

Compression schemes, stable definable families, and o-minimal structures

H. R. Johnson*

Department of Mathematics & CS

John Jay College

CUNY

M. C. Laskowski†

Department of Mathematics

University of Maryland

June 18, 2010

Abstract

We show that any family of sets uniformly definable in an o-minimal structure has an extended compression scheme of size equal to the number of parameters in the defining formula.

As a consequence, the combinatorial complexity (or density) of any

*hujohnson@jjay.cuny.edu

†Partially supported by NSF grant DMS-0600217.

definable family in a structure with a o-minimal theory is bounded by the number of parameters in the defining formula.

Extended compression schemes for uniformly definable families corresponding to stable formulas are also shown to exist.

1 Introduction

Problems concerning the combinatorial complexity of set systems arise in many mathematical disciplines. They appear in context of range-space searching [9] in computational geometry, PAC learning in computational learning theory, and the study of types in mathematical logic. The basic situation of interest is: Given a collection of subsets \mathfrak{C} from some universe X , analyze the relationship of $A \in \mathfrak{C}$ to the finite subsets of X . Warmuth and Littlestone [12] noted a structural characteristic of some set systems \mathfrak{C} which allows this question to be reduced to questions about subsets of X of some uniformly bounded size. They referred to their discovery as a *sample compression scheme*.

Compression schemes have many applications. It has been shown that the existence of a compression scheme on a class of sets is sufficient to ensure PAC learnability of the class [7, 12], and that the size of a compression scheme can serve to replace the VC dimension of a class in PAC sample size bounds [7]. Moreover, there are several kinds of learning machine which operate by using compression sets to make predictions (e.g. [13, 14]).

Warmuth and Littlestone proposed the following definition. Start with a (possibly infinite) set X of elements and a set \mathfrak{C} of subsets of X , which repre-

sent concepts. Warmuth and Littlestone say that \mathfrak{C} admits a d -dimensional compression if, given any finite subset F of X , and any set $A \in \mathfrak{C}$, there is a d -element subset S of F such that the set $A \cap F$ can be recovered from the sets $S \cap A$ and $S \setminus A$. For example, if X is the line of real numbers then the set \mathfrak{C} of open intervals has a 2-dimensional compression. Given any finite set F of real numbers and any open interval I , membership in all of $F \cap I$ can be deduced by knowing the leftmost and rightmost elements of the set $F \cap I$. Arguing similarly, if X denotes \mathbb{R}^2 , the set of ordered pairs of real numbers and \mathfrak{C} denotes the set of open axis-parallel rectangles, then \mathfrak{C} has a 4-dimensional compression.

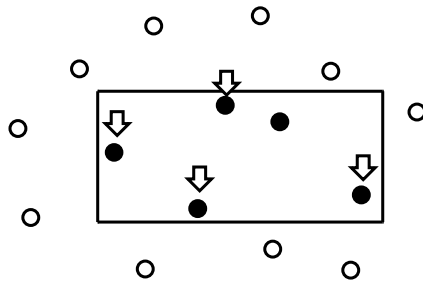


Figure 1: Select the leftmost, rightmost, lowest and highest points included in the rectangle.

While compression schemes are useful, more is needed to achieve closure under the standard set theoretic operations (*i.e.* union, complement, *etc.*). To this end, Floyd and Warmuth [7] proposed the notion of an extended compression scheme with b extra bits. They conceive of storing some uniformly bounded amount of information, represented as a binary string, in addition to $S \cap A$ and $S \setminus A$. From this, as in an ordinary compression scheme, one must recover the trace of A on F . Ben-David and Litman's notion of a *size d -array compression* [4] is a modification of the same idea.

Here we investigate how these notions compare with existing notions from model theory, which is a branch of mathematical logic. To enable this correspondence, it is useful to represent a subset A of X by its characteristic function $f_A : X \rightarrow \{0, 1\}$. so a set of concepts \mathfrak{C} should be thought of as a subset of ${}^X\{0, 1\}$, the set of all characteristic functions with domain X .

For d any integer, X^d denotes the set of (ordered) d -tuples from X and $[X]^d$ (resp. $[X]^{\leq d}$) denotes the d -element (resp. $\leq d$ -element) subsets of X .

For $B \subseteq X$, the notation $\mathfrak{C}|_B$ denotes the set of restrictions $\{f|_B : f \in \mathfrak{C}\}$ and

$$\mathfrak{C}|_{\text{fin}} = \bigcup \{\mathfrak{C}|_B : B \text{ a finite subset of } X \text{ with } |B| \geq 2\}$$

The requirement that $|B| \geq 2$ is technical and is used in the verification of Proposition 2.1. We write $f \sqsubseteq g$ if and only if f is a restriction of g , i.e., $\text{dom}(f) \subseteq \text{dom}(g)$ and $f(x) = g(x)$ for all $x \in \text{dom}(f)$.

Definition Fix $\mathfrak{C} \subseteq {}^X\{0, 1\}$. \mathfrak{C} is said to have an *extended d -compression* if there is a *compression function* $\kappa : \mathfrak{C}|_{\text{fin}} \rightarrow [X]^{\leq d}$ and a finite set \mathcal{R} of *reconstruction functions* $\rho : [X]^{\leq d} \rightarrow {}^X\{0, 1\}$ such that for every $f \in \mathfrak{C}|_{\text{fin}}$

1. $\kappa(f) \subseteq \text{dom}(f)$
2. $f \sqsubseteq \rho(\kappa(f))$ for at least one $\rho \in \mathcal{R}$.

We say that \mathfrak{C} has an *extended d -sequence compression* if there there is a compression function $\kappa : \mathfrak{C}|_{\text{fin}} \rightarrow X^d$ and a finite set \mathcal{R} of *reconstruction functions* $\rho : X^d \rightarrow {}^X\{0, 1\}$ such that for every $f \in \mathfrak{C}|_{\text{fin}}$, $\text{range}(\kappa(f)) \subseteq \text{dom}(f)$, and $f \sqsubseteq \rho(\kappa(f))$ for at least one $\rho \in \mathcal{R}$.

With Proposition 2.1 we will show that for any concept class \mathfrak{C} , the existence of either of these d -compressions is equivalent to having a Floyd-Warmuth extended d -compression with b extra bits for some b .

In model theory, one is concerned with *uniformly definable families* of an algebraic structure. The idea of definability is important in logic, and can be described as follows.

A language L is a collection of function, relation, and constant symbols, together with the familiar logical operators, quantifiers, and variables, which operate according to the axioms of first-order logic. One recursively defines the set of L -formulas. An L -structure \mathcal{M} consists of a nonempty universe M , together with an interpretation of each of the symbols in L . It is straightforward to recursively define *truth in \mathcal{M}* and we write $\mathcal{M} \models \varphi(\bar{a})$ if $\varphi(\bar{a})$ is true in \mathcal{M} .

We generally partition the free variables of a formula into two sequences and write $\varphi(\bar{x}; \bar{y})$ to denote this partition. Intuitively, we think of \bar{x} as the *free variables* and \bar{y} as the *parameters* or *instantiated instances*.

If $\bar{y} = (y_1, \dots, y_m)$ is an ordered set of variables, then we denote the length of \bar{y} by $\text{lg}(\bar{y}) = m$.

Suppose $\varphi(\bar{x}; \bar{y})$ is a partitioned formula with $\text{lg}(\bar{x}) = k$ and $\text{lg}(\bar{y}) = m$. For any $A \subseteq M^k$ and $\bar{b} \in M^m$, we define

$$\varphi(A; \bar{b}) = \{\bar{a} \in A : \mathcal{M} \models \varphi(\bar{a}; \bar{b})\}$$

which is visibly a subset of A . The *uniformly definable family*

$$\mathfrak{C}_{\varphi(\bar{x};\bar{y})} = \{\varphi(M^k; \bar{b}) : \bar{b} \in M^m\}$$

is a set of subsets of M^k .

In this paper, we prove two theorems. The notion of a *stable* formula is defined in section 3.

Theorem 1.1. *If $\varphi(\bar{x};\bar{y})$ is any stable formula in a structure \mathcal{M} , then the uniformly definable family*

$$\mathfrak{C}_{\varphi(\bar{x};\bar{y})} = \{\varphi(M^k; \bar{c}) : \bar{c} \in M^m\}$$

has an extended d -compression for some d .

The proof of 1.1 is an application of a certain kind of *type definition* from model theoretic stability theory. The relationship between type definitions and compression schemes is intimate, as will be made clear in Proposition 3.3.

Theorem 1.2. *If \mathcal{M} is an o-minimal structure and $\varphi(\bar{x};\bar{y})$ is any partitioned formula, then*

$$\mathfrak{C}_{\varphi(\bar{x};\bar{y})} = \{\varphi(M^k; \bar{c}) : \bar{c} \in M^m\}$$

has an extended m -compression, where $m = \lg(\bar{y})$.

Theorem 1.2 generalizes results from Floyd-Warmuth [7] and Basu [3], who proved a similar result for o-minimal expansions of the real field. The

proof follows by an induction on parameters, and makes use of the finitary character of o-minimal systems.

As an example of the utility of these theorems, consider a ‘generalized polynomial’ or ‘fewnomial’ in the sense of Khovanskii [10]

$$p(x; a_1, \dots, a_n, b_1, \dots, b_n) = a_1 x^{b_1} + \dots + a_n x^{b_n}$$

in which the exponents themselves are given by positive real parameters. For each $2n$ -tuple (\bar{a}, \bar{b}) of real numbers, let $A(\bar{a}, \bar{b}) = \{x > 0 : p(x; \bar{a}, \bar{b}) \geq 0\}$, and let $\mathfrak{C} = \{A(\bar{a}, \bar{b}) : (\bar{a}, \bar{b}) \in \mathbb{R}^{2n}\}$. Then Theorem 1.2, together with the fact that the sets $A(\bar{a}, \bar{b})$ form a uniformly definable family in an o-minimal structure, (the real exponential field) imply that \mathfrak{C} has an extended $2n$ -compression. This extends results of Floyd and Warmuth [7], who proved a similar result for polynomials in which the exponents were fixed.

2 Extended compression schemes and combinatorial density

Here we compare our two notions of a compression with preexisting notions and show how their existence relates to the combinatorial complexity of a family of sets.

Proposition 2.1. *Let \mathfrak{C} be any set of subsets of a universe X . For any positive integer d , the following notions are equivalent:*

1. \mathfrak{C} has an extended d -compression;

2. For some b , \mathfrak{C} has a Floyd-Warmuth extended d -compression with b extra bits;
3. \mathfrak{C} has an extended d -sequence compression.

Proof. First, the equivalence of (1) and (2) is immediate. Given an extended d -compression, take any b such that $2^b \geq |\mathcal{R}|$. Use the same compression function κ , and use the b bits to code which reconstruction function to use. Conversely, given a Floyd-Warmuth extended compression, let \mathcal{R} be the set of reconstruction functions coded by the b extra bits.

To obtain the equivalence of (1) and (3), suppose that \mathfrak{C} has an extended d -compression. To show that \mathfrak{C} has an extended d -sequence compression, fix an arbitrary linear ordering $<$ on X . Then form a d -element sequence coding $\kappa(f)$, which is a subset of $\text{dom}(f)$ of size at most d , by writing $\kappa(f)$ in ascending order, repeating the largest element as needed to make the length d , and adjust the reconstruction functions accordingly. Conversely, fix an extended d -sequence compression of \mathfrak{C} . Form an extended d -compression by taking the new compression function to be $\text{range}(\kappa(f))$. Since there are at most d^d length d sequences of these elements, we can replace each of the original reconstruction functions by a set of at most d^d new reconstruction functions coding each of the possibilities. \square

Furthermore, if one allows the d 's to vary, it is routine to show that a set \mathfrak{C} of concepts has an extended d -sequence compression for some d if and only if \mathfrak{C} has a size d' -array compression for some d' in the sense of Ben-David and Litman [4].

We compare the existence of an extended d -compression with the measure of combinatorial density introduced by Assouad in [1].

Definition For $\mathfrak{C} \subseteq {}^X\{0, 1\}$, the *combinatorial density* of \mathfrak{C} , denoted $\text{dens}(\mathfrak{C})$, is a real number defined as the infimum over all $r > 0$ for which there is a number N for which

$$|\mathfrak{C}|_A \leq N \cdot |A|^r$$

for all finite subsets $A \subseteq X$.

One of the primary virtues of VC dimension is that it provides an upper bound for combinatorial density. See, for instance, Chapter 4 of [6]. The following shows that compression schemes can perform a similar function.

Lemma 2.2. *Suppose \mathfrak{C} has an extended d -compression. Then the combinatorial density of \mathfrak{C} is at most d .*

Proof. Fix any finite $A \subseteq X$. Since \mathfrak{C} has an extended d -compression it also has an extended d -sequence compression. Fix one such compression and let N denote the number of reconstruction functions. Then any $g \in \mathfrak{C}|_A$ is uniquely determined by its compression $\kappa(g) \in A^d$ and any reconstruction function $\rho \in \mathcal{R}$ satisfying $g \sqsubseteq \rho(\kappa(g))$. That is, $|\mathfrak{C}|_A \leq N \cdot |A|^d$, so $\text{dens}(\mathfrak{C}) \leq d$. \square

Finally, we call an extended d -compression *consistent* if the image of each reconstruction function ρ is an element of \mathfrak{C} as opposed to an arbitrary function $g : X \rightarrow \{0, 1\}$. Consistent compressions are the subject of Subsection 4.1.

3 Structures and uniformly definable families

We are concerned with finding sufficient conditions on φ and \mathcal{M} which will guarantee that uniformly definable families $\mathfrak{C}_{\varphi(\bar{x};\bar{y})}$ have extended d -compressions.

As notation, given a partitioned formula $\varphi(\bar{x};\bar{y})$, its *dual* is the formula $\varphi^*(\bar{y};\bar{x})$. Formally, the dual φ^* is the same formula as φ , but with the roles of the free variables and parameters reversed. Thus,

$$\mathfrak{C}_{\varphi^*(\bar{y};\bar{x})} = \{\varphi(\bar{a}; M^m) : \bar{a} \in M^k\}$$

is a uniformly definable family of subsets of M^m .

3.1 Stable formulas

There are many equivalent definitions of a partitioned formula $\varphi(\bar{x};\bar{y})$ being stable in a structure \mathcal{M} . Stability of a formula morally means that the parameters can not be used to linearly order arbitrarily large sets of points from the domain. For example, the formula $x < y$ interpreted in any infinite linear order is not stable, whereas $x = y$ is stable in any structure.

We call $\varphi(\bar{x};\bar{y})$ *stable* if, for some integer N , there are no elements $\{(\bar{a}_i, \bar{b}_i) : 1 \leq i \leq N\}$ from $M^k \times M^m$ such that for all $1 \leq i, j \leq N$

$$\mathcal{M} \models \varphi(\bar{a}_i; \bar{b}_j) \quad \text{if and only if} \quad i < j$$

An example of a non-trivial stable formula is $f(\bar{x};\bar{y}) = 0$, where f is a parameterized polynomial, when evaluated in any field $(F, +, \cdot)$. In this case,

$\mathfrak{C}_{f(\bar{x};\bar{y})=0}$ is the parameterized family of zero sets of $f(\bar{x};\bar{a})$ for various \bar{a} from F . The reader may want to verify this fact for the simplest case, $lg(\bar{x}) = 1$, where it follows from the fact that the cardinality of a zero set is uniformly bounded by $\deg(f)$ over all choices of parameters (coefficients).

To distinguish this section from the subsequent one, we should say that o-minimal structures, since they are ordered, always have many unstable formulas. In fact most natural geometric families are unstable. For example, the family of all 2-discs in \mathbb{R}^2 is unstable, because any finite sequence of points in the plane arranged along a line can be linearly ordered (in the sense of stability) by discs emanating from one of the endpoints.

There are many texts giving the salient features of stable formulas. See e.g., [17] (Chapter 1, Lemma 2.1), or [18] for proof of the following easy fact.

Fact 3.1. *For every L -structure \mathcal{M} the set of stable formulas is closed under Boolean combinations and duality. That is, if $\varphi(\bar{x};\bar{y})$ and $\psi(\bar{x};\bar{z})$ are both stable, then so are the dual $\varphi^*(\bar{y};\bar{x})$, $\neg\varphi(\bar{x};\bar{y})$, and $(\varphi \wedge \psi)(\bar{x};\bar{y},\bar{z})$.*

There are many consequences of stability. For our purposes, the most relevant one is the existence of uniform type definitions, which we now introduce.

Definition A formula $\varphi(\bar{x};\bar{y})$ with $lg(\bar{x}) = k$, $lg(\bar{y}) = m$ has a *uniform type definition* if there is an integer $d \geq 1$ and a formula $\psi(\bar{y};\bar{z}_1,\dots,\bar{z}_d)$ with $lg(\bar{z}_i) = m$ for each i such that for all $\bar{a} \in M^k$ and all $B \subseteq M^m$ with more than one element, there is $(\bar{c}_1,\dots,\bar{c}_d) \in B^d$ such that

$$\varphi(\bar{a}; B) = \psi(B; \bar{c}_1, \dots, \bar{c}_d)$$

Thus, the dual formula $\varphi^*(\bar{y}; \bar{x})$ having a uniform type definition means that there is a formula $\psi(\bar{x}; \bar{z}_1, \dots, \bar{z}_d)$ such that for all $\bar{b} \in M^m$ and all $A \subseteq M^k$ with more than one element, there is $(\bar{e}_1, \dots, \bar{e}_d) \in A^d$ such that $\varphi(A^k; \bar{b}) = \psi(A^k; \bar{e}_1, \dots, \bar{e}_d)$.

The following fact is well known. See e.g., Theorem II 2.2 of [18] for a proof.

Fact 3.2. *If $\varphi(\bar{x}; \bar{y})$ is stable then it has a uniform type definition.*

The following Proposition simply amounts to unpacking the definitions.

Proposition 3.3. *If the dual formula $\varphi^*(\bar{y}; \bar{x})$ has a uniform type definition via the formula $\psi(\bar{x}; \bar{z}_1, \dots, \bar{z}_d)$, then the uniformly definable family $\mathfrak{C}_{\varphi(\bar{x}; \bar{y})}$ has an extended d -compression.*

Proof. We will show that $\mathfrak{C}_{\varphi(\bar{x}; \bar{y})}$ has an extended d -sequence compression. Fix $A \subseteq M^k$ finite. For each $\bar{b} \in M^m$, let $\chi_{\varphi(A; \bar{b})} : A \rightarrow \{0, 1\}$ denote the characteristic function of $\varphi(A; \bar{b})$. Define the compression function κ to send $\chi_{\varphi(A; \bar{b})}$ to any $(\bar{e}_1, \dots, \bar{e}_d) \in A^d$ such that $\varphi(A^k; \bar{b}) = \psi(A^k; \bar{e}_1, \dots, \bar{e}_d)$. Then define the unique reconstruction function ρ by taking $\rho(\bar{e}_1, \dots, \bar{e}_d)$ to be the characteristic function of $\psi(A^k; \bar{e}_1, \dots, \bar{e}_d)$. \square

Theorem 3.4. *For any structure \mathcal{M} and any stable formula $\varphi(\bar{x}; \bar{y})$, the uniformly definable family $\mathfrak{C}_{\varphi(\bar{x}; \bar{y})} = \{\varphi(M^k; \bar{b}) : \bar{b} \in M^m\}$ of subsets of M^k has an extended compression scheme.*

Proof. By Fact 3.1 the dual formula $\varphi^*(\bar{y}; \bar{x})$ is stable, hence has a uniform type definition by Fact 3.2. Thus $\mathfrak{C}_{\varphi(\bar{x}; \bar{y})}$ has an extended compression scheme by Proposition 3.3. \square

There is no hope of relating the size of a compression scheme of a stable formula to the number of its parameters. For example, for any d , let \mathcal{M}_d be a structure in the language $L = \{U, R\}$, in which $U^{\mathcal{M}_d}$ is interpreted as an infinite set X , $(\neg U)^{\mathcal{M}_d}$ is interpreted as $[X]^d$, and $R(x, y)$ is interpreted as membership $x \in y$. Then $R(x, y)$ is a stable formula with a d -compression, but no extended d' -compression for any $d' < d$.

It may be the case that the size of an extended compression on a stable formula can be bounded by some other combinatorial characteristic. A leading candidate would be the Shelah 2-rank of the formula.

4 O-minimal structures

Definition Let $L = \{<, \dots\}$ and let \mathcal{M} be an L -structure. We say that \mathcal{M} is *o-minimal* if $<$ is interpreted as a dense linear order without endpoints, and for every partitioned L -formula $\psi(x; \bar{y})$ in a single variable x , the set $\psi(\mathcal{M}; \bar{c})$ is a finite union of points and intervals for every $\bar{c} \in M^m$.

An excellent reference for o-minimality is [5]. The most important o-minimal structures for applications are expansions of the ordered real numbers, but there are other examples, in particular various structures built on the rationals.

It follows from the Tarski-Seidenberg theorem that the real field is o-minimal. Two other important cases are due to results by Wilkie and Gabrielov. They are:

- The real field with a function symbol added for exponentiation [21].

- The real field with symbols added for the analytic functions, restricted respectively to the unit cube of appropriate dimension. [22, 8].

Since the real field with exponentiation is o-minimal, our result shows that any feedforward sigmoidal neural network with standard sigmoid has an extended compression scheme of size w , where w is the number of weights. Similarly, concept classes associated with sets of positivity of exponential expressions such as

$$e^{e^{x_1 y_1 + x_2^2}} + x_1^2 e^{y_2} \geq 0$$

will be subject to Corollary 4.5, as will any “exponential polynomial” or fewnomial in the sense of Wilkie and Khovanskii [16, 21, 10]. It follows from [4], as well as from our result, that any family of real algebraic hypersurfaces defined by polynomials with k non-zero monomials admits an extended k -compression.

Note that if \mathcal{M} is o-minimal and $\varphi(x; \bar{y})$ is any partitioned formula with a single x -variable, then the boundary of $\varphi(M; \bar{c})$, which we denote by $\partial\varphi(M; \bar{c})$, is a finite subset of M . A proof of the following Fact is given in Chapter 3 of [5].

Fact 4.1. *Suppose \mathcal{M} is o-minimal and $\psi(x; \bar{y})$ is any partitioned formula with a single x -variable. Then there is some integer N such that for every $\bar{c} \in M^m$, the boundary $\partial\psi(M; \bar{c})$ of the set $\psi(M; \bar{c})$ consists of fewer than N points.*

Lemma 4.2. *For any o-minimal structure M and for any formula $\theta(\bar{x}; y, \bar{z})$ with $\text{lg}(\bar{x}) = k$ and $\text{lg}(y) = 1$, there is a finite set \mathcal{F}_θ of formulas, each of the*

form $\psi(\bar{x}; \bar{w}, \bar{z})$ with $\text{lg}(\bar{w}) = k$, such that for every finite $A \subseteq M^k$, $c \in M$ and $\bar{e} \in M^{\text{lg}(\bar{z})}$ there is $\bar{a} \in A$ and $\psi \in \mathcal{F}_\theta$ such that $\theta(A; c, \bar{e}) = \psi(A; \bar{a}, \bar{e})$.

Proof. Fix an o-minimal structure \mathcal{M} and a formula $\theta(\bar{x}; y, \bar{z})$. By o-minimality, choose an integer N so that the boundary $\partial\theta(\bar{a}, M, \bar{e})$ has size at most N for all choices of \bar{a} and \bar{e} . For each $i < N$, let $f_i(\bar{a}, \bar{e})$ denote the i^{th} boundary point (with respect to the ordering on M) of $\partial\theta(\bar{a}, M, \bar{e})$ if it exists. Note that each partial function f_i is definable in \mathcal{M} . We define four groups of formulas:

$$\begin{aligned} \psi_{i,1}(\bar{x}; \bar{w}, \bar{z}) &:= \theta(\bar{x}; f_i(\bar{w}, \bar{z}), \bar{z}) \\ \psi_{i,2}(\bar{x}; \bar{w}, \bar{z}) &:= \forall u[u < f_i(\bar{w}, \bar{z}) \rightarrow \exists v(u < v < f_i(\bar{w}, \bar{z}) \wedge \theta(\bar{x}; v, \bar{z}))] \\ \psi_{i,3}(\bar{x}; \bar{w}, \bar{z}) &:= \forall u[u > f_i(\bar{w}, \bar{z}) \rightarrow \theta(\bar{x}; u, \bar{z})] \\ \psi^*(\bar{x}; \bar{w}, \bar{z}) &:= \forall u\theta(\bar{x}; u, \bar{z}) \end{aligned}$$

Let \mathcal{F}_θ be this finite set of formulas.

Now choose a finite $A \subseteq M^k$, $c \in M$, and $\bar{e} \in M^{\text{lg}(\bar{z})}$. Let

$$D = \bigcup_{\bar{a} \in A} \partial\theta(\bar{a}, M, \bar{e})$$

Since A is finite, D is a finite subset of M . The argument now splits into four cases:

Case 1. $c \in D$.

Then $\theta(A; c, \bar{e}) = \psi_{i,1}(A; \bar{a}, \bar{e})$ for any $\bar{a} \in A$ and $i < N$ such that $c = f_i(\bar{a}, \bar{e})$.

Case 2. $c \notin D$, but $c < d$ for some $d \in D$.

Since D is finite, we can choose $d^* \in D$ least such that $c < d^*$. By definition of D , for any $\bar{a} \in A$, the truth of $\theta(\bar{a}, y, \bar{e})$ is invariant for y 's taken from the half-open interval $[c, d^*)$. Thus $\theta(A; c, \bar{e}) = \psi_{i,2}(A; \bar{a}, \bar{e})$ for any $\bar{a} \in A$ and $i < N$ such that $f_i(\bar{a}, \bar{e}) = d^*$.

Case 3. $D \neq \emptyset$ but $c > d$ for every $d \in D$.

Let d^* be the maximum element in D . For every $\bar{a} \in A$, the truth of $\theta(\bar{a}, y, \bar{e})$ is invariant on the open interval (d^*, ∞) , so $\theta(A; c, \bar{e}) = \psi_{i,3}(A; \bar{a}, \bar{e})$ for any $\bar{a} \in A$ and $i < N$ such that $f_i(\bar{a}, \bar{e}) = d^*$.

Case 4. $D = \emptyset$.

In this case, for any $\bar{a} \in A$, the truth of $\theta(\bar{a}, y, \bar{e})$ is invariant on all of M , hence $\varphi(A; c, \bar{e}) = \psi^*(A; \bar{a}, \bar{e})$ for any \bar{a} from A . \square

Note that there was no constraint on $\text{lg}(\bar{z})$ in the previous Lemma. Thus, if we are given a formula $\varphi(\bar{x}; \bar{y})$ with $\text{lg}(\bar{y}) = m$, we can inductively shave off elements from \bar{y} in favor of elements from A . More precisely, we have the following Proposition.

Proposition 4.3. *Let \mathcal{M} be any o-minimal structure. For all formulas $\varphi(\bar{x}; \bar{y})$ with $\text{lg}(\bar{x}) = k$ and $\text{lg}(\bar{y}) = m$, and for all $1 \leq r \leq m$, there is a finite set \mathcal{F}_φ^r of formulas, each of the form $\psi(\bar{x}; y_1, \dots, y_{m-r}, \bar{w}_1, \dots, \bar{w}_r)$, such that for all finite $A \subseteq M^k$ and all $\bar{c} \in M^m$, there are $(\bar{a}_1, \dots, \bar{a}_r) \in A^r$ and $\psi \in \mathcal{F}_\varphi^r$ such that $\varphi(A; \bar{c}) = \psi(A; c_1, \dots, c_{m-r}, \bar{a}_1, \dots, \bar{a}_r)$.*

Proof. By induction on $r \leq m$. When $r = 1$ this follows immediately from Lemma 4.2 applied to φ . Assuming the result holds for $r < m$, in order to

establish it for $r + 1$ one simply applies Lemma 4.2 to each of the formulas $\psi \in \mathcal{F}_\varphi^r$ and taking $\mathcal{F}_\varphi^{r+1} = \bigcup_{\psi \in \mathcal{F}_\varphi^r} \mathcal{F}_\psi$. \square

By taking $r = m$ in the previous Proposition we obtain the following theorem.

Theorem 4.4. *Let \mathcal{M} be any o-minimal structure. For all formulas $\varphi(\bar{x}; \bar{y})$ with $\text{lg}(\bar{x}) = k$ and $\text{lg}(\bar{y}) = m$, there is a finite set \mathcal{F} of formulas, each of the form $\psi(\bar{x}; \bar{w}_1, \dots, \bar{w}_m)$, such that for all finite $A \subseteq M^k$ and all $\bar{c} \in M^m$, there is $(\bar{a}_1, \dots, \bar{a}_m) \in A^m$ and $\psi \in \mathcal{F}$ such that $\varphi(A; \bar{c}) = \psi(A; \bar{a}_1, \dots, \bar{a}_m)$.*

Corollary 4.5. *For \mathcal{M} any o-minimal structure, every uniformly definable family $\mathfrak{C}_{\varphi(\bar{x}; \bar{y})}$ has an extended m -compression, where $m = \text{lg}(\bar{y})$.*

Proof. By Proposition 2.1 it suffices to exhibit an extended m -sequence compression. Let \mathcal{F} be the finite set of formulas from Theorem 4.4. Fix a finite $A \subseteq M^k$. Define the compression function by

$$\kappa(\chi_{\varphi(A; \bar{c})}) = (\bar{a}_1, \dots, \bar{a}_m)$$

where $(\bar{a}_1, \dots, \bar{a}_m) \in A^m$ is found via Theorem 4.4. Take $\mathcal{R} = \{\rho_\psi : \psi \in \mathcal{F}\}$, where $\rho_\psi(\bar{a}_1, \dots, \bar{a}_m) = \chi_{\psi(A; \bar{a}_1, \dots, \bar{a}_m)}$. \square

The next Corollary follows immediately from Corollary 4.5 and Lemma 2.2. It generalizes a result of Basu [3], which makes a similar assertion for o-minimal expansions of real closed fields.

To see the relation between our result and that of Basu, note that Basu considers arrangements of n objects, thought of as the fibers of some fixed

definable map $\pi : T \rightarrow \mathbb{R}^l$ where T is a definable subset of \mathbb{R}^{l+k} . He concludes that the combinatorial complexity of any such finite arrangement (*i.e.* the number of cells) is $O(n^k)$. (This is phrased as a result on the Betti numbers of the cells in Theorem 2.2 of [3]). This is dual to our result on density, though we phrase things in terms of relations rather than projections. Fundamentally, the number of subsets traced on a set of points, and the number of cells given by the fibers of these points under a projection, are the same.

Our result is therefore equivalent, but has a slightly increased generality stemming from the fact that we do not assume (as Basu does) that our o-minimal structure is an expansion of a real closed field.

Corollary 4.6. *For \mathcal{M} any o-minimal structure, every uniformly definable family $\mathfrak{C}_{\varphi(\bar{x};\bar{y})}$ has combinatorial density at most $\lg(\bar{y})$.*

4.1 Consistent compressions

If one is willing to double the size of the data set, then in any o-minimal structure \mathcal{M} , every uniformly definable family $\mathfrak{C}_{\varphi(\bar{x};\bar{y})}$ has a *consistent* d -compression, where $d = 2 \cdot \lg(\bar{y})$. That is, the range of the reconstruction functions can be taken to be $\{\chi_{\varphi(A;\bar{c})} : \bar{c} \in M^{\lg(\bar{y})}\}$.

The proof is analogous to the series of statements 4.2–4.5 given above.

Lemma 4.7. *For any o-minimal structure M and for any formula $\theta(\bar{x}; y, \bar{z})$ with $\lg(\bar{x}) = k$ and $\lg(y) = 1$, there is a finite set \mathcal{F}_θ of formulas, each of the form $\psi(y; \bar{w}_1, \bar{w}_2, \bar{z})$ with $\lg(\bar{w}_1) = \lg(\bar{w}_2) = k$ such that for every finite $A \subseteq M^k$, $c \in M$ and $\bar{e} \in M^{\lg(\bar{z})}$ there is $(\bar{a}_1, \bar{a}_2) \in A^2$ and $\psi \in \mathcal{F}_\theta$ such that*

1. *The set $\psi(M; \bar{a}_1, \bar{a}_2, \bar{e})$ is connected *i.e.*, an interval*

2. $c \in \psi(M; \bar{a}_1, \bar{a}_2, \bar{e})$ and

3. $\theta(A; c, \bar{e}) = \theta(A; d, \bar{e})$ for every $d \in \psi(M; \bar{a}_1, \bar{a}_2, \bar{e})$.

Proof. The proof is similar to that of Lemma 4.2. Fix \mathcal{M} , θ , the integer N , and definable functions f_i as in the proof of 4.2.

Here we define five groups of formulas, each with free variables among $(y, \bar{w}_1, \bar{w}_2, \bar{z})$:

$$\psi_{i,1} := y = f_i(\bar{w}_1, \bar{z})$$

$$\psi_2 := y = y \quad \text{i.e., always true}$$

$$\psi_{i,3} := y < f_i(\bar{w}_1, \bar{z})$$

$$\psi_{i,4} := y > f_i(\bar{w}_1, \bar{z})$$

$$\psi_{i,j,5} := f_i(\bar{w}_1, \bar{z}) < y < f_j(\bar{w}_2, \bar{z})$$

Let \mathcal{F}_θ be this finite set of formulas.

As before, choose a finite $A \subseteq M^k$, $c \in M$, and $\bar{e} \in M^{\text{lg}(\bar{z})}$. Let

$$D = \bigcup_{\bar{a} \in A} \partial\theta(\bar{a}, M, \bar{e})$$

Since A is finite, D is a finite subset of M . Here the argument now splits into five cases:

Case 1. $c \in D$.

Then $\psi_{i,1}(M; \bar{a}, \bar{e})$ works for any $\bar{a} \in A$ and $i < N$ such that $c = f_i(\bar{a}, \bar{e})$.

Case 2. $D = \emptyset$.

Then the always true formula $y = y$ suffices for ψ .

Case 3. $D \neq \emptyset$ but $c < d$ for every $d \in D$.

Let d^* be the smallest element of D . Then $\psi_{i,3}(M; \bar{a}, \bar{e})$ works for any $\bar{a} \in A$ and $i < N$ such that $f_i(\bar{a}, \bar{e}) = d^*$.

Case 4. $D \neq \emptyset$ but $c > d$ for every $d \in D$.

Let d^* be the maximum element in D . Take $\psi_{i,4}(M; \bar{a}, \bar{e})$ for any $\bar{a} \in A$ and $i < N$ such that $f_i(\bar{a}, \bar{e}) = d^*$.

Case 5. $c \notin D$, but there are elements of D above and below c .

In this case, let d^- be the maximum element of D below c and d^+ be the minimum element of D above c . Take $\psi_{i,j,5}(M; \bar{a}_1, \bar{a}_2, \bar{e})$ for any \bar{a}_1 such that $d^- = f_i(\bar{a}_1, \bar{e})$ and any \bar{a}_2 so that $d^+ = f_j(\bar{a}_2, \bar{e})$. \square

As in the proof of Proposition 4.3, a routine induction yields the following Proposition for any m and any $1 \leq r \leq m$.

Proposition 4.8. *Let \mathcal{M} be any o-minimal structure. For all formulas $\varphi(\bar{x}; \bar{y})$ with $\text{lg}(\bar{x}) = k$ and $\text{lg}(\bar{y}) = m$, and for all $1 \leq r \leq m$, there is a finite set \mathcal{F}_φ^r of formulas, each of the form $\psi(y_1, \dots, y_r; \bar{w}_1, \dots, \bar{w}_{2r})$, such that for all finite $A \subseteq M^k$ and all $\bar{c} \in M^m$, there are $(\bar{a}_1, \dots, \bar{a}_{2r}) \in A^{2r}$ and $\psi \in \mathcal{F}_\varphi^r$ such that the set $B = \psi(M^r; \bar{a}_1, \dots, \bar{a}_{2r})$ satisfies*

1. B is an r -dimensional box i.e., an r -fold product of intervals
2. $(c_{m-r+1}, \dots, c_m) \in B$ and
3. $\varphi(A; c_1, \dots, c_{m-r}, \bar{d}) = \varphi(A; c_1, \dots, c_{m-r}, \bar{d}')$ for all $\bar{d}, \bar{d}' \in B$.

Taking $r = m$ yields

Theorem 4.9. *Let \mathcal{M} be any o-minimal structure. For all formulas $\varphi(\bar{x}; \bar{y})$ with $\text{lg}(\bar{x}) = k$ and $\text{lg}(\bar{y}) = m$, there is a finite set \mathcal{F} of formulas, each of the form $\psi(\bar{y}; \bar{w}_1, \dots, \bar{w}_{2m})$ such that for all finite $A \subseteq M^k$ and all $\bar{c} \in M^m$ there are $(\bar{a}_1, \dots, \bar{a}_{2m}) \in A^{2m}$ and $\psi \in \mathcal{F}$ such that the set $B = \psi(M^m; \bar{a}_1, \dots, \bar{a}_{2m})$ is an m -dimensional box containing \bar{c} on which the set $\varphi(A; \bar{b})$ is invariant among $\bar{b} \in B$.*

Corollary 4.10. *Let \mathcal{M} be any o-minimal structure. For all formulas $\varphi(\bar{x}; \bar{y})$ with $\text{lg}(\bar{y}) = m$, $\mathfrak{C}_{\varphi(\bar{x}; \bar{y})}$ has a consistent extended $2m$ -compression scheme.*

Proof. As in the proof of Corollary 4.5 it suffices to produce a consistent extended $2m$ -sequence compression. Let \mathcal{F} be the finite set of formulas given by Theorem 4.9. Let \mathcal{R} be the set reconstruction functions indexed by $\psi \in \mathcal{F}$, where

$$\rho_{\psi}(\bar{a}_1, \dots, \bar{a}_{2m}) = \chi_{\varphi(A; \bar{d})}$$

for some $\bar{d} \in \psi(M^m; \bar{a}_1, \dots, \bar{a}_{2m})$.

Given any finite $A \subseteq M^k$ and any $\bar{c} \in M^m$, the compression function is defined by

$$\kappa(\chi_{\varphi(A; \bar{c})}) = (\bar{a}_1, \dots, \bar{a}_{2m})$$

where $(\bar{a}_1, \dots, \bar{a}_{2m})$ are obtained by an application of Theorem 4.9. \square

5 Discussion

Results in [11] increased the number of known VC classes by considering parallel results in mathematical logic concerning independence dimension. Here we have sought to make an analogous connection between compression

schemes and type definability. The matching in this case is less symmetric, but interesting nonetheless.

Warmuth and Littlestone conjectured, in their original unpublished manuscript on compression schemes, that all set systems with finite VC dimension admit an extended compression scheme of some finite size. In this paper we have confirmed that all set systems which are stable and all those definable in o-minimal structures have extended compressions. It is of longstanding interest to determine whether, in fact, every class of finite VC dimension has an extended compression scheme, or, in model theoretic language, whether the theories with uniform type definitions over finite sets are exactly those which are dependent.

It is also of interest to relate the size of the extended compression, both its dimension and the number of reconstruction functions to other metrics of the class. In the stable case, the Shelah 2-rank [18] is the obvious possibility.

Our results depend modestly but necessarily on appeals to the compactness theorem. Work remains to be done to translate our results into a workable method in which both $|\mathcal{R}|$ and the size of the compression are managed. This will clearly have to be done on a situational basis, but many of the pieces already exist for the real field (Milnor bounds) and the real field with exponentiation (Khovanskii's fewnomial bounds). One can also approach compression schemes in o-minimal structures from the point of view of finite cell decomposition.

Most previous work on compression schemes has relied on the nice properties of set systems which are maximum, which means that on any finite set of points, they trace out the maximum number of sets allowed by Sauer's

lemma. Previous research (*e.g.* [7, 4]) has shown that definable families which are natural (intervals, half-spaces, algebraic sets) have a tendency to be maximum, or nearly so. Our work has made no use of the maximum condition, but the link between maximum families and the families we have considered is largely unexplored. It may be the case, for instance, that any family definable in an o-minimal structure at some level embeds in a maximum family. Ben-David and Litman showed a similar fact for algebraic sets.

References

- [1] P. Assouad, Densité et dimension. *Annales de l'institut Fourier*, 33 no. 3, (1983), 233–282.
- [2] J. Baldwin, *Fundamentals of Stability Theory*, Springer-Verlag, New York, 1988.
- [3] S. Basu, Combinatorial complexity in o-minimal geometry, *Proc. of the 39th ann. ACM symp. on Theory of Computing*, (2007), 47 – 56.
- [4] S. Ben-David, A. Litman, Combinatorial variability of Vapnik-Chervonenkis classes with applications to sample compression schemes, *Discrete Applied Mathematics* 86(1), (1998), 3-25.
- [5] L. van den Dries, *Tame Topology and O-minimal Structures*, Number 248 in London Mathematical Society Lecture Notes Series, Cambridge University Press, Cambridge, 1998.

- [6] Dudley, R.M., *Uniform Central Limit Theorems*, Cambridge University Press, New York, 1999.
- [7] S. Floyd, and F. Warmuth, Sample compression, learnability, and the Vapnik-Chervonenkis dimension, *Machine Learning*, Vol.21 (3), (1995), 269–304.
- [8] A. Gabrielov, Projections of semi-analytic sets, *Functional Anal. Appl.* 2 (1968), 282-291.
- [9] D. Haussler, E. Welzl, ϵ -Nets and Simplex Range Queries. *Discrete & Computational Geometry* 2, (1987), 127-151.
- [10] A. Hovanskii, On a class of systems of transcendental equations, *Soviet Math. Dokl.* 22, (1980), 762–765.
- [11] M. C. Laskowski, Vapnik-Chervonenkis classes of definable sets, *J. London Math. Soc.* 45 (2), (1992), 377-384.
- [12] N. Littlestone and F. Warmuth, Unpublished notes.
- [13] M. Marchland and J. Shaw-Taylor. The Set Covering Machine. *Journal of Machine Learning Research*, 3, (2002), 723-746.
- [14] M. Sokolova, M. Marchland, N. Japkowicz, and J. Shaw-Taylor. The Decision List Machine. *Advances in Neural Information Processing Systems*, 15, (2003), 921-928.
- [15] D. Marker, *Model Theory: An Introduction*, Springer, New York, 2002.

- [16] D. Marker, Model theory and exponentiation, *Notices Amer. Math. Soc.* 43 (1996), 753–759.
- [17] A. Pillay, *Geometric Stability Theory*, Oxford University Press, 1996.
- [18] S. Shelah, *Classification theory*, Second edition, North Holland, 1990.
- [19] A. Tarski, *A Decision method for Elementary Algebra and Geometry*, second edition revised, Rand Corporation, Berkeley and Los Angeles, 1951.
- [20] Vapnik, V., Chervonenkis, A., On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and its Applications*, 16(2), (1971), 264–280.
- [21] A. J. Wilkie, Model completeness results for expansions of the real field by restricted Pfaffian functions and the exponential function, *J. Amer. Math. Soc.* 9, (1996), 1051–1094.
- [22] A. J. Wilkie, A theorem of the complement and some new o-minimal structures, *Selecta Mathematica, New Series*, Vol. 5, No. 4, Dec (1999), 397–421.