r := \max_{|\alpha| \leq r} \|D^\alpha f\|.
\]
For \( F : U \rightarrow \mathbb{R}^n \) we set \( \|F\| = \max_i \|F_i\| \) and similarly for \( \|F\|_r \). For a set \( A \subset \mathbb{R}^n \) we write \( U_\varepsilon(A) \) for the \( \varepsilon \)-neighborhood of \( A \) with the \( \ell_\infty \)-metric. For \( A, B \subset \mathbb{R}^n \), we write \( A \subseteq_\varepsilon B \) to mean that \( A \subset B \) and \( B \subset U_\varepsilon(A) \). We say that \( A \) is an \( \varepsilon \)-cover of \( B \).

The main novelty of our approach is the following version of Yomdin’s algebraic lemma. Let \( I := (0, 1) \).

**Lemma 2.** Let \( r \in \mathbb{N} \) and \( \varepsilon > 0 \). Let \( X \subset [0, 1]^n \) be of dimension \( \mu \) with \( X \in \Omega_{\mathcal{T}, D} \). Then there exists a collection \( \{\phi_\eta : I^\mu \rightarrow X\} \) such that \( \|\phi_\eta\|_r \leq 1 \) and \( \cup_\eta \text{Im} \phi_\eta \subseteq X \), and
\[
\#\{\phi_\eta\} \leq \text{poly}_\mathcal{T}(D, r, \log \varepsilon), \quad \forall \eta : \phi_\eta \in \Omega_{\mathcal{T}, (1), \text{poly}_X(D, r)}. \tag{8}
\]

This formulation is similar in spirit to Yomdin’s original formulation [27, 28]. Gromov [11] later refined Yomdin’s work by showing that we can avoid \( \varepsilon \)-covers altogether and cover the set \( X \) completely. However, as we will see, Lemma 2 is sufficient for the applications in Diophantine geometry (as it was for Yomdin’s original application in dynamics). The weaker formulation with \( \varepsilon \)-covers enables us (as was already the case in Yomdin’s original work) to restrict to affine reparametrizations at some crucial moments, where Gromov’s approach involves nonlinear terms. The linearity of the reparametrizing maps turns out to allow for crucial technical simplifications related to achieving polynomial growth of \( \#\{\phi_\eta\} \) as a function of \( r \) (specifically in §5.3).
1.3. Background.

1.3.1. The Pila-Wilkie theorem. The origin of the area of point-counting in tame geometry can be traced to the work of Bombieri and Pila [8, 14]. In these papers it was shown that if $\Gamma \subset \mathbb{R}^2$ is a compact analytic curve containing no semialgebraic curves then for every $\varepsilon > 0$ one has $\# \Gamma(\mathbb{Q}, H) = O_{\Gamma, \varepsilon}(H^\varepsilon)$. After some work by Pila on subanalytic surfaces [16, 17], this result was generalized into its canonical form by Pila and Wilkie [15], who proved that the bound $\# \mathcal{X}^{\text{trans}}(\mathbb{Q}, H) < C(\mathcal{X}, \varepsilon) \cdot H^\varepsilon$ holds for any $\mathcal{X}$ definable in an o-minimal structure. This result has had a profound impact on arithmetic geometry, and we refer the reader to [21] for a survey.

For Pfaffian surfaces, Jones and Thomas [13] established an effective form of the Pila-Wilkie theorem. In the general restricted sub-Pfaffian setting, a recent paper by the first author with Jones, Schmidt and Thomas [1] establishes an effective form of the Pila-Wilkie theorem: if $\mathcal{X} \in \Omega^{\text{Pfaff}},D$ then one can take $C(\mathcal{X}, \varepsilon) = \text{poly}^{\mathcal{X}}(D)$.

Many of the technical methods for using $\#\text{o-minimality}$ in our context are inspired by this prior work. Indeed the paper [7] which inspired the notion of $\#\text{o-minimality}$ grew out of an attempt to provide a suitable foundation for the results in [1].

1.3.2. The Wilkie conjecture. Examples by Pila [16, Example 7.5] show that in $\mathbb{R}_{\text{an}}$ the Pila-Wilkie asymptotic is essentially optimal. However, such examples involve “hand-crafted” functions and no “natural” example exhibiting this behavior is known. Wilkie made his conjecture (now Corollary 1) in the original paper [15] as a concrete formulation of this phenomenon. The case of Pfaffian curves was proved by Pila [18], and our approach in the one-dimensional case is indeed somewhat similar to Pila’s approach. Some further examples of surfaces were treated in [20], but general surfaces already seem difficult to treat with this approach.

The key obstacle to progress on Wilkie’s conjecture has been to establish a $C^r$-parametrization lemma with polynomial bounds for the number of charts, as a function of the complexity of the set and the smoothness order $r$. This problem was open even in the semialgebraic case, and was recently resolved in [3] using complex analytic methods (see also [9, 22, 23] for a result on polynomial growth in $r$, without complexity bounds in some o-minimal structures). The problem remains open beyond the semialgebraic case, and the complex analytic methods seem unlikely to directly carry over to the unrestricted exponential case. For general definable sets, the only previously known case of the Wilkie conjecture is [2] by the first two authors. This paper established Wilkie’s conjecture for the structure $\mathbb{R}^{\text{REF}}$ generated by the exponential and sine functions restricted to compact domains. The proofs were based on an approach avoiding smooth parametrizations altogether, replacing it by complex-geometric ideas. This is only applicable for holomorphic-Pfaffian functions, i.e. holomorphic functions whose graph, viewed as a real set, is Pfaffian in the real sense. By comparison, our approach here applies to arbitrary restricted sub-Pfaffian functions without requiring that the complex-analytic continuation is again Pfaffian. The complex-geometric ideas also seem much more difficult to carry out in the presence of unrestricted exponentials.

Remark 3. The proof of Corollary[1] would not be applicable with the methods of [3] because the complex analytic nature of these methods would require us to consider $e^{z_1}, \ldots, e^{z_M}$ restricted to large complex polydiscs $D(0, M)^N$ rather than large cubes $[-M, M]^N$ as we do here. However, while the Pfaffian complexity of $e^z$ on $[-M, M]$ is bounded independently of $M$, the Pfaffian complexity of $e^z$ on $D(0, M)$ is roughly
as evidenced by the fact that the Pfaffian equation $e^z = 1$ admits roughly $M/\pi$ solutions in $D(0,M)$.

1.3.3. Unrestricted exponentials in arithmetic applications. Unrestricted exponentials are used in many of the most spectacular applications of the Pila-Wilkie theorem, where they arise in uniformizing maps of arithmetic quotients around cusps. Extending the more advanced counting techniques to this case is therefore potentially very useful. In particular, a recent paper by the first author \cite{2} establishes a polylogarithmic counting result in the spirit of Wilkie’s conjecture for sets defined using algebraic foliations (not necessarily Pfaffian) over number fields. This result has played an important role in recent progress on the André-Oort conjecture for general Shimura varieties. It was used by the first author, Schmidt and Yafaev \cite{6} to establish Galois lower bounds for special points conditional on certain height bounds. These height bounds were subsequently proved by Pila, Shankar and Tsimerman thus finishing the proof of André-Oort in general.

The approach of \cite{2} is based on the complex geometric ideas of \cite{3} and suffers from the same limitation to restricted analytic situations, and this leads to technical complications in \cite{6} and in further potential applications of this result in arithmetic geometry. It seems plausible that the new approach developed in the present paper could also lead to progress on unrestricted exponentials in this non-Pfaffian situation, and we have formulated our results in the more general *o-minimal context with this in mind.

1.4. Asymptotic notation. In this paper each appearance of an expression $Z = O_X(Y)$ should be interpreted as shorthand notation for $Z \leq C_X \cdot Y$ where $X \rightarrow C_X$ is a universally fixed, positive valued real function. Similarly we write $Z = \text{poly}_X(Y)$ as shorthand for $Z \leq P_X(Y)$ where $X \rightarrow P_X$ is a universally fixed mapping and $P_X$ is a polynomial with positive coefficients. However we suppress dependence of the constants on $(\delta, \Omega)$, which we consider to be universally fixed throughout the text.

1.5. Acknowledgments. It is our pleasure to thank Yosi Yomdin for insightful discussions on the algebraic lemma, and Alex Wilkie for alerting us of the potential relevance of his notes \cite{25}. In these notes Wilkie introduces a method for obtaining $C^r$-parametrizations in the one-dimensional case with a single reparametrization, rather than by the more traditional induction on $r$. While we did not eventually use this directly in our text, our approach is a kind of discrete version of this idea (so that we can use linear reparametrizations similar to Yomdin’s approach) and certainly inspired by it. Pila \cite{18} has used a similar approach earlier for Pfaffian curves, and his idea also inspires our approach. We note further that the interpolation method that we use to control $X(g,H)$ efficiently as a function of $g$ was also introduced in Wilkie’s important notes \cite{25}.

2. Sharply o-minimal structures

2.1. #o-minimal structures. In this section we introduce the notion of a sharply o-minimal structure (abbreviated *o-minimal). To start, a format-degree filtration (abbreviated FD-filtration) on a structure $S$ is a collection $\Omega = \{\Omega_{\tau,D}\}_{\tau, D \in \mathbb{N}}$ such that each $\Omega_{\tau,D}$ is a collection of definable sets (possibly of different ambient dimensions), with $\Omega_{\tau,D} \subset \Omega_{\tau+1,D} \cap \Omega_{\tau,D+1}$ and $\cup_{\tau,D}\Omega_{\tau,D}$ is the collection of all
definable sets in \( S \). We call the sets in \( \Omega_{\mathcal{F},D} \) sets of format \( \mathcal{F} \) and degree \( D \). We will assume \( \Omega_{\mathcal{F},D} \) only contains subsets of \( \mathbb{R}^n \) for \( n \leq \mathcal{F} \).

A \( \mathbf{o} \)-minimal structure is a pair \( \Sigma := (S, \Omega) \) consisting of an \( \mathbf{o} \)-minimal expansion of the real field \( S \) and an \( \mathbf{FD} \)-filtration \( \Omega \); and for each \( \mathcal{F} \in \mathbb{N} \) a polynomial \( P_{\mathcal{F}}(\cdot) \) such that the following holds:

1. If \( A \in \Omega_{\mathcal{F},D} \) with \( A \subset \mathbb{R}^n \) then \( A^c, \pi_{n-1}(A), A \times \mathbb{R} \) and \( \mathbb{R} \times A \) lie in \( \Omega_{\mathcal{F}+1,D} \).
2. If \( A_1, \ldots, A_k \subset \mathbb{R}^n \) with \( A_j \in \Omega_{\mathcal{F},D_j} \) then \( \cup_i A_i \in \Omega_{\mathcal{F},D} \) and \( \cap_i A_i \in \Omega_{\mathcal{F}+1,D} \) where \( D = \sum_j D_j \).
3. If \( P \in \mathbb{R}[x_1, \ldots, x_n] \) then \( \{ P = 0 \} \in \Omega_{n,\deg P} \).
4. If \( A \in \Omega_{\mathcal{F},D} \) with \( A \subset \mathbb{R} \) then it has at most \( P_{\mathcal{F}}(D) \) connected components,

Axioms 1-2 bear a close analogy to the standard axioms of a first-order structure, keeping track of the formats and degrees of sets defined using the logical operations. Axiom 3 ensures compatibility with the standard notion of degree in the (semi-)algebraic case. Finally Axiom 4 replaces the mere finiteness postulated in standard \( \mathbf{o} \)-minimality by polynomial bounds in degrees.

2.2. Sharp cell decomposition. The following notion is crucial for working with \( \mathbf{o} \)-minimal structures.

**Definition 4.** We say that \((S, \Omega)\) has sharp cell decomposition if for every \( \mathcal{F} \in \mathbb{N} \) there are

\[
a_{\mathcal{F}} \in \mathbb{N}, \quad b_{\mathcal{F}} \in \mathbb{N}[D,k], \quad c_{\mathcal{F}} \in \mathbb{N}[D]
\]

such that the following holds. For every \( \mathcal{F}, D \in \mathbb{N} \) and every \( X_1, \ldots, X_k \in \Omega_{\mathcal{F},D} \) subsets of \( \mathbb{R}^n \), there exists a cylindrical decomposition \( \{C_\eta\} \) of \( \mathbb{R}^n \) compatible with \( X_1, \ldots, X_k \) such that

\[
\#\{C_\eta\} \leq b_{\mathcal{F}}(D,k), \quad \forall \eta : C_\eta \in \Omega_{a_{\mathcal{F}},c_{\mathcal{F}}(D)}.
\]  

We use the following notation for cells from \([4]\). For \( C \subset \mathbb{R}^{n-1} \) and \( a, b : C \to \mathbb{R} \) we set

\[
C \odot \{a(z)\} := \{(z,w) : z \in C, \ w = a(z)\},
\]

\[
C \odot \{a(z), b(z)\} := \{(z,w) : z \in C, \ a(z) < w < b(z)\}.
\]

We will also allow \( a(z) \equiv -\infty \) and \( b(z) \equiv \infty \) in the second case above.

In an upcoming paper we prove, based on ideas from \([7]\), that for every \( \mathbf{o} \)-minimal structure \((S, \Omega)\) there is another \( \mathbf{FD} \)-filtration \( \Omega^* \) with \( \Omega_{\mathcal{F},D} \subset \Omega^*_{\mathcal{F},D} \) for every \( \mathcal{F}, D \) such that \((S, \Omega^*)\) is \( \mathbf{o} \)-minimal with sharp cell decomposition. This implies that Theorem \([2]\) and its consequences actually apply without explicitly assuming that \((S, \Omega)\) has sharp cell decomposition. However to keep matters clear and avoid dependence on our upcoming text we keep this as an extra condition. Our main example \( \mathbb{R}_{\mathbf{Pfaff}} \) does, in any case, admit sharp cell decomposition as explained in \([3]\).

2.3. Some consequences of \( \mathbf{o} \)-minimality and sharp cell decomposition.

The axioms of \( \mathbf{o} \)-minimality imply that whenever \( X_1, \ldots, X_k \in \Omega_{\mathcal{F},D} \) and \( \psi \) is a first-order formula of depth \( \ell \) with basic predicates \( x \in X_j \) then the set \( X \) defined by \( \phi \) satisfies

\[
X \in \Omega_{\mathcal{O}_{\mathcal{F},\ell}(1),\mathbf{poly}_{\mathcal{O},\ell}(D,k)}.
\]

see \([7]\) Section 1.3] for a more precise treatment. Together with sharp cell decomposition, this can be used to effectivize many of the classical constructions of
o-minimality in a rather routine fashion. We record a few instances used in our text to familiarize the reader with this technique.

**Proposition 5** (Connected components). Let \( X \in \Omega_{\mathcal{F}, D} \). Then each connected component of \( X \) is in \( \Omega_{O_{\mathcal{F}}(1), \text{poly}_{\mathcal{F}}(D)} \), and their number is \( \text{poly}_{\mathcal{F}}(D) \).

**Proof.** Perform a cell decomposition. Each connected component is a union of cells. \( \square \)

**Proposition 6** (Stratification). Let \( X \in \Omega_{\mathcal{F}, D} \). Then \( X \) is a disjoint union \( \cup_{\eta} S_{\eta} \) where each \( S_{\eta} \) is connected and regular, and

\[
\#\{S_{\eta}\} = \text{poly}_{\mathcal{F}}(D), \quad \forall \eta : S_{\eta} \in \Omega_{O_{\mathcal{F}}(1), \text{poly}_{\mathcal{F}}(D)}.
\] (13)

**Proof.** Let \( \mu := \dim X \). Let \( S \subset X \) be the \( \mu \)-regular part of \( X \), i.e. the set of points \( p \in X \) such that for some linear projection \( L : \mathbb{R}^{n} \to \mathbb{R}^{\mu} \), the map \( L|_{X} \) is locally invertible at \( p \), and the inverse \( L' : (\mathbb{R}^{\mu}, L(p)) \to X \subset \mathbb{R}^{n} \) is locally \( C^1 \) with Jacobian of rank \( \mu \). This can be written out explicitly as a first-order formula in \( \varepsilon,\delta \)-type language, so the axioms of \( * \)-o-minimality give \( S \in \Omega_{O_{\mathcal{F}}(1), \text{poly}_{\mathcal{F}}(D)} \). Each connected component of \( S \) is a top-dimensional stratum, and the remaining set \( X \setminus S \) can be handled by induction on \( \mu \). \( \square \)

**Proposition 7** (Definable choice). Let \( X \subset \Lambda \times \mathbb{R}^{n} \) with \( X \in \Omega_{\mathcal{F}, D} \), and suppose \( X_\lambda \neq \emptyset \) for every \( \lambda \in \Lambda \). Then there is a map \( F : \Lambda \to \mathbb{R}^{n} \) with \( \text{gr} F \in \Omega_{O_{\mathcal{F}}(1), \text{poly}_{\mathcal{F}}(D)} \) such that \( \text{gr} F \subset X \).

**Proof.** Perform a cylindrical decomposition of \( \Lambda \times \mathbb{R}^{n} \) compatible with \( X \). In particular we obtain a cylindrical decomposition \( \{C_{\eta}\} \) of \( \Lambda \), and over each \( C_{\eta} \) a cylindrical decomposition of \( C_{\eta} \times \mathbb{R}^{n} \) by cells projecting to \( C_{\eta} \). It will be enough to handle each \( C_{\eta} \) separately and then take the unions of the corresponding graphs. Moreover, we may as well consider just one of the cells over \( C_{\eta} \) that is contained in \( X \) for the purpose of defining the choice function. So assume without loss of generality that \( X \) is a cell.

Write \( X = C \odot (a(z), b(z)) \) where \( C \in \Omega_{O_{\mathcal{F}}(1), \text{poly}_{\mathcal{F}}(D)} \) is a cell in \( \Lambda \times \mathbb{R}^{n-1} \) and \( a(z), b(z) : C \to \mathbb{R} \). The cases \( a(z) = -\infty, b(z) = \infty \) and \( C \odot \{a(z)\} \) are treated similarly. We have \( \text{gr} a(z), \text{gr} b(z) \in \Omega_{O_{\mathcal{F}}(1), \text{poly}_{\mathcal{F}}(D)} \) since they can be defined using first-order formulas as the infimum and supremum of the fiber \( C_{z} \). Then we find a choice function \( \hat{F}(\lambda) \) on \( C \) by induction on \( n \), and

\[
F(\lambda) := \left( \hat{F}(\lambda), \frac{a(\lambda, \hat{F}(\lambda)) + b(\lambda, \hat{F}(\lambda))}{2} \right)
\] (14)

is a choice function for \( X \). \( \square \)

2.4. **Sharp derivatives.** If \( f : X \to Y \) is a definable function we will write \( f \in \Omega_{\mathcal{F}, D} \) as shorthand for \( \text{gr} f \in \Omega_{\mathcal{F}, D} \).

**Definition 8.** We say that \((\mathcal{S}, \Omega)\) has sharp derivatives if for every \( \mathcal{F} \in \mathbb{N} \) there are

\[
a_{\mathcal{F}} \in \mathbb{N}, \quad b_{\mathcal{F}} \in \mathbb{N}[D, k]
\] (15)

such that the following holds. Given a definable \( f : \mathbb{R}^{n} \to \mathbb{R} \) with \( f \in \Omega_{\mathcal{F}, D} \), we have for every \( \alpha \in \mathbb{Z}_{\geq 0} \nabla \)

\[
f^{(\alpha)} \in \Omega_{a_{\mathcal{F}}, b_{\mathcal{F}}(D, |\alpha|)}.
\] (16)
Here \(f^{(\alpha)}\) denotes the function with domain of definition equal to the interior of the locus where \(f\) is continuously differentiable to order \(|\alpha|\).

**Remark 9.** In every \(\ast\)-o-minimal structure we have \(f^{(\alpha)} \in \Omega_{a,f,|\alpha|,b_{x-f,|\alpha|}}(D)\) with \(b_{x-f,|\alpha|} \in \mathbb{N}[D]\). Sharp derivatives means that as we take derivatives of high order, the format remains fixed and the degree depends polynomially on the order. We do not know whether this holds for general \(\ast\)-o-minimal structures.

### 3. The restricted sub-Pfaffian structure \(\mathbb{R}_{rPfaff}\)

In this section we let \(\Omega\) denote the \(\ast\)-o-minimal filtration on \(\mathbb{R}_{rPfaff}\) introduced in [7]. The main result of loc. cit. is that \((\mathbb{R}_{rPfaff}, \Omega)\) is a \(\ast\)-o-minimal structure admitting sharp cell decomposition.

**Remark 10.** A small technical issue is that in [7] we considered only subsets of \([0,1]^n\), whereas in the \(\ast\)-o-minimal setting it is of course customary to work in \(\mathbb{R}^n\). Sharp derivatives means that as we take derivatives of high order, the format remains fixed and the degree depends polynomially on the order. We do not know whether this holds for general \(\ast\)-o-minimal structures.

Let \(U \subset \mathbb{R}^n\) and \(f : U \to \mathbb{R}\) with \(f \in \Omega_{\mathcal{F},D}\) and \(\Gamma := \text{gr} f\). By the definition from [7],

\[
\Gamma = \bigcup_i \pi_{n+1}X_i^0
\]

where i) each \(X_i^0\) is a connected component of a semi-Pfaffian \(X_i \subset \mathbb{R}^N\) of degree \(\text{poly}_{\mathcal{F}}(D)\) for some \(N = N(\mathcal{F})\), and ii) the number of \(X_i\) is \(\text{poly}_{\mathcal{F}}(D)\). Moreover according to [7, Lemma 18] we may assume that the projection \(\pi_{n+1}|_{X_i}\) has finite fibers.

Fix one \(X = X_i\) with \(\pi_{n+1}(X_i^0)\) of full dimension in \(\Gamma\). The general case easily reduces to this at the end. By [10] we may further assume that \(X\) is effectively smooth, i.e. is defined by a semi-Pfaffian system

\[
\{F_1 = \ldots = F_{N-n} = 0\} \cap \{G_1 > 0, \ldots, G_M > 0\}
\]

with the differential \(dF_1 \wedge \cdots \wedge dF_{N-n}\) non-vanishing on \(X\). The degrees of the \(F_i, G_j\) in [10] are \(\text{poly}_{\mathcal{F}}(D)\). Removing a smaller-dimensional part, we may assume that the projection \(\pi_n|_X\) is everywhere submersive.

Denote the coordinates on \(\mathbb{R}^N\) by \((x, y)\) where

\[
x = (x_1, \ldots, x_n), \quad y = (y_1, \ldots, y_{N-n}).
\]
By the implicit function theorem and our setup above, around every point in \( X \) one can express \( y \) as a smooth function of \( x \), and

\[ F(x, y) = 0 \implies \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y} \cdot \frac{\partial y}{\partial x} = 0 \implies \frac{\partial y}{\partial x} = -\left( \frac{\partial F}{\partial y} \right)^{-1} \frac{\partial F}{\partial x}. \] (21)

where \( F = (F_1, \ldots, F_{N-n}) \). Note that \( \frac{\partial F}{\partial y} \) is invertible everywhere on \( X \) by our setup. Note \( f(x) = y_1(x) \) on \( \pi_n(X^o) \). Since \( F \) is a vector of Pfaffian functions, all the derivatives in the right hand side are again Pfaffian, and using \( A^{-1} = \det^{-1}(A) \text{adj}(A) \) we can write each \( \frac{\partial F}{\partial x_j} \) in the form \( P_j/Q \) where \( P_j \) is a Pfaffian function and \( Q = \det \frac{\partial F}{\partial y} \). Using this, one can rewrite \( f^{(\alpha)}(x) = y_1^{(\alpha)}(x) \) as a ratio of Pfaffian functions \( P_\alpha/Q^{2|\alpha|} \) with \( \deg P_\alpha = \text{poly}_T(D) \cdot |\alpha| \), the asymptotic constants depending on the Pfaffian chain used to define \( f \) (which are part of the format \( \mathcal{F} \)). This ratio is not formally Pfaffian, but adding a variable \( z \) and an equation \( Q^{2|\alpha|}z = P_\alpha \) to the equations of \( X \) gives a set \( Z \subset \mathbb{R}^{n+1} \) with a connected component \( Z^o \) lying over \( X^o \), such that the projection of \( Z \) to \( (x, z) \) is the graph of \( f^{(\alpha)} \) over \( \pi_n(X^o) \).

Recall that the union of \( \text{poly}_T(D) \) sets \( X^o \) as above define a dense subset of \( \Gamma \). We have thus seen how to define a dense subset of \( \text{gr} f \) as a smooth function of \( x \) at \( x_0 \in U \) equal to the interior of the locus where \( \Gamma_\alpha \) is the graph of a continuously differentiable function is also in \( \Omega_{\text{poly}_T(D),|\alpha|} \) by \( \# o \)-minimality. Setting

\[ \Gamma_\alpha' = \Gamma_\alpha \cap \bigcap_{|\beta| < |\alpha|} D_\beta \] (22)

defines the graph of \( y^{(\alpha)} \) with the correct domain of definition, and the format and degree bounds follow by \( \# o \)-minimality.

4. Norms on \( C^r \)-functions

For

\[ P = \sum_{|\alpha| \leq r} a_\alpha t^\alpha \in \mathbb{R}[t_1, \ldots, t_m] \] (23)

we denote by \( MP = \sum_{|\alpha| \leq r} |a_\alpha| t^\alpha \) the majorant. We set \( \| \cdot \| : = MP(1, \ldots, 1) \).

For a \( C^r \)-smooth function \( f : U \to \mathbb{R} \) on a domain \( U \subset \mathbb{R}^m \) we and \( x_0 \in U \) we denote by

\[ j^r_x f = \sum_{|\alpha| \leq r} \frac{f^{(\alpha)}(x_0)}{\alpha!} t^\alpha \] (24)

the \( r \)-jet of \( f \) at \( x_0 \). We define two norms on \( f \) as follows,

\[ \| f \|_r := \max_{|\alpha| \leq r} \| D^\alpha f \|, \quad \| f \|_{T,r} := \sup_{x \in U} \| j^r_x f \|. \] (25)

As in §12 we extend this to \( F : U \to \mathbb{R}^n \) by coordinate-wise maximum.

For our purposes these two norms are essentially equivalent. Indeed, on the one hand we have

\[ \| f \|_{T,r} \leq e^m \| f \|_r. \] (26)

On the other hand the following lemma is immediate.
Lemma 11. Suppose $f : I^n \to \mathbb{R}$ with $\|f\|_{T,r} \leq 1$. Let $\phi : I^n \to I^n$ be a diagonal affine map with $\text{Im } \phi$ a cube of side-length $1/r$. Then $\|f \circ \phi\|_r \leq 1$.

As a consequence of Lemma 11, given functions of unit $(T, r)$-norm on $I^n$ we can always rescale to obtain bounded $r$-norms using $r^n$ charts.

We usually state our results with $\|f\|_r$, but in some cases $\|f\|_{T,r}$ is more technically convenient, mainly because of the following submultiplicativity and subcompositionality properties.

Lemma 12. The following estimates for products and compositions hold:

1. Let $f, g : U \to \mathbb{R}$ be $C^r$-smooth. Then $\|fg\|_{T,r} \leq \|f\|_{T,r} \cdot \|g\|_{T,r}$.
2. Let $F : U \to \mathbb{R}^n$ and $g : V \to \mathbb{R}$ be $C^r$-smooth with $\text{Im } f \subset V$. Suppose $\|F_i\|_{T,r} \leq 1$ for $i = 1, \ldots, n$. Then $\|g \circ F\|_{T,r} \leq \|g\|_{T,r}$.

Proof. Part i follows from

$$\|M_j^r(fg)(1, \ldots, 1) \| \leq \|M_j^r f(1, \ldots, 1) \| \cdot \|M_j^r g(1, \ldots, 1) \|$$

which holds since $j^r_x(fg)$ is just $j^r_x(f)$, $j^r_x(g)$ truncated to degree $r$. Part ii follows in a similar fashion, this time noting that $j^r_x(g \circ F)$ is just $j^r_x(g)(j^r_x F_1, \ldots, j^r_x F_n)$ truncated to degree $r$. \hfill \Box

5. Proof of the algebraic lemma

We will prove the algebraic lemma in the following equivalent form which is more suitable for an inductive argument. Below, we think of maps $F : \Lambda \times I^n \to I^k$ as definable families of maps $\{F_\lambda : I^n \to I^k\}_{\lambda \in \Lambda}$, where $F_\lambda := F(\lambda, \cdot)$.

Lemma 13. Let $r \in \mathbb{N}$ and $\varepsilon > 0$. Let $F : \Lambda \times I^n \to I^k$ with $F_j \in \Omega_{T,r}$ for $j = 1, \ldots, k$. Then there exists a collection $\{\phi_\eta : \Lambda \times I^n \to I^n\}$ such that for every $\lambda \in \Lambda$ we have:

1. $\|F_\lambda \circ \phi_{\eta,\lambda}\|_r \leq 1$, ii) $\cup_\eta \text{Im}(F_\lambda \circ \phi_{\eta,\lambda}) \subseteq \varepsilon \text{Im } F_\lambda$, and iii)
2. $\#\{\phi_\eta\} \leq \text{poly}_r(D, r, k, |\log \varepsilon|)$, $\forall \eta : \phi_\eta \in \Omega_{O_\varepsilon(1),\text{poly}_g(D, r)}$.

Lemma 13 implies the following family version of Lemma 12.

Lemma 14. Let $r \in \mathbb{N}$ and $\varepsilon > 0$. Let $X \subset \Lambda \times I^n$ with $\mu := \max_\lambda \dim X_\lambda$ and $X \in \Omega_{T,r}$. Assume $X$ has no empty fibers. Then there exists a collection $\{\phi_\eta : \Lambda \times I^\mu \to X\}$ such that for every $\lambda \in \Lambda$ we have:

1. $\|\phi_{\eta,\lambda}\|_r \leq 1$, ii) $\cup_\eta \text{Im}(\phi_{\eta,\lambda}) \subseteq \varepsilon X_\lambda$, and iii)
2. $\#\{\phi_\eta\} \leq \text{poly}_r(D, r, |\log \varepsilon|)$, $\forall \eta : \phi_\eta \in \Omega_{O_\varepsilon(1),\text{poly}_g(D, r, |\log \varepsilon|)}$.

Proof. First perform a cell decomposition of $\Lambda \times I^n$ compatible with $X$, to cover $X_\lambda$ by $\text{poly}_r(D)$ images $\text{Im } F_\theta \lambda$ for $F_\theta : \Lambda \times I^n \to I^n$. The non-empty fibers are required to guarantee we can always do this. Then apply Lemma 13 to each of these maps. The collection of all resulting $F_\theta \circ \phi_\eta$ establishes the conclusion the lemma. \hfill \Box

Remark 15. Suppose $\Lambda = \Lambda_1 \cup \cdots \cup \Lambda_N$ is a definable subdivision of $\Lambda$ with $N = \text{poly}_r(D, r, k, |\log \varepsilon|)$ and $\Lambda_i \in \Omega_{O_\varepsilon(1),\text{poly}_g(D, r, |\log \varepsilon|)}$. Suppose we prove Lemma 13 for $F$ restricted to each $\Lambda_i$ separately, say giving collections

$$\{\phi_{i,j} : \Lambda_i \times I^n \to I^n\} \quad i = 1, \ldots, N, \quad j = 1, \ldots, M$$

allowing repetitions to make these collections have the same size $M$. Then the collection $\{\phi_j := \cup_i \phi_{i,j}\}_j$ proves Lemma 13 for $\Lambda$ (the degree and format bounds
follow from *o-minimality). A similar remark applies for Lemma 14. In the proof below we will often use this subdivision argument without explicit mention.

To make the notation more suggestive, we sometimes denote the coordinates on \( \mathbb{R}^n \) by \((x, y_1, \ldots, y_{n-1})\). The proof of Lemma 14 will occupy the remainder of this section. We proceed by induction on \( n \), treating the base case \( n = 1 \) in the following subsection. We record a simple lemma that is useful in many stages of our argument.

**Lemma 16.** Let \( F : X \to Y \) be 1-Lipschitz, and suppose \( \{ \phi_\eta : D_i \to X \} \) satisfies \( \cup_\eta \operatorname{Im} \phi_\eta \subseteq X \). Then \( \cup_\eta \operatorname{Im}(F \circ \phi_\eta) \subseteq \varepsilon \operatorname{Im} F \).

In particular this implies that when every \( F_\lambda \) is 1-Lipschitz we can replace the condition \( \cup_\eta \operatorname{Im}(F_\lambda \circ \phi_{\eta, \lambda}) \subseteq \varepsilon \operatorname{Im} F_\lambda \) in Lemma 13 by \( \cup_\eta \operatorname{Im} \phi_{\eta, \lambda} \subseteq \varepsilon \). We will often use this remark after performing a pullback to satisfy the 1-Lipschitz condition.

### 5.1. The case \( n = 1 \)

The main difficulty in proving Lemma 13 is to get polynomial growth with respect to \( r \). For a fixed \( r \) the classical proof of Yomdin-Gromov gives the same statement, even with a true cover in place of the \( \varepsilon \)-cover. We record below the \( C^1 \)-version that we will need in the sequel.

**Lemma 17.** Let \( F : \Lambda \times I \to I^k \) with \( F_i \in \Omega_{\tau, D} \). Then there exists a collection \( \{ \phi_\eta : \Lambda \times I \to I \} \) such that for every \( \lambda \in \Lambda \) we have i) \( \|F_\lambda \circ \phi_{\eta, \lambda}\|_1 \leq 1 \), ii) \( \cup_\eta \phi_{\eta, \lambda}(I) = I \), and iii) \( \#\{\phi_\eta\} \leq \operatorname{poly}_{\tau}(D, k) \).

**Proof.** Assume without loss of generality that \( f(x) = x \) is among the \( F_i \). Denote \((\cdot)' = \frac{\partial}{\partial x} (\cdot)\). Perform a cell decomposition of \( \Lambda \times I \) compatible with the sets of zeros of all the functions \( |F_i'| - |F_j'| \) for \( i, j = 1, \ldots, k \) as well as with the sets of points where \( F_i' \) is undefined. We have \( \operatorname{poly}_{\tau}(D, k) \) cells, each in \( \Omega_{\tau, (1, \operatorname{poly}_{\tau}(D))} \).

It will suffice to handle each cell separately. For cells of the form \( C \cap \{ a(\lambda) \} \) one can cover their image by a constant map, so consider a cell \( C \cap \{ a(\lambda), b(\lambda) \} \). Since each \( |F_i'| - |F_j'| \) is either identically vanishing or identically non-vanishing on the cell, there is one \( F_i \), without loss of generality \( F_1 \), such that
\[
|F_1'| \gg |F_j'| \quad \forall j = 2, \ldots, k
\]
uniformly over the cell. In particular \( |F_1| \gg 1 \). Set
\[
F_1(\lambda, I) = (A(\lambda), B(\lambda))
\]
and define \( \tilde{\phi} : C \circ (A(\lambda), B(\lambda)) \to I \) by \( \tilde{\phi}(\lambda, s) = (\lambda, F_1^{-1}(s)) \). By \([29]\) we have
\[
|(F_i \circ \tilde{\phi}(\lambda, s))'| = |F_{i, 1}(\tilde{\phi}(\lambda(s)))|\tilde{\phi}(\lambda(s))'|
\]
so \( \|F_\lambda \circ \tilde{\phi}_\lambda\|_1 \leq 1 \). Finally, let \( \phi : C \times I \to I \) be the pullback of \( \tilde{\phi} \) by a linear rescaling map \( C \times I \to C \circ (A(\lambda), B(\lambda)) \). Since \( (A(\lambda), B(\lambda)) \subset I \) this only decreases derivatives, so the collection of maps \( \phi \) thus obtained satisfies the conditions of the lemma.

We first apply Lemma 17 to \( F \). Pulling back \( F \) by each of the \( \phi_\eta \) thus obtained, we may assume without loss of generality that \( \|F_\lambda\|_1 \leq 1 \) for every \( \lambda \in \Lambda \). In particular, each \( F_\lambda \) is 1-Lipschitz (with respect to the \( \ell_\infty \)-norm).
Perform a cell decomposition of $\Lambda \times I$ compatible with the sets of zeros of all the functions $F^{(j)}$ for $i = 1, \ldots, k$ and $j = 0, \ldots, r + 1$, as well as the sets of points where these functions are undefined. We have $\Omega_{\mathcal{O}(1), \text{poly}_D(D,r)}$. It will suffice to handle each cell separately. For cells of the form $C \cap \{a(\lambda)\}$ there is nothing to prove, so we consider cells $C \cap (a(\lambda), b(\lambda))$. Pulling back by the affine map $C \cap I \to C \cap (a(\lambda), b(\lambda))$ only decreases derivatives, so without loss of generality it now suffices to prove Lemma 13 assuming each $F_\lambda$ is 1-Lipschitz and has constant-signed derivatives up to order $r + 1$. The result now follows from the following lemma.

**Lemma 18.** Let $f : I \to I$ such that $f^{(j)}$ has constant sign for $j = 0, \ldots, r + 1$. Then for every $M > 1$ and $j = 0, \ldots, r$ we have

$$|f^{(j)}(x)| < M^j \text{ whenever } \text{dist}(x, \partial I) > j/M. \quad (35)$$

**Proof.** We proceed by induction, the case $j = 0$ being vacuous. Suppose the claim is proved for $f^{(j)}$. Assume $f^{(j+1)}$ is weakly-increasing (or weakly-decreasing, which is analogous) and suppose toward contradiction that $f^{(j+1)}(x) \geq M^{j+1}$ for some $x$ with $\text{dist}(x, \partial I) > (j+1)/M$ (the case $f^{(j+1)}(x) \leq -M^{j+1}$ being analogous). Then $f^{(j+1)} > M^{j+1}$ throughout the interval $[x, x+1/M]$. Thus $f^{(j)}(x+1/M) − f^{(j)}(x) \geq M^j$. This contradicts the inductive hypothesis, since both $x$ and $x + 1/M$ have distance at least $j/M$ to $\partial I$, and $f^{(j)}$ is constant-signed and bounded in absolute value by $M^j$ at both points.

It follows from Lemma 13 that if $\phi : I \to I$ is an affine translation with the length of $\phi(I)$ smaller than $\text{dist}(\phi(I), \partial I)/r$ then $\|F_\lambda \circ \phi\|_r \leq 1$ for every $\lambda$. It is an elementary exercise that with $\text{poly}(r, |\log \varepsilon|)$ such maps we can cover $I_\varepsilon := (\varepsilon, 1−\varepsilon)$. Finally, since $F_\lambda$ is 1-Lipschitz for every $\lambda \in \Lambda$ we have $F_\lambda(I_\varepsilon) \subseteq \varepsilon F_\lambda(I)$ so this covering satisfies the conditions of Lemma 13.

We record a corollary of the proof above for later use, where we cover the domain $I$ by linear maps but skip the 1-Lipschitz preparation step, so we only get an $\varepsilon$-cover of the domain $I$ but not necessarily of the image under $F$.

**Corollary 19.** Let $r \in \mathbb{N}$ and $\varepsilon > 0$. Let $F : \Lambda \times I \to I^k$ with $F_j \in \Omega_{\mathcal{F}, D}$ for $j = 1, \ldots, k$. Then there exists a collection $\{\phi_\eta : \Lambda \times I \to I\}$ such that for every $\lambda \in \Lambda$ we have i) $\|F_\lambda \circ \phi_\eta\|_r \leq 1$, ii) $\cup_{\eta} \phi_\eta(I) \subseteq I$, iii) $\phi_{\eta, \lambda}$ is affine, and iv)

$$\#\{\phi_\eta\} \leq \text{poly}_{\mathcal{F}}(D, r, k, |\log \varepsilon|), \quad \forall \eta : \phi_\eta \in \Omega_{\mathcal{O}(1), \text{poly}_D(D,r)}. \quad (36)$$

**5.2. Reduction to bounded $y$-derivatives.** We now continue with the case of general $n$, assuming that the case $n − 1$ is already established.

Apply the inductive hypothesis to $F : (\Lambda \times I_x) \times I_y^{n-1} \to I^k$ with $\varepsilon/2$ in place of $\varepsilon$, viewing $x$ as an additional parameter. Let $\{\phi_\eta\}$ be the resulting collection. It will suffice to establish the conclusion of Lemma 13 with $\varepsilon/2$ in place of $\varepsilon$ and each $F \circ (\lambda, x, \phi_\eta)$ in place of $F$. In other words, without loss of generality it suffices to prove Lemma 13 assuming that $\|F_\lambda(x, \cdot)\|_r \leq 1$ for every $\lambda \in \Lambda$ and every $x \in I$. Below we keep working with $\varepsilon$ rather than $\varepsilon/2$ to simplify notations, and we make similar reductions later in the proof.

**5.3. Reduction to 1-Lipschitz.** Recall the following lemma from [5].

**Lemma 20 ([5]).** Let $F : I_x \times I_y^{n-1} \to I$ be definable and suppose that for each $i = 1, \ldots, n−1$ the derivative $f_{yi}(x,y)$ is uniformly bounded over all $y$ where it is
defined, for almost every \( x \in I_x \). Then \( f_x(x, y) \) is also uniformly bounded for all \( y \) where it is defined, for almost every \( x \in I_x \).

Proof. This is essentially [5, Lemma 16], but we sketch the argument to illustrate the connection with the material above. Suppose toward contradiction that \( f_x(x, y) \) is unbounded in \( y \) for every \( x \in J \subset I \). Without loss of generality assume \( J = I \). Then by definable choice one can choose a function \( \gamma : I \to I^{n-1} \) such that each \( f_y(x, \gamma(x)) \) is defined and \( f_x(x, \gamma(x)) > M \) for every \( x \in I \) (or \( f(x, \gamma(x)) < -M \), which is analogous). Moreover the format and degree of \( \gamma \) are bounded independently of \( M \). By Corollary 19 applied to \( \gamma \) and \( f(x, \gamma(x)) \) with \( r = 1 \) and say \( \varepsilon = 0.1 \) we find a subinterval \( I' \subset I \) where both \( \|\gamma'(x)\| \) and \( |f(x, \gamma(x))'| \) are bounded from above by a constant independent of \( M \). One can take the longest of the intervals \( \phi_\varepsilon(I) \) for example, which has length bounded from below uniformly in \( M \). This is now a contradiction for \( M \gg 1 \) because

\[
M < f_x(x, \gamma(x)) = f(x, \gamma(x))' - \sum_{i=1}^{n-1} f_{y_i}(x, \gamma(x))\gamma_i'(x)
\]

and the right-hand side is uniformly bounded. \( \square \)

For \( i = 1, \ldots, k \) we define \( S_i \subset \Lambda \times I^n \) by

\[
S_i := \{ (\lambda, x, y) : \| (F_{i})_x(\lambda, x, y) \| \geq \frac{1}{2} \sup_{y' \in I^{n-1}} \| (F_{i})_x(\lambda, x, y') \| \}
\]

where the supremum is taken over the points where \( (F_{i})_x(\lambda, x, y) \) is defined. By Lemma 20 the supremum is finite, for each \( \lambda \in \Lambda \), for almost every \( x \). For \( x \) where the supremum is infinite we consider that the condition is vacuous, i.e. every \( (\lambda, x, y) \) is included in \( S_i \) in this case. Clearly \( S_i \in \Omega_{O_{\varepsilon}(1),poly_{\varepsilon}(D)} \).

By definable choice we may choose subsets \( \Gamma_i \subset S_i \) such that \( \Gamma_i \) contains exactly one \((\lambda, x, y)\) for every \((\lambda, x)\). In particular, sharp cylindrical decomposition shows that \( \Gamma_i \in \Omega_{O_{\varepsilon}(1),poly_{\varepsilon}(D)} \). By definition \( \Gamma_i \) is a graph of an \((n-1)\)-tuple of functions

\[
\gamma_{i,1}, \ldots, \gamma_{i,n-1} : \Lambda \times I_x \to I
\]

and \( \gamma_{i,j} \in \Omega_{O_{\varepsilon}(1),poly_{\varepsilon}(D)} \) as well.

We apply Lemma 17 to the tuple including the functions \( \gamma_i \) and \( x \), as well as \( F \circ (\lambda, x, \gamma_i) \) for every \( i = 1, \ldots, k \). For every \( \phi_\eta \) thus obtained let

\[
\Phi_\eta(\lambda, t, y) = (\lambda, \phi_\eta(\lambda, t), y).
\]

(40)

Denote \( F_\eta := F \circ \Phi_\eta \). It will suffice to establish the conclusion of Lemma 13 for each \( F_\eta \) in place of \( F \). Moreover we obviously still have \( \| F_{\eta,\lambda}(t, \cdot) \|_\varepsilon \leq 1 \) for every \( \lambda \in \Lambda \) and every \( t \in I \). Denote \( \Gamma_{i,\eta} := \Phi_{\eta}^{-1}(\Gamma_i) \). Then \( \Gamma_{i,\eta} \) is the graph of an \((n-1)\)-tuple of functions \( \gamma_{i,j,\eta} := \gamma_{i,j} \circ (\lambda, \phi_\eta) \) with \( \| \gamma_{i,j,\eta,\lambda} \|_1 \leq 1 \) for every \( \lambda \in \Lambda \). Note that for every \((\lambda, t, y) \in \Gamma_{i,\eta} \) we have

\[
\| (F_{i,\eta})_t(\lambda, t, y) \| = \| (F_{i})_x \circ \Phi_\eta(\lambda, t, y) \cdot (\phi_\eta)_t(\lambda, t) \| \geq \frac{1}{2} \sup_{y' \in I^{n-1}} \| (F_{i})_x(\lambda, \phi_\lambda(t), y') \cdot (\phi_\eta)_t(\lambda, t) \| = \frac{1}{2} \sup_{y' \in I^{n-1}} \| (F_{i,\eta})_t(\lambda, t, y') \|.
\]

(41)
In other words, $\Gamma_{i,\eta}$ satisfies the same definition in the $(\lambda, t, y)$ coordinates as $\Gamma_i$ in the $(\lambda, x, y)$ coordinates. Clearly the sets and functions defined above have format $O_r(1)$ and degree poly$_r(D)$.

We claim that $|(F_{i,\eta},t)| = O_n(1)$ whenever it is defined. According to (41) it is enough to check this on the curves $\Gamma_{i,\eta}$. We compute the derivative of $F_{i,\eta}$ along this curve,

$$1 \geq |(F_{i,\eta}(\lambda, t, \gamma_i, \eta))_{t}| = |(F_{i,\eta})_{t} + \sum_{j=1}^{n-1}(F_{i,\eta})_{y_j}(\gamma_i, \eta, \lambda, \lambda)| \geq |(F_{i,\eta})_{t}| - n - 1, \quad (42)$$

and rearranging we see that $|(F_{i,\eta})_{t}| = O_n(1)$ as claimed.

According to the following lemma, each $F_{\eta}(\lambda, \gamma)$ is $O_n(1)$-Lipschitz for each $\lambda \in \Lambda$ after a subdivision of $I_x$ into intervals.

**Lemma 21.** Let $f : I_x \times I^{n-1}_y \to I$ be definable with $\|f(x, \cdot)\|_1 \leq 1$ everywhere and $|f_x| \leq 1$ whenever $f_x$ is defined. Then the locus of discontinuity of $f$ is contained in finitely many hyperplanes $x_1, \ldots, x_N$, and $f$ is $O_n(1)$-Lipschitz on each $(x_i, x_{i+1}) \times I^{n-1}_y$ (where we take $x_1 = 0$ and $x_N = 1$).

**Proof.** Since $f$ is $O_n(1)$-Lipschitz in the $y$-direction, it is easy to check that if $f$ is discontinuous at $(x_0, y_0)$ it is also discontinuous at every point of $(x_0, y_0) \times U_{\delta}(y_0)$ for some $\delta \ll 1$. Since the locus of discontinuity has empty interior, the first claim follows.

Let $V \subset (x_i, x_{i+1}) \times I^{n-1}_y$ denote the locus where $f_x$ is undefined. Consider two points in $(x_i, x_{i+1}) \times I^{n-1}_y$ and the straight line $\gamma$ connecting them. If $\gamma \cap V$ is finite then $f|_{\gamma}$ is piecewise $O_n(1)$-Lipschitz and continuous, so it is 1-Lipschitz. In general, we deform $\gamma$ into a curve $\gamma + v$ with $v \in \mathbb{R}^n$. For the same reason as above, we can choose $\|v\|$ arbitrarily small so that $(\gamma + v) \cap V$ is finite, and $f$ is $O_n(1)$-Lipschitz on $\gamma + v$. The claim follows by continuity of $f$. \hfill \Box

Perform a cell decomposition of $\Lambda \times I_x$ compatible with the projections of the loci of discontinuity of $F_{i,\eta}$ for each $i = 1, \ldots, k$ to $\Lambda \times I_x$, giving cells of the form either $C = C_\lambda \cap (a(\lambda))$ or $C = C_{\lambda, \alpha} \cap (a(\lambda), b(\lambda))$ where in the latter case $F_{\eta,\lambda}$ is $O_n(1)$-Lipschitz in $(a(\lambda), b(\lambda)) \times I^{n-1}$ for every $\lambda \in \Lambda$. In the former case we can handle $F_{\eta,\lambda}C \times I^{n-1}$ by the inductive hypothesis. In the latter case, rescaling $(a(\lambda), b(\lambda))$ back to $(0, 1)$ only improves the Lipschitz constant in $F_{\eta,\lambda}(C \times I^{n-1})$. Applying a further linear subdivision in the $x, y$ coordinates we may further reduce the Lipschitz constant to 1 to simplify our notations.

Returning to the original notation, we conclude that it will suffice to establish the conclusion of Lemma 16 assuming that $\|F_\lambda(x, \cdot)\|_r \leq 1$ for every $\lambda \in \Lambda$ and every $x \in I$, and $F_\lambda$ is 1-Lipschitz for every $\lambda$.

5.4. Controlling higher derivatives. We start off similarly to §5.3. For $i = 1, \ldots, k$ and $\alpha \in \mathbb{Z}^n_{\geq 0}$ with $|\alpha| \leq r$ we define $S_{i,\alpha} \subset \Lambda \times I^n$ by

$$S_{i,\alpha} := \{(\lambda, x, y) : |F^{(\alpha)}(\lambda, x, y)| \geq \frac{1}{2} \sup_{y' \in I^{n-1}} |F^{(\alpha)}(\lambda, x, y')|\} \quad (43)$$

where the supremum is restricted to those points where $F^{(\alpha)}(\lambda, x, y')$ is defined. Repeatedly applying Lemma 20 and using the fact that for $\alpha_1 = 0$ all the $F^{(\alpha)}$ are
Lemma 22. Let \( \lambda \in \Lambda \). Let \( \Gamma \) be one of \( (\Gamma_{i,\alpha,\eta})_{\lambda} \) and \( \gamma : I \rightarrow \Gamma \) be the tuple \( \gamma_{i,\alpha,\eta,\lambda} \). Let \( G = F_{i,\eta,\lambda} \) for some \( l, \alpha, \eta \). Then for each \( \beta \in \mathbb{Z}^{n}_{\geq 0} \) with \( |\beta| \leq r \) we have

\[
\|G^{(\beta)} \circ (t, \gamma)\|_{x, r - \beta} \leq E(\beta_1), \quad E(\beta_1) := e \cdot (r + (n - 1)e)^{\beta_1}.
\]

Proof. We will work by induction on \( \beta_1 \). For \( \beta_1 = 0 \) the estimate (48) follows from the fact that \( F^{(\beta)} \circ (\lambda, x, \gamma_{i,\alpha}) \) was included in the tuple of functions to which...
Corollary 19 was applied. Indeed, $F^{(\beta)} \circ \Phi_\gamma = (F \circ \Phi_\gamma)^{(\beta)}$ and it follows that
\[
G^{(\beta)} \circ (t, \gamma) = (F_1 \circ \Phi_\eta)^{(\beta)} \circ (\lambda, t, \gamma_{i, \alpha, \eta, \lambda}) \\
= F_1^{(\beta)} \circ \Phi_\eta \circ (\lambda, t, \gamma_{i, \alpha, \eta, \lambda}) \\
= F_1^{(\beta)} \circ (\lambda, \phi_\eta(\lambda, t), \gamma_{i, \alpha} \circ \phi_\eta(\lambda, t)) \\
= F_1^{(\beta)} \circ (\lambda, x, \gamma_{i, \alpha}) \circ \phi_\eta(\lambda, t).
\]
so the $(T, r)$-norm of the left hand side is bounded according to Corollary 19 and (20).

Suppose now that (48) is proved for all $\beta'$ with $\beta'_1 < \beta$. Compute
\[
(G^{(\beta-e_1)} \circ (t, \gamma))' = G^{(\beta)} \circ (t, \gamma) + \sum_{j=2}^{n} (G^{(\beta-e_1+e_j)} \circ (t, \gamma)) \cdot \gamma_j',
\]
where $e_1, \ldots, e_n \in \mathbb{Z}_{\geq 0}$ and $e_{i,j} = \delta_{i,j}$. By the inductive hypothesis
\[
\|G^{(\beta-e_1)} \circ (t, \gamma)\|_{T, r - \beta_1 + 1} \leq E(\beta_1 - 1).
\]
Thus the left-hand side of (50) has $(T, r - \beta_1)$-norm bounded by $rE(\beta_1 - 1)$. Similarly
\[
\|G^{(\beta-e_1+e_j)} \circ (t, \gamma)\|_{T, r - \beta_1} \leq E(\beta_1 - 1)
\]
by induction and $\|\gamma_j\|_{T, r - \beta_1} \leq e$ since $\|\gamma_j\|_r \leq 1$. Rearranging and using Lemma 12 we conclude that
\[
\|G^{(\beta)} \circ (t, \gamma)\|_{T, r - \beta_1} \leq rE(\beta_1 - 1) + (n - 1) e E(\beta_1 - 1)
\]
as claimed.

Finally we conclude that $|F^{(\alpha)}_{i, \eta, \lambda}| = O_n(|\alpha|)$ whenever it is defined, for any $\lambda \in \Lambda$, any $i = 1, \ldots, k$ and any $|\alpha| \leq r$. Indeed, according to (17) it is enough to check this on the curve $\Gamma_{i, \alpha, \eta}$ and there it holds by Lemma 22 taking $G = F^{(\alpha)}_{i, \eta, \lambda}$ since the $(T, r)$-norm bounds the maximum norm. A further linear subdivision into cubes of length $1/r$ as in Lemma 11 then gives $\|F_{i, \eta, \lambda}\|_r \leq 1$, whenever the derivatives are defined.

Returning again to the original notation, we conclude that it will suffice to establish the conclusion of Lemma 13 assuming that $F_{\lambda}$ is 1-Lipschitz and $\|F_{\lambda}\|_r \leq 1$ for every $\lambda \in \Lambda$.

5.5. Final clean up. By now we have satisfied the conclusions of Lemma 13 except that some derivatives of $F$ may be undefined at some points. Let $V_{i, \alpha}$ denote the locus where $F^{(\alpha)}_{i, \eta}$ is undefined, for $i = 1, \ldots, k$ and $|\alpha| \leq r$. Perform a cell decomposition of $\Lambda \times I^n$ compatible with every $V_{i, \alpha}$ giving poly$^F(D, r, k)$ cells, each in $\Omega_{O(1), \text{poly}(D, r)}$. This induces a cell decomposition $\{C_\nu\}$ of $\Lambda \times I^{n-1}$.

By Lemma 14 we get a collection $\{\phi_{\nu, \eta} : \Lambda \times I^{n-1} \to C_\nu\}$ such that
\[
\bigcup_{\nu} (\phi_{\nu, \eta}(I^{n-1})) \subseteq C_{\nu, \lambda}
\]
and $\|\phi_{\nu, \eta}\|_r \leq 1$. Moreover,
\[
\#(\phi_{\nu, \eta}) \leq \text{poly}_T(D, r, |\log \epsilon|), \quad \forall \nu, \eta : \phi_{\nu, \eta} \in \Omega_{O(1), \text{poly}(D, r)}.
\]
Since $F_{\lambda}$ is 1-Lipschitz for every $\lambda$ it will be enough to prove the claim for each pullback $F \circ (\phi_{\nu, \eta}, x_\eta)$, and we may further restrict to the case where $C_{\nu, \lambda}$ has full dimension in $I^{n-1}$. Then $\|F_{\lambda} \circ (\phi_{\nu, \eta, \lambda}, x_\eta)\|_r$ is bounded for every $\lambda$ by Lemma 12.
and by linear subdivision as in Lemma 11 one can return to unit $r$-norms (and after further subdivision to 1-Lipschitz).

In other words we may replace $F$ by each $F \circ (\phi_{\nu,\eta} \cdot x_{\nu})$ and assume without loss of generality that $V := \cup_{i,a} V_{i,a}$, i.e. the locus of non-smoothness of $F$, is already given by graphs of functions $G_1, \ldots, G_q \in \Omega_D(1)$ with $q = \poly(D, k, r)$, where $G_i : \Lambda \times I^{n-1} \rightarrow I$. We also may assume $G_1 = 0$ and $G_q = 0$ for simplicity.

Now apply the inductive hypothesis, i.e. Lemma 13 in dimension $n - 1$, to the tuple $x_1, \ldots, x_{n-1}$ and to $G_1, \ldots, G_q$ on $\Lambda \times I^{n-1}$. Making the same reduction as above, we may now assume without loss of generality that $\|G_{1,\lambda}\|_r, \ldots, \|G_{q,\lambda}\|_r \leq 1$ for every $\lambda$. We cover

$$\Lambda \times I^n \setminus V = \bigcup_{i=1}^q (\Lambda \times I^{n-1}) \circ (G_i, G_i+1)$$

(56)

by images of affine maps

$$(\lambda, x_1, \ldots, x_{n-1}, t) \rightarrow (\lambda, x_1, \ldots, x_{n-1}, tG_i + (1-t)G_i)$$

(57)

with similarly bounded $r$-norms, and finally by another pullback step as above reduce to the case $V = \emptyset$, finishing the proof.

6. Point counting

In this section we give the proof of Theorem 24. Since the argument is fairly standard by now we focus mostly on the novel parts. The key proposition is the following.

Proposition 23. There are appropriate choices of $r, d = \poly_m(g, \log H)$ and $\log \varepsilon = -\poly_m(g, \log H)$ such that the following holds. Let $\phi : I^m \rightarrow X$ such that $\phi(I^m \subseteq \varepsilon) \subseteq I^{m+1}$ and $\|\phi\|_r \leq 1$. Then there exists a polynomial $P \in \mathbb{R}[x_1, \ldots, x_{m+1}] \setminus \{0\}$ of degree $d$ such that $X(g, H) \subseteq \{P = 0\}$.

Proof. There are two approaches to proving such a statement. The first, due to Bombieri-Pila [8], is based on interpolation determinants. The second, due to Wilkie [25] is based on the Siegel lemma. Both of these are normally stated with $\phi(I^m) = X$. We briefly show that for both methods the weaker assumption $\phi(I^m \subseteq \varepsilon) \subseteq \epsilon X$ is sufficient. After cutting into $2^m$ pieces we may assume that the domain of $\phi$ is $J^m$ where $J := (0,1/2)$ instead of $I^m$.

We start with the interpolation determinant method, which directly applies to the case $g = 1$, i.e. to $X(\mathbb{Q}, H)$. Recall that an interpolation determinant is

$$\Delta^d(p_1, \ldots, p_\mu) = \det(p_i^{\alpha}) \quad \text{for} \quad \alpha \in \mathbb{R}_{\geq 0}^{m+1}, |\alpha| \leq d$$

(58)

where $p_i \in \mathbb{R}^{m+1}$ and $\mu$ is the dimension of the space of polynomials of degree at most $d$ in $m + 1$ variables. To prove $X(\mathbb{Q}, H) \subseteq \{P = 0\}$ it is enough to show that $\Delta^d(\cdot)$ vanishes for any $\mu$-tuple of points in $X(\mathbb{Q}, H)$. This is done as follows. First, for any such tuple $p$ one estimates the height of $\Delta^d(p)$, concluding that either $\Delta^d(p) = 0$ or $|\Delta^d(p)| \geq H^{-d\mu}$. On the other hand an analytic estimate shows that for an appropriate choice of $r, d$ as above we have $|\Delta^d(p)| < \frac{1}{2} H^{-d\mu}$. These contradicting estimates force $\Delta^d(p) = 0$ on $\phi(J^m)$ and finish the proof.

In our context, we would like to extend this to the case $p_1, \ldots, p_\mu \in U_{\varepsilon}(\phi(J^m))$. Let $q_1, \ldots, q_\mu \in \phi(J^m)$ with $\dist(p_i, q_i) \leq \varepsilon$. Then $\Delta^d(q) < \frac{1}{2} H^{-d\mu}$ as above, and
if show $|\Delta^d(q) - \Delta^d(p)| < \frac{1}{2}H^{-d\mu}$ we can finish as above. One easily estimates $\|\partial\Delta^d(p)/\partial p\| \leq d!!$ in $\mathbb{P}^{m^\infty(n+1)}$, so choosing $\varepsilon \sim \frac{1}{2}H^{-d\mu}/(d(\mu + 2)!)$ suffices.

We now consider Wilkie’s approach. Using Siegel’s lemma one constructs a polynomial $P \in \mathbb{Z}[x_1, \ldots, x_{m+1}]$ of degree $d$ and coefficients bounded by some $N$ such that $P(\phi_1, \ldots, \phi_{m+1})$ has many small Taylor coefficients. Liouville’s inequality gives for any $x \in X(g, H)$ that either $P(x) = 0$ or

$$|P(x)| \geq (d^{m+1}N\cdot HH^{d(m+1)})^g.$$  

(59)

Denote the right-hand side by $R$. Wilkie shows that for an appropriate choice of $r, d$ as above (and also log $N = \text{poly}_m(g, \log H)$) we have $|P(x)| < R/2$ whenever $x \in \phi(J^m)$, thus forcing $P$ to vanish on $x$. In fact Wilkie considered $X(Q, H)$ and the one-dimensional case, but see [12] Proposition 16 or [2] Proposition 28 for a treatment of the general case.

If we take $x \in U_\varepsilon(\phi(I^m))$ instead, and $x_0 \in \phi(I^m)$ with $\text{dist}(x, x_0) \leq \varepsilon$, then as above it will suffice to show that $|P(x_0) - P(x)| < R/2$. The bounds on $d, N$ easily imply log $|P/\partial x| \leq \text{poly}_m(g, \log H)$ in $[0, 1]^n$, so choosing an appropriate log $\varepsilon = -\text{poly}_m(g, \log H)$ suffices to finish the proof. \hfill $\Box$

We proceed to the proof of Theorem 2, which is very similar to [19] and [11] modulo the sharper Proposition 23. We proceed by induction on dimension $m := \dim X$, the zero-dimensional case being trivial by °o-minimality. Suppose the claim is proved for $X$ of dimension smaller than $m$.

Up to inverting and negating some coordinates (which does not affect height) one can cut $X$ into pieces contained in $[0, 1]^n$, so assume this without loss of generality. Apply Lemma 2 to $X$ to obtain a collection $\{\phi_\eta\}$ of size $\text{poly}_d(D, g, \log H)$ and $\cap_\eta\phi_\eta(I^m) \subseteq \varepsilon X$ where $m = \dim X$. It will be enough to consider each $\phi_\eta(I^m)$ separately, so fix one $\phi = \phi_\eta$ and suppose $\phi(I^m) \subseteq \varepsilon X$. Using Proposition 23 we can find for each $J \subset \{1, \ldots, n\}$ of size $m + 1$ a polynomial $P_J$ of degree $d = \text{poly}_m(g, \log H)$ in the coordinates $(x_i : i \in J)$ vanishing on $X(g, H)$. The zero loci of these polynomials cut out an algebraic variety $V$ of dimension at most $m$ in $\mathbb{R}^n$. Stratify $V$ (e.g. using °o-minimality of $\mathbb{R}^{\text{alg}}$) and let $\{S_i\}$ denote the top dimensional strata and $S'$ the union of the rest. Note $S_i, S' \in \text{poly}_d(D, g, \log H)$ by °o-minimality. The points in $S' \cap X$ are handled by induction on $m$. Now stratify $X \cap S_i$ and denote by $\{B_{i,j}\}$ the top dimensional strata and by $B'$ the union of the rest (over all $S_i$). Note # $\{B_{i,j}\}$ is again $\text{poly}_d(D, g, \log H)$ by °o-minimality. Finally $B' \cap X$ is similarly handled by induction on $m$, while $B_{i,j}$ are by definition basic blocks with semialgebraic closures $S_i$, which finishes the proof.

REFERENCES

[1] G. Binyamini, G. Jones, H. Schmidt, and M. Thomas. Effective Pila-Wilkie in the restricted sub-Pfaffian structure. In preparation, 2021.

[2] Gal Binyamini. Point counting for foliations over number fields. Forum Math. Pi, to appear.

[3] Gal Binyamini and Dmitry Novikov. Wilkie’s conjecture for restricted elementary functions. Ann. of Math. (2), 186(1):237–275, 2017.

[4] Gal Binyamini and Dmitry Novikov. Complex cellular structures. Ann. of Math. (2), 190(1):145–248, 2019.

[5] Gal Binyamini and Dmitry Novikov. The Yomdin-Gromov algebraic lemma revisited. Arnold Math. J., 7(3):199–430, 2021.

[6] Gal Binyamini, Harry Schmidt, and Andrei Yafaev. Lower bounds for Galois orbits of special points on Shimura varieties: a point-counting approach. Math. Ann., to appear.
[7] Gal Binyamini and Nicolai Vorobjov. Effective cylindrical cell decompositions for restricted sub-Pfaffian sets. *Int. Math. Res. Not. IMRN*, 11 2020.

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

[9] Raf Cluckers, Jonathan Pila, and Alex Wilkie. Uniform parameterization of subanalytic sets and diophantine applications. *Ann. Sci. Éc. Norm. Supér. (4)*, 53(1):1–42, 2020.

[10] A. Gabrièlov and N. Vorobjov. Complexity of stratifications of semi-Pfaffian sets. *Discrete Comput. Geom.*, 14(1):71–91, 1995.

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

[12] P. Habegger. Diophantine approximations on definable sets. *Selecta Math. (N.S.)*, 24(2):1633–1675, 2018.

[13] Gareth O. Jones and Margaret E. M. Thomas. Effective Pila-Wilkie bounds for unrestricted Pfaffian surfaces. *Math. Ann.*, 381(1-2):729–767, 2021.

[14] J. Pila. Geometric postulation of a smooth function and the number of rational points. *Duke Math. J.*, 63(2):449–463, 1991.

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

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

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

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

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

[20] Jonathan Pila. Counting rational points on a certain exponential-algebraic surface. *Ann. Inst. Fourier (Grenoble)*, 60(2):489–514, 2010.

[21] Thomas Scanlon. A proof of the Andrè-Oort conjecture via mathematical logic [after Pila, Wilkie and Zannier]. *Astérisque*, 2010/2011(348):Exp. No. 1037, ix, 299–315, 2012. Séminaire Bourbaki: Vol. 2010/2011. Exposés 1027–1042.

[22] Siegfried Van Hille. On a family of mild functions. *Int. J. Number Theory*, 17(6):1379–1390, 2021.

[23] Siegfried Van Hille. Smooth parameterizations of power-subanalytic sets and compositions of Gevrey functions. *Proc. Edinb. Math. Soc. (2)*, 64(3):532–554, 2021.

[24] 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.

[25] A. J. Wilkie. Rational points on definable sets. In *O-minimality and diophantine geometry*, volume 421 of *London Math. Soc. Lecture Note Ser.*, pages 41–65. Cambridge Univ. Press, Cambridge, 2015.

[26] Y. Yomdin. $C^k$-resolution of semialgebraic mappings. Addendum to: “Volume growth and entropy”. *Israel J. Math.*, 57(3):301–317, 1987.

[27] Y. Yomdin. Volume growth and entropy. *Israel J. Math.*, 57(3):285–300, 1987.