ABSTRACT

Title of Dissertation: Definable families of finite Vapnik Chervonenkis dimension

Hunter R. Johnson, Doctor of Philosophy, 2008

Dissertation directed by: Professor M. C. Laskowski
Department of Mathematics

Vapnik Chervonenkis dimension is a basic combinatorial notion with applications in machine learning, stability theory, and statistics. We explore what effect model theoretic structure has on the VC dimension of formulas, considered as parameterized families of sets, with respect to long disjunctions and conjunctions. If the growth in VC dimension is linear in the number of disjunctions, then the theory under consideration has a certain kind of good structure. We have found a general class of theories in which this structure obtains, as well as situations where it fails.

We relate "compression schemes" of computational learning theory to model theoretic type definitions, and explore the model theoretic implications. All stable definable families are shown to have finite compression schemes, with specific bounds in the case of NFCP theories.

Notions of maximality in VC classes are discussed, and classified according to their first order properties. While maximum classes can be characterized in first-order logic, maximal classes can not.

Definable families of finite Vapnik Chervonenkis dimension

by

Hunter R. Johnson

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Professor M. C. Laskowski, Chairman/Advisor
Professor David W. Kueker
Professor Edgar G.K. Lopez-Escobar
Professor James A. Schafer
Professor William I. Gasarch

DEDICATION

For my father, Jim Johnson,

and my mother, Marchetta Johnson

TABLE OF CONTENTS

LIST OF FIGURES

0.1 Preface

In [38], two Russian probabilists made a discovery which relates certain desirable behavior in a family of sets[1] to a combinatorial property of the sets. This is now known as Vapnik Chervonenkis (VC) dimension in their honor. In one of the strange coincidences in the history of mathematics, their discovery was echoed by independent papers from several authors shortly afterward [33, 31].

The idea of Vapnik Chervonenkis dimension has turned out to have implications for almost all disciplines in which parameterized families of sets are studied. This ranges from modern model theory, to empirical processes, theoretical probability, computational learning theory, and combinatorics.[2]

Recently the term "VC theory" has been introduced to refer to questions involving VC dimension; several articles have been produced which study VC dimension in its own right [1, 26]. There is no treatise on the subject, however, which joins the results in various fields into an organic presentation. This paper is certainly not such a work, but it does seek to relate at least a few of the above sets of practices: those of mathematical logic, computational learning theory, and combinatorics.

There has been some success in this line already. In [23], after the central role of VC dimension in Valiant's model of PAC learning had been realized, the observation was made that VC dimension was already a well-studied phenomenon in logic, under the guise of the Independence Property. This immediately answered

[1] In the case of Vapnik and Chervonenkis, this was a uniform convergence of relative frequency to probability. Other interesting conditions are also implied by finite VC dimension, such as a law of large numbers, and a central limit theorem on boolean valued functions.

[2] See, respectively, [1, 10, 38, 5, 26].

at least one interesting question, whether feedforward sigmoidal neural networks always have finite VC dimension [19]. More generally, the laborious quest of specific arguments to show finite VC dimension in certain set systems was no longer necessary, since most interesting systems were already known to be definable in so-called dependent theories.[3]

Unfortunately, while the existing logical theory provides proofs of finitude, it does not give reasonable bounds on VC dimension. Most theories known to be dependent have been shown to be so through a "sufficiency of one variable" argument [23, 33]. This means that every formula in a theory has finite VC dimension iff the statement is true for every formula in a single variable. Since the one dimensional definable families are frequently more simple[4] than the definable families in higher dimensions, this gives a straightforward route for proofs of dependence. Unfortunately the collapse to a single variable invokes Ramsey's theorem, and the resulting bounds on complexity, while finite, are large.

As an attempt to establish practical bounds, we try to find maps from syntactical complexity to VC dimension. In an ideal world, tight bounds on VC dimension of formulas would follow simply by knowing the VC dimension of the atomic formulas, and the syntactical composition of the formula in question. Good mappings of this type, for specific situations, can be found in [19, 16, 34, 11, 5], and in many other papers.

For the sake of simplicity, we consider a single parameterized formula, and analyze changes in VC dimension under long disjunctions and conjunctions. The

---

[3]These are the theories in which all definable parameterized set systems have finite VC dimension.

[4]In o-minimal, weakly o-minimal and strongly minimal formulas this is particularly true.

conjunctions and disjunctions fix the variables, but the number of parameters increases linearly with the length of the composition; every new conjunction brings an independent set of new parameters for the original formula. The papers cited above, as we will see, show that the growth in VC dimension in such a situation is at worst log-linear. For many years it was unclear whether the log-linear upper bound ever held, but [12] demonstrated that this is indeed the case.

This essentially unique counter-example is interesting for two reasons. Firstly, the family produced is stable, following from the fact that the elements in the family are almost disjoint.[5] Secondly, aside from the almost disjointness, the family has very little structure, having been shown to exist by a probabilistic method.

In most applications, the log-linear bound is used because of its generality [16, 5, 26], and possibly because the performance difference between linear and log-linear growth is relatively unsubstantial. This leaves open the question, however, whether there is a qualitative dichotomy between structures in which the log-linear bound is possible, and those in which it is not. Here we give the first example of a non-trivial theory in which the log-linear bound is not attainable, and where in fact growth in VC dimension (in the above sense) is linearly bounded. A plausible goal is to characterize (algebraically, or model theoretically) the division between the two types of behavior. In other words:

- What are some algebraic factors which determine when a set system combines with itself in a simple way, with respect to VC dimension?

In exploring this question, we show that for long disjunctions of the above

[5] The size of intersection is uniformly bounded by a natural number.

type, the independence dimension (dual VC dimension) is linear in the number of parameters. These results constitute Chapter 2.

0.1.1

The remaining two chapters do not answer questions with such a long history. We first discuss Chapter 3.

In model theory, a *type* is the signature, with respect to first-order formulas, of an element of the universe of some structure. The type of an element gives a complete description of the first-order relationships between the element and the other elements in the universe. The compactness theorem makes it possible to answer many questions about types by considering not all formulas, but rather instances of a single formula, where *instance* means that the only difference between two formulas in the type is a reassignment of parameters, and perhaps a negation. When restricted to a formula, the information contained in a type amounts to a specification of which tuples of the universe provide parameters corresponding to positive (or negative) instances in the type. The type of an element, in this light, is a subset of the universe. If we fix the restricting formula and consider other element types, we get a family of subsets. In other words, we get a concept class, or set system.

It has long been observed [32] that in the presence of strong structure, the information contained in a type, restricted to a formula, has a finite description, and in fact this description can be given in terms of the restricting formula. There is a name and a large body of work devoted to such strong structures: They are called *stable*, and the study of them is referred to as Stability Theory.

A similar, but different, phenomenon can be observed in collections of geometric objects. Consider the set of all rectangles in the plane, with sides parallel to one of the two cartesian axes.[6] If one considers a rectangle from this class and its relationship with a finite set of points, the properties of the rectangle (from the point of view of the finite set) can be captured by considering a set of only four points, which describe the boundary of the rectangle. Similar "compressions" can be found, in all dimensions, for balls, half-spaces, and even semi-algebraic sets.[7] In each case, a uniformly bounded amount of finite information is sufficient to describe the boundary of an object, on any finite set, however large.

These compressions are known as *compression schemes*, and were presented in [24], then later developed in [13, 4, 14, 22]. A small compression scheme on a set system of finite VC dimension can give better bounds on the number of examples needed for PAC learning than VC dimension alone, and also allows concepts to be learned in a space-bounded manner [13, 14].

The set of rectangles, as above, can be described uniformly by a single formula. In fact compression schemes are operations on types, just as type definitions are. In the rectangular case, the "element" whose type is considered is a tuple determining a certain rectangle (say, two diagonal corners), the restricting formula is the formula for rectangles, and the type gives a description of the relationship between the tuple and all points in the plane, *vis a vis* the "rectangle" relation. The picture of a rectangle in the plane in this sense *is* a type.

This suggests that the existence of a type compression (or definition) has

[6] See Figure A.1.

[7] There is a proviso here that the discrete set is in "general position."

a certain relation with the existence a geometric compression scheme on the same type. The main theorem in Chapter 3 establishes this formally. The class of "geometrically compressible" set systems of finite VC dimension is thereby extended to include all stable families. An account of the non-stable geometrically compressible families can be found in [4].

An interesting question is raised: Are dependent theories characterized by the existence of compression schemes, in a way similar to the characterization of stable theories in terms of type definitions?

0.1.2

Chapter 4 explores a maximality condition on a set system in a first order context. For any set system of a certain VC dimension, we can imagine adding new sets until it is not possible to add more sets without increasing the VC dimension. Sauer's lemma (p. 20) establishes that the size of a class of VC dimension d on a finite domain of size n is always bounded by a polynomial in n of degree d. Many natural classes attain this maximum size. On the other hand, it is possible to reach a "dead end" before the class is maximum, in which case the class is traditionally said to be maximal.[8] In the infinite case, a class is maximum if it is maximum when restricted to any finite domain.

For many interesting questions about VC dimension, it suffices to consider only the maximal examples.[9] While it is clear that the property of being maximum is first-order, the property of being maximal is less clear. If a class is not maximum, there must be a local reason, but for maximality this is not the

[8]See the footnote on page 23.

[9]For example, the existence of compression schemes.

case. The argument in Chapter 4 gives a formal argument that there is no first order sentence expressing the maximality of a definable family. We also show that maximality for any *definable* family is in fact a very strong condition, which never holds many natural situations. This partly explains a previously remarked upon absence of natural examples of maximal families.

Chapter 1

Introduction

1.1 Concept classes and Stone spaces

There has been a trend in the history of VC dimension of authors in different disciplines independently discovering the same results in different contexts. As a result, there has not much standardization of notation. Along with different notations, there are concepts in the literature of VC dimension which can be morally identified with the same concept. There are hypergraphs in graph theory, range spaces and concept classes in computer science, definable families in mathematical logic, and other notions. They are all ways of describing, albeit with different emphases, an abstract set system, that is, a set X together with some $\mathfrak{C} \subseteq \mathcal{P}(X)$.[1]

Here we adopt arguably the broadest possible notation for studying VC dimension, the formulas of first-order logic. In certain situations, however, logical notation carries an extra burden of complication, and we occasionally use other descriptions. We try to make the equivalence between these different ways of describing clear.

[1] This is the power set, $\mathcal{P}(X) = \{A : A \subseteq X\}$.

In the following definition, and throughout, we use ω to denote the set of all finite ordinal numbers.

Vapnik Chervonenkis Dimension: Let X any set, and $\mathfrak{C} \subseteq \mathcal{P}(X)$. Then, for any $n \in \omega$, we say that $\mathrm{VC}(\mathfrak{C}) \geq n$ if there is some $A \subseteq X$, $|A| = n$, such that $\{c \cap A : c \in \mathfrak{C}\} = \mathcal{P}(A)$. That is, $\mathrm{VC}(\mathfrak{C}) \geq n$ if some size n subset of X is *shattered* by $\mathfrak{C}$. Then $\mathrm{VC}(\mathfrak{C}) = n$ if $\mathrm{VC}(\mathfrak{C}) \geq n$ and $\mathrm{VC}(\mathfrak{C}) \not\geq n+1$. If $\mathrm{VC}(\mathfrak{C}) \geq n$ for all $n \in \omega$, we say $\mathrm{VC}(\mathfrak{C}) = \infty$.

The model-theoretic notion of a definable family will be main object of interest in what follows. In order to introduce it, we give a brief description of model theory. For a reference, see [25] or [8].

By a *language*, denoted L, we mean a collection of relation, function and constant symbols, together with the usual logical symbols, *i.e.* quantifiers, variable symbols, conjunction, disjunction, negation and equality. The relation and function symbols are of a fixed arity, and there is a syntax[2] for composing well-formed formulas from the symbols. By a *L-formula* we mean a string of symbols, admissible with respect to the syntax.[3] A *sentence* is a formula with no free variables. A *theory*, or axiom system, is a collection of sentences, which we will always assume to be consistent. A theory is said to be *complete* if the truth value of every sentence is determined by the theory.

A *model* $\mathcal{M}$ of a theory (in a language L) interprets the symbols of L in a way which is consistent with the axioms of T, and provides a set of elements,

[2]In this paper we will use only the standard first-order syntax.

[3]The atomic L-formulas consist of single relation symbols from L, with terms (compositions of functions) as inputs. If $\phi(x)$ and $\psi(z)$ are L-formulas, then so are $\phi(x) \wedge \psi(z)$, $\exists(x)\phi(x)$, and $\neg\psi(z)$. There are also rules for substitution of variable symbols, *etc.*

with respect to which L formulas and sentences are either true or false. These elements are the *universe* of $\mathcal{M}$. The notation $\mathcal{M} \models T$ expresses that $\mathcal{M}$ is a model of T.

Fix a language L and a formula $\varphi(v_1, \ldots, v_n)$. We will deal with φ as a partitioned formula $\varphi(\bar{x}; \tilde{y})$ (or $\varphi(\bar{x}, \tilde{y})$.) This means that, $\varphi(\bar{x}; \tilde{y}) = \varphi(v_1, \ldots, v_n)$, but there is a separation of $(v_1, \ldots, v_n)$ into object and parameter variables. We always consider a formula φ with respect to some L-theory T, which is almost always assumed to be complete, with an infinite model.

To simplify things, we choose a large model $\mathbb{C}$ of a complete theory T, which is saturated in some extremely large cardinal κ, and regard any model discussed as being an elementary substructure of $\mathbb{C}$.[4] This is sometimes referred to as the "monster model," and we will denote ours by $\mathbb{C}$. By adopting a monster, we can speak of ambient sets A, B, C *etc*, without worrying very much about which universe they belong to; they belong to the universe of the monster.

We will sometimes write, $\models \varphi(a_1, \ldots, a_n)$, rather than the more formal $\mathbb{C}_{(a_1,\ldots,a_n)} \models \varphi(a_1, \ldots, a_n)$. Either of these says that $\varphi(a_1, \ldots, a_n)$, with $(a_1, \ldots, a_n) \in \mathbb{C}^n$, is true in $\mathbb{C}_{(a_1,\ldots,a_n)}$, where the subscript indicates that the language has been expanded to include names for the a_i.

For any formula $\psi(\bar{z})$, $\bar{z} = (z_1, \ldots, z_n)$, it is standard to let, for any model $\mathcal{M}$ of T, $\psi(\mathcal{M}) := \{\bar{a} \in \mathcal{M}^n : \models \psi(\bar{a})\}$, and if $D = \psi(\mathcal{M})$, to say that D is a definable subset of $\mathcal{M}$, defined by ψ. There is an obvious and analogous meaning for $\psi(A)$, when A is some unadorned subset of $\mathcal{M}$ (or $\mathcal{M}^n$.) To make the case where $A \subseteq \mathcal{M}^n$ clear, by $\psi(A)$, when $A \subseteq \mathcal{M}^{|\bar{x}|}$, we mean $\{\bar{a} \in A : \mathcal{M} \models \psi(\bar{a})\}$.

[4] $\mathcal{M}$ is an elementary substructure of $\mathbb{C}$ iff $\mathcal{M} \subseteq \mathbb{C}$, and any sentence with constants from $\mathcal{M}$ is true in $\mathcal{M}$ iff it is true in $\mathbb{C}$.

Given a partitioned formula $\varphi(\bar{x};\tilde{y})$, and a set $A \subseteq \mathcal{M}^{|\bar{x}|}$, the concept class, or *definable family*, associated with φ on A is,

$$\mathfrak{C}_\varphi^{\mathcal{M}}(A) := \{\varphi(A,\tilde{b}) : \tilde{b} \in \mathcal{M}^{|\tilde{y}|}\}.$$

If no model $\mathcal{M}$ is specified, then $\mathfrak{C}_\varphi(A)$ is taken to be $\{\varphi(A,\tilde{b}) : \tilde{b} \in \mathbb{C}^{|\tilde{y}|}\}$, where $\mathbb{C}$ is the monster model. There is a subtle but important difference in $\mathfrak{C}_\varphi^{\mathcal{M}}(A)$ and $\mathfrak{C}_\varphi(A)$, when $A \subseteq \mathcal{M}$; in the former case, the parameters must come from $\mathcal{M}$. This will not effect the VC dimension of the class, but it will be used in Chapter 4.

Note that $\mathfrak{C}_\varphi(A) \subseteq \mathcal{P}(A)$, and so $(A, \mathfrak{C}_\varphi(A))$ can be regarded as a familiar hypergraph, set system, etc. It thus has a well defined VC dimension. By the VC dimension of φ on A, denoted $\mathrm{VC}_A(\varphi)$, we mean the VC dimension of $\mathfrak{C}_\varphi(A)$.

Proposition 1.1.1. *For any models $\mathcal{M}$ and $\mathcal{N}$ of a complete theory T, VC($\mathfrak{C}_\varphi(\mathcal{M})$) = VC($\mathfrak{C}_\varphi(\mathcal{N})$).*

Proof. The following sentence, in T, says that the VC dimension of φ is at least d:

$$\exists \bar{y}_1 \ldots \exists \bar{y}_{2^d} \exists \bar{x}_1 \ldots \bar{x}_d \left(\bigwedge_{w \subseteq [d]} \left[\bigwedge_{i \in w} \varphi(x_i; y_{f(w)}) \wedge \bigwedge_{i \in [d] \setminus w} \neg\varphi(x_i; y_{f(w)}) \right] \right).$$ [5]

It is clear from the above, and the definition of VC dimension, that VC(φ) can be expressed in a sentence. Since $\mathcal{M}$ and $\mathcal{N}$ agree on all sentences, they agree on VC(φ). ☐

Definition Let $\mathfrak{A}, \mathfrak{B} \subseteq \mathcal{P}(X)$. We say that $\mathfrak{B}$ is an *extension* of $\mathfrak{A}$ if $\mathfrak{A} \subseteq \mathfrak{B}$.

[5] Here f is any bijection between $\mathcal{P}([d])$ and $[2^d]$.

Definition Let $\mathfrak{A} \subseteq \mathcal{P}(X)$, and $X_0 \subseteq X$. Then the *restriction* of $\mathfrak{A}$ to X_0, denoted $\mathfrak{A} \restriction_{X_0}$, is defined by

$$\mathfrak{A} \restriction_{X_0} := \{a \cap X_0 : a \in \mathfrak{A}\}.$$

While an extension increases the parameter set (or number of sets) in a concept class, a restriction alters the domain. The VC dimension is monotone with respect to extensions and restrictions in the following way.

Proposition 1.1.2. *For any X, and any $\mathfrak{C}_1, \mathfrak{C}_2 \subseteq \mathcal{P}(X)$, if $\mathfrak{C}_1 \subseteq \mathfrak{C}_2$, then $VC(\mathfrak{C}_1) \leq VC(\mathfrak{C}_2)$.*

Proof. Suppose $A \subseteq X$ is shattered by sets in $\mathfrak{C}_1$. Then A is shattered by the same sets in $\mathfrak{C}_2$. □

Proposition 1.1.3. *For any X, any $X_0 \subseteq X$, and $\mathfrak{A} \subseteq \mathcal{P}(X)$,*

$$VC(\mathfrak{A} \restriction_{X_0}) \leq VC(\mathfrak{A}).$$

Proof. Suppose $B \subseteq X_0$ is shattered by $\mathfrak{A} \restriction_{X_0}$. Then $\mathfrak{A}$ also shatters this set as a subset of X. □

We note that, for any set A whatever, $\mathrm{VC}(\mathfrak{C}_\varphi(A)) \leq \mathrm{VC}(\mathfrak{C}_\varphi(\mathbb{C}))$.

There are various types of operations defined on VC classes, such as simple set theoretical operations, so called 'shifts', cartesian products and others. These are investigated in [11] and [15]. We will be primarily interested in the "box union" and "box intersection" operations. The following notation, or something similar, can be found in [11], [5] and [12].

Proposition 1.1.4. *Suppose $\mathfrak{A}$ and $\mathfrak{B}$ are concept classes on X. Then the following are also concept classes on X.*

1. $\mathfrak{A} \sqcup \mathfrak{B} := \{a \cup b : a \in \mathfrak{A},\ \text{and } b \in \mathfrak{B}\}$.

2. $\mathfrak{A} \sqcap \mathfrak{B} := \{a \cap b : a \in \mathfrak{A},\ \text{and } b \in \mathfrak{B}\}$.

3. $\neg\mathfrak{A} := \{X \setminus a : a \in \mathfrak{A}\}$.

Also, for any formulas $\varphi(\bar{x}, \tilde{y})$, $\psi(\bar{x}, \tilde{z})$, and set $A \subseteq \mathbb{C}^{|\bar{x}|}$,

1. $\mathfrak{C}_\varphi(A) \sqcup \mathfrak{C}_\psi(A) = \mathfrak{C}_{\varphi(\bar{x},\tilde{y}) \vee \psi(\bar{x},\tilde{z})}(A)$.

2. $\mathfrak{C}_\varphi(A) \sqcap \mathfrak{C}_\psi(A) = \mathfrak{C}_{\varphi(\bar{x},\tilde{y}) \wedge \psi(\bar{x},\tilde{z})}(A)$.

3. $\neg\mathfrak{C}_\varphi(A) = \mathfrak{C}_{\neg\varphi}(A)$.

At several points we will be interested in the 'asymptotic' VC dimension of a formula. By this we mean the behavior of, say,

$$\mathrm{VC}(\bigwedge_{i=1}^{m} \varphi(\bar{x}, \tilde{y_i})),$$

as m grows arbitrarily large. Proposition 1.1.4 establishes that this is the same question as the one studied on box products in [11], [5] and [12]. Note that while the $\bar{x}$ variables are fixed, every new conjunction provides an independent set of parameters.

The following, which will figure largely in the next chapter, has been known since at least [5]. It is true for any boolean combination (in $\sqcap$, $\sqcup$, and $\neg$) though for simplicity we write the statement with the disjunction symbol.

Proposition 1.1.5. *Let $\mathfrak{C}_i \subseteq \mathcal{P}(X)$, $VC(\mathfrak{C}_i) = d$, for $i \in \omega$. Let, for any $m \in \omega$,*

$$\beta(m) = VC(\bigsqcup_{i=1}^{m} \mathfrak{C}_i).$$

Then $\beta(m)$ is $\mathcal{O}(m \log m)$, where the log *function has base 2.*

Proof. The bound will come from the Sauer-Shelah lemma, Lemma 1.2.1. Let $A \subseteq X$, finite, $|A| = n$. Then

$$| \left(\bigsqcup_{i=1}^{m} \mathfrak{C}_i \right) \restriction_A | \leq |\mathfrak{C}_1 \restriction_A | \cdot \cdots \cdot |\mathfrak{C}_m \restriction_A | \leq \left(\sum_{i=1}^{d} \binom{n}{i} \right)^m .$$

Since $\sum_{i=1}^{d} \binom{n}{i}$ is $\mathcal{O}(n^d)$, there is some $K_0 \in \omega$, depending only on d, such that the number of subsets of A cut out by $\sqcup_{i=1}^{m} \mathfrak{C}$ is at most $K_0^m n^{dm}$.

We claim that for all but finitely many values of m,

$$2^{(2d)m \log m} > K_0^m ((2d)m \log m)^{dm}.$$

Because,

$$2^{(2d)m \log m} > K_0^m ((2d)m \log m)^{dm}$$

$\iff$

$$m^{(2d)m} > K_0^m ((2d)m \log m)^{dm}$$

$\iff$

$$m^{2d} > K_0 ((2d)m \log m)^{d}$$

$\iff$

$$m^{d} > K_0 ((2d) \log m)^{d}$$

The last inequality clearly holds for all sufficiently large m.

Now we have shown, $\beta(m)$ is $\mathcal{O}(m \log m)$. For if not, a contradiction occurs when a set of size $(2d)m \log m$ is shattered, for sufficiently large m. □

It has recently been shown[6] in [12] that this bound is tight, provided $d \geq 5$. In fact, something stronger is shown, because in the argument of Eisenstat and

[6] We make a straightforward adaptation that smooths the result to the infinite case in section 2.6 of Chapter 2.

Angluin (as well as our adaptation) the bound is achieved using a single definable family, whereas in the above, the ω many classes have no necessary relationships.

1.1.1 Types

In chapters 3 and 4 we will make use of an object which is dual, in the abstract, to $\mathfrak{C}_\varphi(A)$. We call this a φ-type over some parameter set.

The word *type* describes a set of formulas, all in the same variables, assumed here to be consistent. Our φ-types will just be consistent types in which the only formulas allowed are formulas of the form $\varphi(\bar{x}, \tilde{a})$ or $\neg\varphi(\bar{x}, \tilde{a})$, where $\tilde{a}$ is some parameter set. The definitions which follow use the set theoretical properties of ordinal numbers (eg, $2 = \{0, 1\}$), and the notation ${}^{A}B$, which represents the set of all functions from A to B.

Definition 1. Suppose $\varphi(\bar{x}, \tilde{y})$ is a formula. For $i \in \{0, 1\}$,

$$\varphi(\bar{x}, \tilde{y})^i := \begin{cases} \varphi(\bar{x}, \tilde{y}) & \text{if i = 1,} \\ \neg\varphi(\bar{x}, \tilde{y}) & \text{if i = 0.} \end{cases}$$

2. If A is a set, $|\tilde{y}| = l$, and $\eta \in {}^{A^l}2$, let

$$r_\eta^{A,\varphi}(\bar{x}) := \{\varphi(\bar{x}, \tilde{a})^{\eta(\tilde{a})} : \tilde{a} \in A^l\}.$$

3. The set of formulas $r_\eta^{A,\varphi}(\bar{x})$ is *consistent* iff for every finite subset $\{\varphi(\bar{x}, \tilde{a_1})^{i_1}, \ldots, \varphi(\bar{x}, \tilde{a_n})^{i_n}\}$ of $r_\eta^{A,\varphi}(\bar{x})$, $i_j \in \{0, 1\}$,

$$\models \exists \bar{x} \bigwedge_{j=1}^{n} \varphi(\bar{x}, \tilde{a_j})^{i_j}.$$

(By the compactness theorem and the $|A|^+$-saturation of $\mathbb{C}$, this is equivalent to the existence of some $\bar{b} \in \mathbb{C}$ such that $\models r_\eta^{A,\varphi}(\bar{b})$. Such a $\bar{b}$ is said to *realize* the type $r_\eta^{A,\varphi}(\bar{x})$.)

4. For a formula $\varphi(\bar{x}, \tilde{a})$ and set A, the *Stone space* of φ over A is defined as

$$S_\varphi(A) := \{r_\eta^{A,\varphi}(\bar{x}) : \eta \in {}^{A^l}2 \text{ and } r_\eta^{A,\varphi}(\bar{x}) \text{ is consistent}\}.$$

As we alluded earlier, the Stone space of φ and the concept class associated with φ are dual. We now make the relationship precise. By the dual of a formula $\varphi(\bar{x}, \tilde{y})$, we mean $\varphi^*(\tilde{y}, \bar{x}) := \varphi(\bar{x}, \tilde{y})$. That is, the formula is syntactically the same, but we regard the object variables and parameters as reversed.

Definition For $\varphi(\bar{x}, \tilde{y})$ a formula, $A \subseteq \mathbb{C}^{|\bar{x}|}$ a set, define

$$[\mathfrak{C}_\varphi(A)] := \{f_c : c \in \mathfrak{C}_\varphi(A)\},$$

where $f_c \in {}^A2$ is the indicator function for c. Similarly, let

$$[S_\varphi(A)] := \{\eta : r_\eta^{A,\varphi}(\bar{x}) \in S_\varphi(A)\}.$$

Proposition 1.1.6. *If $\varphi(\bar{x}, \tilde{y})$ is a formula, $\varphi^*(\tilde{y}, \bar{x})$ its dual, and A any set, then*

$$[\mathfrak{C}_{\varphi^*}(A)] = [S_\varphi(A)].$$

Proof. Let $f_c \in [\mathfrak{C}_{\varphi^*}(A)]$, where $c = \varphi^*(A, \bar{b})$ for some $\bar{b} \in \mathbb{C}^{|\bar{x}|}$. Then $r_{f_c}^{A,\varphi}(\bar{x}) = \{\varphi(\bar{x}, \tilde{a})^{f_c(\tilde{a})} : \tilde{a} \in A^{|\tilde{y}|}\}$. But this is consistent because $\models r_{f_c}^{A,\varphi}(\bar{b})$. Thus $r_{f_c}^{A,\varphi}(\bar{x}) \in S_\varphi(A)$, and so $f_c \in [S_\varphi(A)]$. This gives $[\mathfrak{C}_{\varphi^*}(A)] \subseteq [S_\varphi(A)]$.

For the other direction, take $\eta \in {}^{A^{|\tilde{y}|}}2$, such that $r_\eta^{A,\varphi}(\bar{x})$ is consistent. Then by the compactness theorem and the $\aleph_0 + |A|^+$-saturation of $\mathbb{C}$, there is some $\bar{b} \in \mathbb{C}^{|\bar{x}|}$ realizing $r_\eta^{A,\varphi}(\bar{x})$. Then for any $\tilde{a} \in A^{|\tilde{y}|}$, we have $\eta(\tilde{a}) \iff \models \varphi(\bar{b}, \tilde{a}) \iff \models \varphi^*(\tilde{a}, \bar{b})$. Thus if $c = \varphi^*(A, \bar{b})$, $f_c = \eta$, and so $\eta \in [\mathfrak{C}_{\varphi^*}(A)]$.

Therefore $[\mathfrak{C}_{\varphi^*}(A)] = [S_\varphi(A)]$. □

We make some well known remarks about the relations between VC($\mathfrak{C}_{\varphi^*}(A)$) and VC($\mathfrak{C}_\varphi(A)$) [23, 28]. First we define the notion of so-called set-theoretic independence. Intuitively, it is the largest 'topological' Venn-diagram that can be created from elements of a concept class. We use the ordinal definition of a natural number, $k := \{0, 1, 2, \ldots, k-1\}$.

Definition Let X a set, $\mathfrak{C} \subseteq \mathcal{P}(X)$. Then the independence dimension of $\mathfrak{C}$, denoted IN($\mathfrak{C}$) is the largest $n \in \omega$ such that there are $c_0, c_1, ..., c_{n-1}$ in $\mathfrak{C}$ such that, for any $w \subseteq n$,

$$\bigcap_{i \in w} c_i \cap \bigcap_{i \in n \setminus w} X \setminus c_i \neq \emptyset.$$

If there is no such largest n, we say IN($\mathfrak{C}$) $= \infty$.

The independence dimension of a formula $\varphi(\bar{x}, \tilde{y})$ will be the independence dimension of its associated concept class, and will be denoted IN(φ). The independence dimension of a formula $\varphi(\bar{x}, \tilde{y})$ over $A \subseteq \mathbb{C}^{|\tilde{y}|}$ will be the independence dimension of its associated concept class restricted to sets defined by φ parameterized with tuples from A, and will be denoted IN$_A(\varphi)$.

The following shorthand will be used many times.

Definition We write "$B \subseteq_\kappa A$" to mean: $B \subseteq A$ and $|B| < \kappa$.

Proposition 1.1.7. *1. For any formula $\varphi(\bar{x}, \tilde{y})$, and any set $A \subseteq \mathbb{C}^{|\tilde{y}|}$, IN$_A(\varphi)$ = VC$_A(\varphi^*)$. In particular, IN(φ) = VC(φ^*).*

2. Let $\varphi(\bar{x}, \tilde{y})$, $A \subseteq \mathbb{C}^{|\tilde{y}|}$, and $k \in \omega$ be given. The following are equivalent.

(a) IN$_A(\varphi) \geq k$

(b) There is $B \subseteq_{k+1} A$ such that $|\mathfrak{C}_{\varphi^*}(B)| = 2^k$

(c) *There are* $D_0, D_1, ..., D_{k-1}$*, where for all* $i \in k$*,* D_i *is defined by* $\varphi(\bar{x}, \tilde{a}_i)$*, some* $\tilde{a}_i \in A$*, and for any* $w \subseteq k$*,*

$$\bigcap_{i \in w} D_i \cap \bigcap_{i \in k \setminus w} \bar{D}_i \neq \emptyset.$$

3. *Let* $\varphi(\bar{x}, \tilde{y})$ *a formula. Then*

$$VC(\varphi) \leq 2^{IN(\varphi)},$$

and

$$IN(\varphi) \leq 2^{VC(\varphi)}.$$

A proof for the above can be found in [23]. A less model-theoretic formulation may be sought in [37]. Note that 1.1.7 (3) implies that 'finite VC dimension' and 'finite independence dimension' are equivalent.

1.1.2 Topology

$S_\varphi(A)$ (or $\mathfrak{C}_\varphi(A)$) is equipped with a nice topology, called the Stone topology. It makes sense, then, to henceforth think of its elements as *points*, and to denote them by simple p's and q's. Because we will eventually make reference to the Stone space topology, we will make an excursion into its properties.

Definition 1. A $\pm$*instance* of $\varphi(\bar{x}, \tilde{y})$ is a formula $\psi(\bar{x})$ of the form $\varphi(x, \tilde{a})$ or $\neg\varphi(\bar{x}, \tilde{a})$, where $\tilde{a}$ is any tuple (in $\mathbb{C}$).

2. A basic open set of $S_\varphi(A)$ is a set of the form

$$U_{\psi(\bar{x})} := \{p \in S_\varphi(A) : \psi \in p\},$$

where ψ is a $\pm$instance of φ.

The Stone space $S_\varphi(A)$, has the topology induced by the above described base of open sets.

Recall that if $2 = \{0, 1\}$ is regarded as a discrete space, and B is any set, then 2^B (which we will identify with ${}^B 2$) can be equipped with the Tychonov (or product) topology, where the basic open sets are of the form $U_{\eta,b} := \{\mu \in {}^B 2 : \mu(b) = \eta(b)\}$, where η is any member of ${}^B 2$, and $b \in B$. Then the Stone space topology is just the subspace topology on $[S_\varphi(A)] \subseteq {}^{A^{|\bar{y}|}} 2$, where ${}^{A^{|\bar{y}|}} 2$ is given the product topology on 2. More formally,

Proposition 1.1.8. *Let $({}^{A^{|\bar{y}|}} 2, \tau)$ be the topological space given by the product topology on 2, and let $([S_\varphi(A)], \tau') \subseteq ({}^{A^{|\bar{y}|}} 2, \tau)$ have the subspace topology. Then if $(S_\varphi(A), \rho)$ denotes the usual Stone topology on $S_\varphi(A)$, we have the following homeomorphism:*

$$(S_\varphi(A), \rho) \cong ([S_\varphi(A)], \tau').$$

From this we get that $S_\varphi(A)$ is Tychonov[7], since 2 is Tychonov, and the property is preserved under products and subspaces. The content of the compactness theorem (in this local setting) is that $[S_\varphi(A)]$ is a closed, equivalently compact, subspace of the compact Hausdorff space ${}^{A^{|\bar{y}|}} 2$.

1.2 Maximal classes and Sauer's lemma

The most important property a concept class has as a result of finite VC dimension, particularly outside of logic, is the existence of a polynomial bound on size of the trace function of the system. The technical statement of this fact is the Sauer-Shelah-VC lemma.

[7] Completely regular and Hausdorff

Lemma 1.2.1 (Sauer's lemma)**.** *Let X a set $\mathfrak{C} \subseteq \mathcal{P}(X)$. Then for any finite set $A \subseteq X$,*

$$|\mathfrak{C} \restriction A| \leq \sum_{i=0}^{d} \binom{|A|}{i}.$$

where d is the VC dimension of $\mathfrak{C}$.

Proof. We introduce the semi-standard notation $\Phi_d(n) := \sum_{i=0}^{d} \binom{n}{i}$, and note that the following recurrence relation is satisfied:[8]

$$\Phi_d(n) = \Phi_d(n-1) + \Phi_{d-1}(n-1).$$

The argument will be by double induction on d and n. If $d = 0$, then for any finite A, $|\mathfrak{C} \restriction A| \leq 1$ and the statement holds. Also, if $|A| = 0$ then the statement holds for any d.

Suppose the statement holds for any A' of size less than n and any $d' < d$. Let A of size n and suppose $\mathfrak{C}$ has VC dimension d on A. Pick any $a \in A$, and make the following definitions:

$$\mathfrak{C} - a := \mathfrak{C} \restriction_{A \setminus \{a\}}$$

$$\mathfrak{C}^{\{a\}} := \{c \in \mathfrak{C} - a : c \in \mathfrak{C} \text{ and } (c \cup \{a\}) \in \mathfrak{C}\}.$$

Clearly $\mathrm{VC}(\mathfrak{C} - a) \leq d$, and $\mathfrak{C} - a$ is a concept class on the set $A \setminus \{a\}$ of size $n - 1$. The class $\mathfrak{C}^{\{a\}}$ is also a class on $A \setminus \{a\}$, and has VC dimension at most $d - 1$; otherwise if $\mathfrak{C}^{\{a\}}$ shatters B of size d, then $\mathfrak{C} \restriction_A$ shatters $B \cup \{a\}$ of size $d + 1$ $\bigotimes$.

Now we observe, since every element of $\mathfrak{C}^{\{a\}}$ extends to two separate elements of $\mathfrak{C} \restriction_A$,

$$|\mathfrak{C} \restriction_A| = 2|\mathfrak{C}^{\{a\}}| + (|\mathfrak{C} - a| - |\mathfrak{C}^{\{a\}}|) = |\mathfrak{C}^{\{a\}}| + |\mathfrak{C} - a|.$$

[8] If we require $\Phi_d(0) = \Phi_0(m) = 1$, then this relation in fact characterizes $\Phi_d(n)$ [20].

By inductive hypothesis,

$$|\mathfrak{C}^{\{a\}}| + |\mathfrak{C} - a| \leq \Phi_{d-1}(n-1) + \Phi_d(n-1) = \Phi_d(n).$$

This establishes the lemma.

□

A concept class which has the maximum size allowable by Sauer's lemma has interesting properties as a result. See, for instance [39, 13, 14, 22, 15].

Definition Suppose X is a set, $\mathfrak{C} \subseteq \mathcal{P}(X)$, $\mathrm{VC}(\mathfrak{C}) = d$, and for every finite subset A of X, $|\mathfrak{C} \restriction A| = \sum_{i=0}^{d} \binom{|A|}{i}$. Then $\mathfrak{C}$ is *Sauer maximal*, abbreviated as $\mathcal{S}$-maximal.

Not only is the bound given by Sauer's lemma tight, it obtains in many natural situations. We enumerate a few examples of $\mathcal{S}$-maximal families to illustrate.

1. The concept class of all convex subsets in a dense linear order is $\mathcal{S}$-maximal.

2. For any natural number $e+1$, there is a dense subset X of $\mathbb{R}^{e+1}$, such that $(X, \mathfrak{C})$ is a $\mathcal{S}$-maximal class of VC dimension e, where $\mathfrak{C}$ is the set of all positive half-spaces in $\mathbb{R}^{e+1}$, restricted to X [13].

3. $(X, [X]^{\leq d})$, where $d \in \omega$, and X has size at least d.

Model theoretically, such classes can be stable, as in $\mathfrak{C} = [X]^{\leq d}$, or unstable, as in the case of (1).

The notion of *combinatorial density* is from [11, 2].

Definition For $\mathfrak{C} \subseteq \mathcal{P}(X)$, the *combinatorial density* of $\mathfrak{C}$, denoted $dens(\mathfrak{C})$, is a real number defined as:

$$dens(\mathfrak{C}) := inf\{r : r > 0 \text{ and } \exists K \in \omega \ \forall n \in \omega \ \forall A \subseteq_{n+1} X, \ |\mathfrak{C} \restriction_A | \leq K n^r\}.$$

Whereas VC dimension describes, *a priori*, the behavior of $\mathfrak{C}$ on a single set, combinatorial density describes how $\mathfrak{C}$ acts on arbitrarily large sets. A class may have a very large VC dimension, and yet have combinatorial density zero. The fundamental result of Vapnik and Chervonenkis states that the combinatorial density of a class is always at most the VC dimension [38].[9]

Dudley observed that combinatorial density of classes always combines linearly. From the proof of the following proposition, it is clear that the operator does not matter; the same fact is true for conjunction.

Proposition 1.2.2 (Dudley). *Let* $\mathfrak{C}, \mathfrak{A} \subseteq \mathcal{P}(X)$*, with* $r = dens(\mathfrak{C})$ *and* $s = dens(\mathfrak{A})$*. Then* $dens(\mathfrak{C} \sqcup \mathfrak{A}) \leq r + s$*.*

Proof. Fix $A \subseteq X$. Clearly $|\mathfrak{C} \sqcup \mathfrak{A} \restriction_A| \leq |\mathfrak{C} \restriction_A| \cdot |\mathfrak{A} \restriction_A|$. □

Proposition 1.2.3. *Let* $\mathfrak{C} \subseteq \mathcal{P}(X)$ *of VC dimension d, and suppose there is an infinite subset* $X' \subseteq X$ *such that* $\mathfrak{C} \restriction_{X'}$ *is* $\mathcal{S}$*-maximal of VC dimension d. Then* $dens(\mathfrak{C}) = VC(\mathfrak{C})$*.*

Proof. Clear by the definition, and the fact that $\sum_{i=0}^{d} \binom{n}{i}$ is $\mathcal{O}(n^d)$. □

Most of the nice properties of $\mathcal{S}$-maximal classes are preserved under substructure (*ie*, subset if the set system), so it is of interest to know which set systems $(X, \mathfrak{C})$ of VC dimension d can be extended to $\mathcal{S}$-maximal classes of the same VC dimension. We offer the following, which can be omitted with no loss for subsequent chapters.

Proposition 1.2.4. *Let* $\mathfrak{C} \subseteq \mathcal{P}(X)$ *be a concept class of VC dimension d. Then* $\mathfrak{C}$ *extends to a* $\mathcal{S}$*-maximal class on X iff* $\mathfrak{C} \restriction_A$ *extends to a* $\mathcal{S}$*-maximal class on A for every finite* $A \subseteq X$*.*

---

[9]Sauer's lemma can be thought of as a more precise statement of this fact.

Proof. The forward direction is trivial.

For the reverse direction, consider $\mathfrak{C}$ represented as the definable family associated with R in the structure $\mathcal{M} = (X, Y, R(x, y))$, $R \subseteq X \times Y$.

Let $\Delta_{X,Y}(\mathcal{M})$ be the basic diagram of $\mathcal{M}$ with constants from X, Y. For every finite $X_0 \subseteq X$, let θ_{X_0} say that R is $\mathcal{S}$-maximal of dimension d on X_0. We want to show that $\Delta_{X,Y}(\mathcal{M}) \cup \bigcup_{X_0 \subseteq_\omega X} \{\theta_{X_0}\}$ is consistent. It suffices, by compactness, to show that for every X_0,

$$\Delta_{X_0,Y}(\mathcal{M}) \cup \{\theta_{X_0}\}$$

is consistent. But this is true because any finite restriction of $\mathfrak{C}$ has an $\mathcal{S}$-maximal extension of VC dimension d. Thus, using a theorem from basic model theory, there is some structure $\mathcal{M}'$, in the same language as $\mathcal{M}$, such that $\mathcal{M} \subseteq \mathcal{M}'$.[10] Since $\mathfrak{C}_R(\mathcal{M}') \restriction_{\mathcal{M}}$ is $\mathcal{S}$-maximal on $\mathcal{M}$ of VC dimension d, this is the desired extension. □

One might wonder whether any concept class $\mathfrak{C}$ can be extended to a $\mathcal{S}$-maximal class simply by adding new subsets of X, one after another, until it is no longer possible to add a new subset without increasing the VC dimension of $\mathfrak{C}$. Some easy finite counter-examples show that things are not so simple. Thus we need a different notion of 'maximality' to describe a class to which no new subsets can be added, unless such an addition increases the VC dimension.

Definition Let X a set, and $\mathfrak{C} \subseteq \mathcal{P}(X)$, of VC dimension $d \in \omega$. We say that $\mathfrak{C}$ is *Dudley maximal*, or $\mathcal{D}$-maximal, if for any set $A \subseteq X$, $A \in \mathcal{P}(X) \setminus \mathfrak{C} \implies \mathrm{VC}(\mathfrak{C} \cup \{A\}) > d$.[11]

[10] That is, $\mathcal{M}$ is a model theoretic substructure of $\mathcal{M}'$.

[11] What we have defined as $\mathcal{S}$-maximal classes are elsewhere known as 'complete' [39], or max-

1.3 Stability

In this section we will define what it means for a concept class to be stable, in the model theoretic sense. The standard definition is related to the notion of a 'definable type,' as discussed in Chapter 3. The notion of stability and the propositions in this section are due to Shelah.

Let $\varphi(\bar{x}, \tilde{y})$ a L formula, $\mathcal{M}$ a L model, and $B \subseteq \mathcal{M}$.

Definition For $C \in \mathfrak{C}_\varphi(B)$, $A \subseteq \mathcal{M}$, say that C is *definable* over A if there is an L formula $\delta(\bar{x})$, with parameters only from A, such that for all $\bar{b} \in B^{|\bar{x}|}$,

$$\bar{b} \in C \iff \models \delta(\bar{b}).$$

We will define stability in terms of definability. First we need a way to fix a language for an arbitrary concept class.

Definition For $\mathfrak{A} \subseteq \mathcal{P}(X)$, the *natural structure* associated with $\mathfrak{A}$ will be the model $\mathcal{M} = (X, Y, R(x, y))$, where $Y^{\mathcal{M}} = \mathfrak{A}$, and for all $b \in X$, $C \in Y$,

$$\mathcal{M} \models R(b, C) \iff b \in C.$$

Note that $\mathfrak{A} = \mathfrak{C}_R^{\mathcal{M}}(\mathcal{M})$, when $\mathcal{M}$ is the natural structure associated with $\mathfrak{A}$.

Definition Let $\mathfrak{A} \subseteq \mathcal{P}(X)$, and $\mathcal{M}$ the associated natural structure. Let $\mathbb{C}$ be the monster model of $Th(\mathcal{M})$. We say that $\mathfrak{A}$ is *stable* if for all $A \subseteq \mathbb{C}$, every $C \in \mathfrak{C}_R(A)$ is definable over A.

---

imum [13, 14]. On the other hand, what we identify as $\mathcal{D}$-maximal classes below are sometimes called simply 'maximal' classes. Since maximum/maximal is difficult to keep straight, and 'complete' has other connotations in model theory, I have adopted the $\mathcal{S}$-maximal/$\mathcal{D}$-maximal scheme. So far as I know, Dudley was the first to investigate classes of the relevant maximality type [11].

We say that a L formula $\varphi(\bar{x}, \tilde{y})$ is *stable* with respect to a complete L theory T if $\mathfrak{C}_\varphi(\mathcal{M})$ is stable for some model $\mathcal{M}$ of T.[12]

The following proposition shows that the notion of stability of a formula is well-defined.

Proposition 1.3.1. *Let $\varphi(\bar{x}, \tilde{y})$ a L formula, T a complete L theory, and $\mathcal{M}, \mathcal{N}$ models of T. Then $\mathfrak{C}_\varphi(\mathcal{M})$ is stable $\iff$ $\mathfrak{C}_\varphi(\mathcal{N})$ is stable.*

Proof. The natural structures associated with $\mathfrak{C}_\varphi(\mathcal{M})$ and $\mathfrak{C}_\varphi(\mathcal{N})$ are elementarily equivalent. Thus if $\mathbb{C} \models Th(\mathfrak{C}_\varphi(\mathcal{N}))$ is the monster model of their natural structure, it contains both as elementary substructures. $\square$

From the above, stability is a 'property of the theory,' in the sense that the stability of a formula is a quality which is invariant across all models of a complete theory. Therefore we say that a complete L theory T is *stable* if every parameterized L formula is stable with respect to T.

Examples of stable theories are [25]:

1. The theory of any algebraically closed field
2. The theory of any differentially closed field
3. The theory of any $\mathbb{Z}$-group

Unfortunately, most of the standard geometric families are unstable, as can be seen by constructing infinite descending chains. Such chains give a family the strict order property.

[12]Note that this is saying that every element of $\mathfrak{C}_\varphi(A)$ is definable by a formula in the language $L \restriction_{\varphi(\bar{x}, \tilde{y})}$ with parameters from A whenever $A \subseteq \mathbb{C}$.

Definition Let $\varphi(\bar{x}, \tilde{y})$ an L-formula, and T a complete L theory. We say that $\varphi(\bar{x}, \tilde{y})$ has the *strict order property* with respect to T, if for some (equivalently every) model $\mathcal{M}$ of T, and every $n \in \omega$, there are $C_1, \ldots, C_n$ in $\mathfrak{C}_{\varphi}^{\mathcal{M}}(\mathcal{M})$, such that

$$C_1 \supsetneq C_2 \supsetneq \cdots \supsetneq C_n.$$

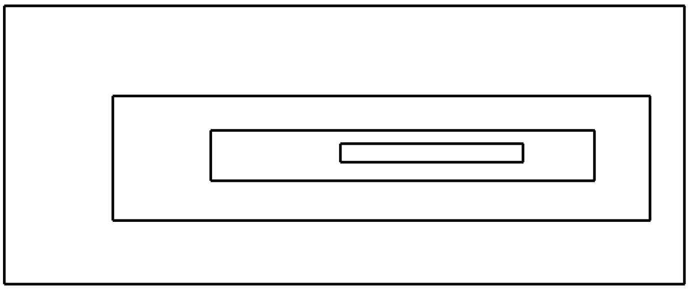

Figure 1.1: Axis parallel rectangles are unstable, because of infinite descending chains.

We take a moment to establish that the strict order property implies the absence of stability.

Proposition 1.3.2. *Let T a complete L theory, and $\varphi(\bar{x}, \tilde{y})$ a L formula. Suppose $\varphi(\bar{x}, \tilde{y})$ has the strict order property with respect to T. Then $\varphi(\bar{x}, \tilde{y})$ is not stable with respect to T.*

Proof. Suppose $\varphi(\bar{x}, \tilde{y})$ is stable with respect to T. Let T' be the complete theory of the natural structure associated with $\mathfrak{C}_{\varphi}(\mathcal{N})$ for some model $\mathcal{N}$ of T. By way of contradiction, assume also that $\varphi(\bar{x}, \tilde{y})$ has the strict order property. Then by the compactness theorem, we can find a countable model $\mathcal{M}$ of T', such that $\mathfrak{C}_{\varphi}(\mathcal{M})$[13] contains an infinite subset S, where S is linearly ordered by inclusion.

[13] We use φ and R interchangeably here.

That is, we can define a linear ordering $<_{inc}$ on S by

$$A <_{inc} B \iff A \subsetneq B,$$

for all $A, B \in S$. In fact, also by compactness, we can do this in such a way that $(S, <_{inc})$ and $(\mathbb{Q}, <)$ are isomorphic as orderings. Fix such an isomorphism $f : (\mathbb{Q}, <) \to (S, <_{inc})$, and for every $p \in \mathbb{Q}$, denote the parameter set defining the corresponding element of S by $\tilde{a}_p$. That is, for all $p \in \mathbb{Q}$, $\varphi(\mathcal{M}, \tilde{a}_p) = f(p)$.

Let $B = \{\tilde{a}_p : p \in \mathbb{Q}\}$. For any $\tilde{a} \in B$, define a formula $\theta_{\tilde{a}}(\tilde{y})$ by

$$\theta_{\tilde{a}}(\tilde{y}) = \exists \bar{x}(\neg\varphi(\bar{x}, \tilde{a}) \wedge \varphi(\bar{x}, \tilde{y})) \wedge \forall \bar{x}(\varphi(\bar{x}, \tilde{a}) \to \varphi(\bar{x}, \tilde{y})).$$

This says that if $\models \theta_{\tilde{a}}(\tilde{b})$, then $\varphi(\mathcal{M}, \tilde{a}) \subsetneq \varphi(\mathcal{M}, \tilde{b})$.

For any real number r, define a set of formulas $\Gamma_r(\tilde{y})$ by

$$\Gamma_r(\tilde{y}) = \{\theta_{\tilde{a_p}}(\tilde{y}) : p \in \mathbb{Q}, p < r\} \cup$$

$$\{\neg\theta_{\tilde{a_p}}(\tilde{y}) : p \in \mathbb{Q}, p \geq r\}.$$

By the properties of dense linear order, $\Gamma_r(\tilde{y})$ is consistent for every $r \in \mathbb{R}$. Also, for any real numbers r, t, $r \neq t$, $\Gamma_r(\tilde{y})$ and $\Gamma_t(\tilde{y})$ are inconsistent. Finally, if $s \in S_{\varphi^*}(\mathcal{M})$, and $\tilde{a}, \tilde{b} \in \mathcal{M}^{|\tilde{y}|}$ are such that $\models s(\tilde{a})$ and $\models s(\tilde{b})$, then for any $r \in \mathbb{R}$, $\models \Gamma_r(\tilde{a}) \iff \models \Gamma_r(\tilde{b})$.

Thus $S_{\varphi^*}(\mathcal{M})$ has cardinality $|\mathbb{R}| = 2^{\aleph_0}$.

On the other hand, since $\varphi(\bar{x}, \tilde{y})$ is stable, every element of $\mathfrak{C}_\varphi(\mathcal{M})$ is definable over $\mathcal{M}$. In fact, if $Fm_{L'}(M)$ denotes the formulas of $L' = \{X, Y, R(x, y)\}$ with parameters in $\mathcal{M}$, then this set is countable. This implies that $\mathfrak{C}_\varphi(\mathcal{M})$ is countable. But since necessarily $|\mathfrak{C}_\varphi(\mathcal{M})| = |S_{\varphi^*}(\mathcal{M})|$, this is a contradiction. □

This shows, in Chapter 3, that the existence of a compression scheme for a concept class is strictly weaker than stability.

There are many equivalent definitions of stability. The following is one of the most common. We will not prove or assume its equivalence.

Definition A formula $\varphi(\bar{x}, \tilde{y})$ is said to have the *order property* with respect to a complete theory T if there are sequences $\langle \bar{a}_i : i \in \omega \rangle$, $\langle \tilde{c}_i : i \in \omega \rangle$ in $\mathbb{C}$ such that for all $i, j \in \omega$,

$$\models \varphi(\bar{a}_i, \tilde{c}_j) \iff i < j.$$

For the proof of Proposition 3.2.1, we will use Conclusion I, 2.11 from [32].

Conclusion I, 2.11: If for some $A \subseteq \mathbb{C}$, $|S_\varphi(A)| > |A| + \aleph_0$, then $\varphi(\bar{x}, \tilde{y})$ has the order property.

From a cardinality argument very similar to the one in the proof of Proposition 1.3.2, it can be seen that the converse of Conclusion I, 2.11 holds as well.

We will need:

Lemma 1.3.3. *If for some $A \subseteq \mathbb{C}$, and some formula $\varphi(\bar{x}, \tilde{y})$, $|S_\varphi(A)| > |A| + \aleph_0$, then there is some $B \subseteq \mathbb{C}$ such that $|\mathfrak{C}_\varphi(B)| > |B| + \aleph_0$.*

Proof. Suppose, for some $A \subseteq \mathbb{C}$, that $|S_\varphi(A)| > |A| + \aleph_0$. By Conclusion I, 2.11, $\varphi(\bar{x}, \tilde{y})$ has the order property. By compactness, we can find sequences $\langle \bar{a}_i : i \in \mathbb{Q} \rangle$, $\langle \tilde{c}_i : i \in \mathbb{R} \rangle$ such that

$$\models \varphi(\bar{a}_i, \tilde{c}_j) \iff i < j.$$

Let B denote the elements of $\langle \bar{a}_i : i \in \mathbb{Q} \rangle$, and D denote the elements of $\langle \tilde{c}_i : i \in \mathbb{R} \rangle$. For every $r \in \mathbb{R}$, define a type $p_r \in S_{\varphi^*}(B)$ by

$$p_r = \{\varphi^*(\tilde{y}, \bar{a}_i) : i < r\} \cup \{\neg\varphi^*(\tilde{y}, \bar{a}_i) : i \geq r\}.$$

Each p_r is consistent, since $\models p_r(\tilde{c}_r)$. Then the mapping $f : D \to S_{\varphi^*}(B)$, $\tilde{c}_r \longmapsto p_r$, is 1-1, and so $|S_{\varphi^*}(B)| \geq 2^{\aleph_0}$. However, it is clear that $|B| = |\mathbb{Q}| = \aleph_0$. Since $|\mathfrak{C}_\varphi(B)| = |S_{\varphi^*}(B)|$, the lemma is proved. □

1.4 Geometric Examples

We go over some standard geometric families, emphasizing their definability or non-definability, and compute their VC dimensions.[14]

Example 1.4.0.1. *Axis Parallel Rectangles*

Let $X = \mathbb{R}^2$*. We can define the class of all axis parallel rectangles in euclidean space as* $\mathfrak{C}_\varphi(\mathbb{R})$*, where*

$$\varphi(x_1, x_2; y_l, y_r, y_b, y_t) := x_1 > y_l \wedge x_1 < y_r \wedge x_2 > y_b \wedge x_2 < y_t,$$

and T *is the complete theory of* $(\mathbb{R}, \leq)$*.*

We show that the VC dimension of φ is 4. We must show that some set of size 4 is shattered by $\mathfrak{C}$, and that $\mathfrak{C}$ shatters no set of size 5. The set of points pictured below will be referred to as A.

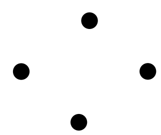

Figure 1.2: The 4 point set A.

According to the definition of VC dimension, we must be able to realize any subset $B \subseteq A$ by intersecting A with an element of $\mathfrak{C}$. Below we represent B by labeling the diagram with the symbols $\{+, -\}$, where the positively labeled points are the elements of B, and the negatively labeled points are the elements of $A \setminus B$.

To cut out the set B from A, it suffices to find an axis parallel rectangle which includes exactly the positive points in the labeled diagram of A. This is done in Figure 1.4, below.

[14]The results and proof techniques in this section are well-known.

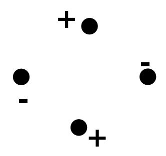

Figure 1.3: The subset B of A.

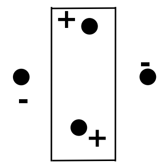

Figure 1.4: Realizing the labeling

There are $2^{|A|} = 16$ possible labellings, or subsets, of A; realizing the others is left as an exercise.

Now that we have shown $\mathfrak{C}$ shatters a set of size 4, we must show it shatters no sets of size 5. It suffices to consider only convex sets of points, for the reason illustrated in Figure 1.5.

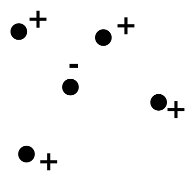

Figure 1.5: A labeling of a size 5 set not realizable with a rectangle.

Since every element of $\mathfrak{C}$ is convex, no element can realize the labeling of a nonconvex set which includes the generators of the convex hull of the set, but omits an interior point.

We now consider a generic convex 5 point set A. We may choose 4 points of A, representing the leftmost, topmost, rightmost, and bottommost elements, as shown in Figure 1.6. It may happen that two distinct points are both the, say, "leftmost" points of A, but we leave this case to the reader, and assume a good general position.

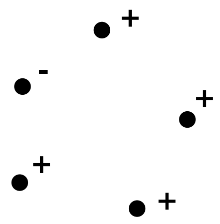

Figure 1.6: A convex five point set, with the extremal points relative to the two axes labeled positively.

Now it is clear that no axis parallel rectangle can realize the configuration in figure 1.6, because any rectangle including the extremal points must also include the negatively labeled point.

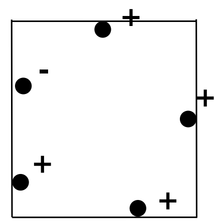

Figure 1.7: A labeling of a size 5 set not realizable with a rectangle.

Thus no size 5 set is shattered, and so the VC dimension of $\mathfrak{C}$ is 4. A similar argument shows that the VC dimension of axis parallel hyper-rectangles in $\mathbb{R}^d$ is at most $2d$.

Example 1.4.0.2. *Convex sets in $\mathbb{R}^d$*

Let $X = \mathbb{R}^d$, and

$$\mathfrak{C}^d_{conv} = \{A : A \subseteq \mathbb{R}^d \wedge A \textit{ is convex}\}.$$

We show $VC(\mathfrak{C}^d_{conv}) = \infty$.

The general case is clear from the proof for $d = 2$. Let A be any finite number of points arranged on a 1-sphere in $\mathbb{R}^2$.

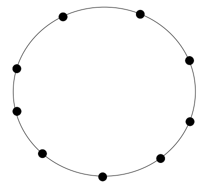

Figure 1.8: A set A of points, arranged on a circle.

For any labeling of the points in figure 1.8, there is a convex set which realizes the labeling. In fact there is a convex polygon which realizes the labeling, showing that the class of all convex polytopes has infinite VC dimension as well.

It is significant that $\mathfrak{C}^d_{conv}$ is not realizable as a definable family for any formula in a natural algebraic structure. Mathematical systems may be separated into those which are *dependent*, meaning that any parameterized formula gives rise to a set system with finite VC dimension, and the remainder, which are said to be *independent*. Important examples of independent systems are natural number arithmetic, $(\mathbb{N}, +, \cdot, 0, 1)$, and $(\mathbb{R}, \cdot, +, sin(x), 0, 1)$. The reader may wish to find a formula for each of these respective systems with infinite VC dimension.

The following venerable theorem, can be found in [26]. It is one of the most important tools in establishing upper bounds on VC dimension in geometric set systems.

Theorem 1.4.1 (Radon's lemma)**.** *Let* A *be a set of* $d + 2$ *points in* $\mathbb{R}^d$*. Then there exist two disjoint subsets* $A_1, A_2 \subseteq A$ *such that*

$$conv(A_1) \cap conv(A_2) \neq \emptyset.$$

Example 1.4.0.3. *Balls in* $\mathbb{R}^d$

Let $X = \mathbb{R}^d$

$\mathfrak{C}^d_\circ = \{\{\bar{x} \in \mathbb{R}^d : ||\bar{x} - \bar{y}|| < r\} : \bar{y} \in \mathbb{R}^d \wedge r \in \mathbb{R}\}$

$VC(\mathfrak{C}^d_\circ) = d + 1$

Since the definition of this family is phrased in terms of the usual euclidian distance metric, it is easy to see that it will be definable in any expansion of the real field. The dependence of the real field therefore provides a quick argument that the VC dimension is finite.

To see that it is the claimed value, suppose by way of contradiction that a set A of size $d+2$ is shattered. By Radon's theorem[15] we can write A as the disjoint union of C and D, where $conv(C) \cap conv(D) \neq \emptyset$. The hypothesis then implies that there are balls $b_C, b_D \in \mathfrak{C}^d_\circ$ such that $b_C \cap A = C$ and $b_D \cap A = D$. Since b_C and b_D are convex, $conv(C) \subseteq b_C$ and $conv(D) \subseteq b_D$. Therefore $b_C \cap b_D \neq \emptyset$.

The intersection of the open balls b_C and b_D determines a unique hyperplane h in $\mathbb{R}^d$, which separates A into disjoint sets C and D. The two open halfspaces (positive h^+ and negative h^-) associated with h are both convex, and so, without loss, $conv(C) \subseteq h^+$ and $conv(D) \subseteq h^-$. But then $h^+ \cap h^- \neq \emptyset$, which is absurd.

On the other hand, the $d+1$ points determined by the unit vectors and the origin are always shattered.[16]

---

[15]This states that any set of d + 2 points in $\mathbb{R}^d$ can be partitioned into two disjoint sets whose convex hulls intersect.

[16]Hint: Shatter the points with half-spaces, then replace the hyperplanes with sufficiently large spheres.

Chapter 2

VC linearity

2.1 Introduction

In this chapter we define what it means for a formula to be linear in VC dimension, and give some examples. We also prove the existence of VC linear theories, including the theory of R-modules, and draw some consequences. The Veronese mapping is studied as a means of transferring linearity results from one class to another. Some uses of Tychonov closures in VC theory are introduced.

2.1.1 Definitions

There is a well known result, which states that the VC dimension of m intervals on the real line is $2m$. From the standpoint of formulas, this says that if $\varphi(x, y_1, y_2) = x < y_1 \wedge y_2 < x$, then, with respect to $(\mathbb{R}, \leq)$,

$$\mathrm{VC}(\bigvee_{i=1}^{m} \varphi(x, (y_1, y_2)_i)) = 2m.$$

Proof. The following inclusion/exclusion pattern on a set of $2m + 1$ points can not be realized:

$\square$

Figure 2.1: $2m+1$ points with an unrealizable labeling.

This is the archetypal example of a formula which is VC linear. We formally define VC linearity as follows.

Definition [1] Let $\psi(\bar{x}, \tilde{y})$ be any formula with object variables $\bar{x}$ and parameter variables $\tilde{y}$. We say that ψ is VC linear if there is a $K \in \omega$ such that for all $m \in \omega$,

$$\mathrm{VC}(\bigwedge_{i=1}^{m} \psi(\bar{x}, \tilde{y}_i)) \leq Km,$$

and

$$\mathrm{VC}(\bigvee_{i=1}^{m} \psi(\bar{x}, \tilde{y}_i)) \leq Km.$$

We will say that a complete theory T is VC linear, if every formula is VC linear with respect to T.

Proposition 1.1.5 and Theorem 2.6.1 show that this definition indeed refines the class of dependent formulas.

2.2 R-modules are VC linear

The model theory of R-modules has been studied by [3, 40, 30], and others. The natural language of R-modules, $L_R = \{+, 0, \{\cdot_r : r \in R\}\}$, is clearly dependent on the pre-chosen ring R; each scalar corresponds to a unique function symbol in

[1]We always work with respect to some complete theory, even when one is not mentioned.

the language.[2]

This hard-coding of scalars has the side-effect of making it impossible to parameterize families in terms of "slope"; as a consequence any definable family consists of a class of subsets which are, in a sense, parallel. The benefit is that families which are definable in such a system are extremely simple, which allows for the following theorem.

Theorem 2.2.1. *Let R any ring, and T the L-theory of an R module in the language $L = \{+, 0, \{\cdot_r : r \in R\}\}$. Then T is a VC linear theory.*

We must show that if $\varphi(\bar{x}, \tilde{y})$ is any L-formula, $\varphi(\bar{x}, \tilde{y})$ is VC linear with respect to T.

To prove this result, we use a quantifier elimination result stated in [3].

Definition A formula $\varphi(\bar{x}) = \varphi(x_1, \ldots, x_n)$ is a *positive primitive* (p.p.) formula, if it is equivalent in T_R to a formula of the form:

$$(\exists \bar{z}) \bigwedge_{i<p} \left(\sum_{j<m} a_{i,j} z_j + \sum_{k<n} b_{i,k} x_k = 0 \right),$$

for some $m, n, p < \omega$, where the a's and b's are scalars.

All formulas in R-modules are equivalent to a boolean combination of p.p. formulas. We summarize some important properties of p.p. formulas, described in more detail in [3, 40, 30].

- A p.p. formula asserts the solvability of a finite system of linear equations.
- Any p.p. formula $\varphi(x_1, \ldots, x_n)$ with no parameters defines a subgroup of $\mathcal{M}^n$ for some (any) model $\mathcal{M}$ of T_R.

[2]Axioms of course specify that the functions behave in the expected way, satisfying the usual axioms of R-modules.

- For any p.p. formula $\varphi(\bar{x}, \tilde{y})$ with parameters, $\varphi(\bar{x}, \tilde{0})$ also defines a subgroup, where $\tilde{0}$ denotes a vector of 0's. Furthermore, if $\tilde{a} \in \mathcal{M}$ is such that $\varphi(\bar{x}, \tilde{a})$ is consistent, then there is a $\bar{b} \in \mathcal{M}$ such that $\varphi(\bar{x}, \tilde{a}) = \bar{b} + \varphi(\mathcal{M}, \tilde{0})$. That is, $\varphi(\bar{x}, \tilde{a})$ is a translate, or coset, of the subgroup $\varphi(\mathcal{M}, \tilde{0})$.

A corollary to the above is:

Lemma 2.2.2. *For any p.p formula $\varphi(\bar{x}, \tilde{y})$, any $\tilde{a}, \tilde{b} \in \mathcal{M}$, either $\varphi(\bar{x}, \tilde{a})$ and $\varphi(\bar{x}, \tilde{b})$ are equivalent, or else they are inconsistent.*

We now give our first combinatorial lemma towards the proof of the theorem.

Lemma 2.2.3. *Let $\psi_1(\bar{x}, \tilde{y}_1), \ldots, \psi_k(\bar{x}, \tilde{y}_k), \rho_1(\bar{x}, \tilde{w}_1), \ldots, \rho_l(\bar{x}, \tilde{w}_l)$ be p.p. formulas, $k, l \in \omega$. For any $m \in \omega$,*

$$VC\left(\bigvee_{i=1}^{m}\left(\bigwedge_{j=1}^{k} \neg\psi_j(\bar{x}, \tilde{y}_{j,i}) \wedge \bigwedge_{r=1}^{l} \rho_r(\bar{x}, \tilde{w}_{r,i})\right)\right) \leq m\xi(k) + 1,$$

where $\xi(k) = 1 + k + k(k-1) + \cdots + k!$.

Proof. Note that for any k, $(\xi(k) - 1)/k = \xi(k-1)$. Suppose, by way of contradiction, that $A \subseteq \mathcal{M}$ is shattered, $|A| > m\xi(k) + 1$, $\mathcal{M} \models T$. For readability, we denote sets defined by instances of $\neg\psi_j(\bar{x}, \tilde{y}_{j,i})$ as $B_{j,i}$, and sets defined by instances of $\rho_r(\bar{x}, \tilde{w}_{r,i})$ as $C_{r,i}$. We denote intersection $R \cap Q$ as RQ.

Since A is shattered, there must be $B_{1,1}, \ldots, B_{k,m}, C_{1,1}, \ldots, C_{l,m}$, such that

$$A\left(B_{1,1} \cdots B_{k,1} C_{1,1} \cdots C_{l,1} \cup \cdots \cup B_{1,m} \cdots B_{k,m} C_{1,m} \cdots C_{l,m}\right) = A.$$

Without loss of generality, and by the pigeon-hole principle, $B_{1,1} \cdots B_{k,1} C_{1,1} \cdots C_{l,1}$ contains some $A^1 \subseteq A$, $|A^1| > \xi(k)$. Suppose $A^1 = \{a^1_0, \ldots, a^1_{\xi(k)}\}$. Again since A is shattered, there must be some $B^1_1, \ldots, B^1_k, C^1_1, \ldots, C^1_l$ such that

$$B^1_1 \cdots B^1_k C^1_1 \cdots C^1_l A^1 = \{a^1_0\}.$$

By Lemma 2.2.2, for all $s = 1, \ldots, l$, $C_s^1 = C_{s,1}$. Since $|A^1 \setminus \{a_0^1\}| > \xi(k) - 1$, without loss B_1^1 omits more than $(\xi(k)-1)/k = \xi(k-1)$ elements of $A^1 \setminus \{a_0^1\}$. Let these be denoted by $A^2 = \{a_0^2, \ldots, a_{\xi(k-1)}^2\}$. Since A is shattered, there must now be $B_1^2, \ldots, B_k^2, C_1^2, \ldots, C_l^2$ such that

$$B_1^2 \cdots B_k^2 C_1^2 \cdots C_l^2 A^2 = \{a_0^2\}.$$

Again, by Lemma 2.2.2, for all $s = 1, \ldots, l$, $C_s^2 = C_{s,1}$. Then $|A^2 \setminus \{a_0^2\}| > \xi(k-1) - 1$. We claim $B_1^2 \supseteq A^2 \setminus \{a_0^2\}$. For if B_1^2 omits, say $a^* \in A^2 \setminus \{a_0^2\}$, then by Lemma 2.2.2, $B_1^2 = B_1^1$. Then $a_0^2 \in B_1^1$ $\otimes$. Therefore, without loss, B_2^2 omits $> (\xi(k-1)-1)/(k-1) = \xi(k-2)$ elements of $A^2 \setminus \{a_0^2\}$. Let these be denoted A^3.

Performing this process $k+1$ times, we get a set A^{k+1} of size $> \xi(k-k) = \xi(0) = 1$. Say $A^{k+1} = \{a_0^{k+1}, a_1^{k+1}\}$. There must be some $B_1^{k+1}, \ldots, B_k^{k+1}, C_1^{k+1}, \ldots, C_l^{k+1}$ such that

$$B_1^{k+1} \cdots B_k^{k+1} C_1^{k+1} \cdots C_l^{k+1} A^{k+1} = \{a_0^{k+1}\}.$$

But a_1^{k+1} can not be omitted by any B_i^{k+1}, contradiction. $\square$

With the following lemmas, the above argument will give VC linearity for a formula in disjunctive normal form.

Lemma 2.2.4 (Dudley)**.** *Suppose X is a set, $\mathfrak{C}, \mathfrak{A} \subseteq \mathcal{P}(X)$, $d = VC(\mathfrak{C})$, and $e = VC(\mathfrak{A})$. Then*

$$VC(\mathfrak{C} \cup \mathfrak{A}) \leq d + e + 1.$$

Lemma 2.2.5 (Pollard [29])**.** *Suppose X is a set, $\mathfrak{C} \subseteq \mathcal{P}(X)$, and $d = VC(\mathfrak{C})$. Then*

$$VC(\mathfrak{C} \sqcup \mathfrak{C}) \leq 10d.$$

Since $\mathfrak{C} \sqcup \mathfrak{A} \subseteq (\mathfrak{C} \cup \mathfrak{A}) \sqcup (\mathfrak{C} \cup \mathfrak{A})$, these two lemmas together imply that $\mathrm{VC}(\mathfrak{C} \sqcup \mathfrak{A}) \leq 10(d + e + 1)$. More generally, the following follows by an easy induction.

Lemma 2.2.6. *Suppose $\theta(\bar{x}, \tilde{y}_1, \ldots, \tilde{y}_n)$ is a simple boolean combination of formulas $\psi_i(\bar{x}, \tilde{y}_i), \ldots, \psi_n(\bar{x}, \tilde{y_n})$. That is, θ is of the form*

$$\psi_1(\bar{x}, \tilde{y}_1)^{\eta(1)} \boxdot_1 \cdots \boxdot_{n-1} \psi_n(\bar{x}, \tilde{y_n})^{\eta(n)},$$

where $\boxdot_i$ is either conjunction or disjunction, and $\eta \in {}^{[n]}2$. Then, if $d_i = VC(\psi_i(\bar{x}, \tilde{y}_i))$,

$$VC(\theta) \leq 10^{n-1}(d_1 + d_2 + \cdots + d_n + n).$$

Lemma 2.2.7. *Suppose*

$$\varphi(\bar{x}, \tilde{y}) = \bigvee_{i=1}^{n} \psi_i(\bar{x}, \tilde{z}_i),$$

where $\tilde{y} = \tilde{z}_1 \cdots \tilde{z}_n$. Assume, for all $i = 1, \ldots, n$, $\psi_i(\bar{x}, \tilde{z}_i)$ is VC linear under disjunctions, with linearity constant K^{ψ_i}. Then for all $m \in \omega$,

$$VC(\bigvee_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)) \leq 10^{n-1} m(K^{\psi_1} + K^{\psi_2} + \cdots + K^{\psi_n} + n).$$

Proof. The proof will be by induction on n. For $n = 1$, there is nothing to show..

Suppose the theorem is true for all values less than n, and that $\psi_1(\bar{x}, \tilde{z}_1), \ldots, \psi_n(\bar{x}, \tilde{z}_n)$ are VC linear under disjunction. Then:

$$\bigvee_{i=1}^{m} \bigvee_{j=1}^{n} \psi_j(\bar{x}, \tilde{z}_{j,i}) = \bigvee_{i=1}^{m} \left(\bigvee_{j=1}^{n-1} \psi_j(\bar{x}, \tilde{z}_{j,i}) \vee \psi_n(\bar{x}, \tilde{z}_{n,i}) \right)$$

$$= \bigvee_{i=1}^{m} \left(\bigvee_{j=1}^{n-1} \psi_j(\bar{x}, \tilde{z}_{j,i}) \right) \vee \bigvee_{i=1}^{m} \psi_n(\bar{x}, \tilde{z}_{n,i}).$$

Then by inductive hypothesis and Lemma 2.2.6, $\mathrm{VC}(\bigvee_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)) \leq 10(10^{n-1} m(K^{\psi_1} + K^{\psi_2} + \cdots + K^{\psi_n} + n) + K^{\psi_{n+1}} m + 1)$, and this is equal to the desired $10^n m(K^{\psi_1} + \cdots + K^{\psi_{n+1}} + n + 1)$. □

Lemma 2.2.8. *If $\varphi(\bar{x}, \tilde{y})$ is VC linear, then so is $\neg\varphi(\bar{x}, \tilde{y})$.*

Proof.

$$\mathrm{VC}(\bigvee_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)) = \mathrm{VC}(\neg \bigvee_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)) = \mathrm{VC}(\bigwedge_{i=1}^{m} \neg\varphi(\bar{x}, \tilde{y}_i)).$$

Similarly for conjunctions. □

Proof of Theorem 2.2.1. Any formula in an R-module can be written as a boolean combination of p.p. formulas in disjunctive normal form:

$$\bigvee_{i=1}^{n} \left(\bigwedge_{j=1}^{k_i} \neg\psi_{j,i}(\bar{x}, \tilde{y}_{j,i}) \wedge \bigwedge_{r=1}^{l_i} \rho_{r,i}(\bar{x}, \tilde{w}_{r,i}) \right) \qquad (*)$$

Lemma 2.2.3 establishes that, for any fixed i,

$$\bigwedge_{j=1}^{k_i} \neg\psi_{j,i}(\bar{x}, \tilde{y}_{j,i}) \wedge \bigwedge_{r=1}^{l_i} \rho_{r,i}(\bar{x}, \tilde{w}_{r,i})$$

is VC linear under disjunctions, with linearity constant $\xi(k_i) + 1$. Let $\theta(\bar{x}, \tilde{\tilde{v}}) := \theta(\bar{x}, \tilde{y}_{1,1}, \ldots, \tilde{y}_{n,k_n}, \tilde{w}_{1,1}, \ldots, \tilde{w}_{n,l_n})$ represent a boolean combination of p.p. formulas in disjunctive normal form, as in (*). By Lemma 2.2.7,

$$\mathrm{VC}(\bigvee_{i=1}^{m} \theta(\bar{x}, \tilde{\tilde{v}}_i) \leq 10^{n-1} m(\xi(k_1) + \cdots + \xi(k_n) + 2n - 1).$$

Since $\neg\theta$ also has a DNF, VC linearity holds under conjunctions as well, by Lemma 2.2.8. This completes the proof.

□

Corollary 2.2.9. *if $\mathcal{G} = (G, +, 0)$ is any abelian group, then the complete theory of $\mathcal{G}$ is VC linear.*

Proof. Any abelian group is a $\mathbb{Z}$-module. □

The following gives a geometric application of the theorem.

Definition By an *unclipped polytope* in $\mathbb{R}^d$ with s faces[3], $s \in \omega$, is defined to be a polytope in $\mathbb{R}^d$ with s faces, for which the hyperplane constituting each respective side is extended indefinitely.

Figure 2.2 illustrates an unclipped polytope in $\mathbb{R}^2$.

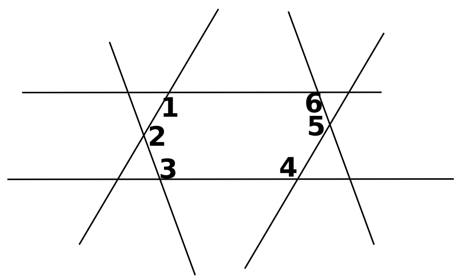

Figure 2.2: An unclipped polytope, with its vertices numbered.

Corollary 2.2.10. *Let P any unclipped polytope in $\mathbb{R}^d$, and define $\mathfrak{C}_P$ to be $\{P\}$ closed under translations and rigid dilations of P. Then $\mathfrak{C}_P$ is VC linear.*

Proof. The polytope P, with s faces of dimension $(d-1)$, can be written in the $\mathbb{R}$-module $(\mathbb{R}, +, 0, \cdot_r)_{r\in\mathbb{R}}$ as

$$\bigvee_{i=1}^{s} \psi_i(x_1, \ldots, x_d, c_i),$$

where, for each i, $\psi_i(x_1, \ldots, x_d, y_i)$ parameterizes a family of parallel hyperplanes, and where c_i represents the x_d intercept of the hyperplane. The coefficients are fixed elements of $\mathbb{R}$, which means we are considering $(\mathbb{R}, +, 0)$ as an $\mathbb{R}$ module in the language from Theorem 2.2.1.

The class resulting from $\{P\}$ closed under translation can then be written as

$$\theta(x_1, \ldots, x_d; c_1, \ldots, c_s, t_1, \ldots, t_d) = \bigvee_{i=1}^{s} \psi(x_1 + t_1, \ldots, x_d + t_d, c_i),$$

[3]We mean only the $d-1$ faces.

where $t_1, \ldots, t_d$ parameterize the translation.

Now consider a fixed translation $t_1, \ldots, t_d$ of P, denoted by $P_{\bar{t}}$. Let $P'_{\bar{t}}$ be some rigid dilation of $P_{\bar{t}}$. Consider a single $(d-1)$-face of $P_{\bar{t}}$, determined by some $\psi_i(x_1+t_1, \ldots, x_d+t_d, c_i)$. There is a corresponding $(d-1)$-face of $P'_{\bar{t}}$, which is parallel to the $(d-1)$-face of $P_{\bar{t}}$ determined by $\psi_i(x_1+t_1, \ldots, x_d+t_d, c_i)$. But the x_d intercept of the hyperplane given by $\psi_i(x_1+t_1, \ldots, x_d+t_d, c_i)$ is controlled by c_i, and as this value changes, the hyperplane defined by the formula varies across all hyperplanes parallel to it. Thus there is c'_i such that $\psi_i(x_1+t_1, \ldots, x_d+t_d, c'_i)$ defines the $(d-1)$-face of $P'_{\bar{t}}$ corresponding under the dilation to the $(d-1)$-face of $P_{\bar{t}}$ determined by $\psi_i(x_1+t_1, \ldots, x_d+t_d, c_i)$. In other words, $P'_{\bar{t}}$ is in the definable family given by

$$\theta(x_1, \ldots, x_d; y_1, \ldots, y_s, t_1, \ldots, t_d) = \bigvee_{i=1}^{s} \psi(x_1+t_1, \ldots, x_d+t_d, y_i).$$

Since $(\mathbb{R}, +, 0)$ as a module over $\mathbb{R}$ has a VC linear theory, $\mathfrak{C}_\theta$ is VC linear. Because the class mentioned in the statement of the proposition is a subclass of $\mathfrak{C}_\theta$, the conclusion follows.

□

The following may sometimes be useful:

Proposition 2.2.11. *Suppose X a set, and $\mathfrak{C}, \mathfrak{A} \subseteq \mathcal{P}(X)$ are both VC linear families. Then $\mathfrak{C} \cup \mathfrak{A}$ is also a VC linear family.*

Proof. Let $K_{\mathfrak{C}}$ and $K_{\mathfrak{A}}$ be the linear proportionality constants for $\mathfrak{C}$ and $\mathfrak{A}$, respectively. Fix $m \in \omega$. The family

$$\underbrace{(\mathfrak{C} \cup \mathfrak{A}) \sqcup \cdots \sqcup (\mathfrak{C} \cup \mathfrak{A})}_{m \text{ times}}$$

is a subfamily of

$$\Big(\underbrace{\mathfrak{C} \sqcup \cdots \sqcup \mathfrak{C}}_{m \text{ times}}\Big) \sqcup \Big(\underbrace{\mathfrak{A} \sqcup \cdots \sqcup \mathfrak{A}}_{m \text{ times}}\Big).$$

The parenthetical terms have VC dimensions bounded by, respectively, $mK_{\mathfrak{C}}$, and $mK_{\mathfrak{A}}$. Therefore, by Lemma 2.2.5, the entire expression has VC dimension at most $10m(K_{\mathfrak{C}} + K_{\mathfrak{A}} + 1)$. The case with conjunctions works similarly.

□

2.3 Special cases

Here we look at special types of VC linearity in formulas, for theories which are o-minimal, or weakly o-minimal.

Definition 1. T is an *o-minimal* theory if every dimension 1 definable subset in any model is a finite union of points and intervals.

2. T is *weakly o-minimal* if every dimension 1 definable subset in any model is a finite union of convex sets.[4]

The notion of combinatorial density (p. 21) from the previous chapter shows that sets on which a formula is not VC linear must be "thin" in a certain sense. Proposition 2.3.3 partly quantifies this intuition. The following lemma is well-known.

Lemma 2.3.1. *Suppose T is (weakly) o-minimal, $\mathcal{M} \models T$, and $\varphi(x, \tilde{y})$ a formula. For $\tilde{a} \in \mathcal{M}^{|\tilde{y}|}$, let $n_{\tilde{a}} \in \omega$ denote the number of connected components of $\varphi(\mathcal{M}, \tilde{a})$. Then there is $n^* \in \omega$ such that*

$$n^* \geq sup\{n_{\tilde{a}} : \tilde{a} \in \mathcal{M}^{|\tilde{y}|}\}.$$

[4]The boundary of any convex set may not actually be in the model.

Proof. By compactness. □

Corollary 2.3.2. *Any formula $\varphi(x, \tilde{y})$ is VC linear in a (weakly) o-minimal structure.*

Proof. The definable family associated with a length m boolean (conjunction) disjunction of φ is contained in

$$\bigsqcup_{i=1}^{mn^*} \mathfrak{C}_{Int},$$

where n^* is as in Lemma 2.3.1, and $\mathfrak{C}_{Int} = \{(a, b) : a, b \in \mathcal{M} \cup \{\infty, -\infty\}\}$.[5] Since $\mathfrak{C}_{Int}$ is VC linear, so is φ. □

Proposition 2.3.3. *Suppose T is (weakly) o-minimal, and $\mathcal{M} \models T$. Then if $A \subseteq \mathcal{M}^{|\bar{x}|}$ is such that $\varphi(\bar{x}, \tilde{y})$ is not VC linear on A, then there is no definable line l in $\mathcal{M}^{|\bar{x}|}$ containing A.*

Proof. By way of contradiction, assume the situation obtains. Let $\gamma(x) : \mathcal{M} \to \mathcal{M}^{|\bar{x}|}$ be a definable function whose image is l. Then $\gamma^{-1}(\varphi(\bar{x}, \tilde{y}))$ is a parameterized relation on $\mathcal{M}$, and is definable. By the above lemma, it is VC linear, with some linearity constant $K \in \omega$. Let $m \in \omega$ be such that $\bigvee_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)$ shatters a set $A_0 \subseteq A$ of size $mK + 1$. Let $A_0' := \gamma^{-1}(A_0)$.[6] Then A_0' is shattered by

$$\bigvee_{i=1}^{m} \gamma^{-1}(\varphi(\bar{x}, \tilde{y}_i)),$$

a contradiction. □

[5]This seems to leave out points and non-open intervals, but since every shattered set is finite, the VC linearity of this class is still sufficient to show the VC linearity of φ. If desired, $\mathfrak{C}_{Int}$ may be replaced by its Tychonov closure (p. 49), in which case the statement becomes exactly precise.

[6]It will not hurt here to assume that γ is injective. Otherwise we take a subset of the inverse image, picking out $|A_0|$ many points that all come from distinct elements of A_0.

Sontag has studied questions related to VC linearity in certain o-minimal structures [34]. VC linearity must occur in any formula in 1-variable for strongly minimal theories, because of the existence of a uniform bound much like 2.3.1. Similarly, any dependent formula will be VC linear on a set or sequence of indiscernibles, also due to uniform bounds.

2.4 Transfer techniques

Given the success of showing the VC linearity of R-modules, it is natural to ask whether we can use the same approach to show the VC linearity of the real field. This theory has a famous quantifier elimination result, which states that any formula is equivalent to a boolean combination of polynomial inequalities. Therefore to show the VC linearity of the real field, it suffices to show the VC linearity of an arbitrary boolean combination of polynomial inequalities.

Unfortunately, this question is still open. In fact it is unknown whether half spaces in $\mathbb{R}^d$ are VC linear when d is higher than 3. Surprisingly, this second seemingly simpler question is equivalent to the first. This is because of a technique known as the Veronese mapping [26], which embeds any concept class associated with a polynomial inequality into a class associated with a linear inequality in a way that preserves VC linearity. The argument is easy, but useful, so we make it explicit. For simplicity we assume we are working in the real field, but it is clear that the technique generalizes to some situations in expansions of the real field, *e.g.* boolean combinations of exponential polynomials.

Definition Let $\mathfrak{M}_{d,D}$ be an ordered set of all nonconstant monomials of degree at most D in $x_1, \ldots, x_d$, for D and d fixed natural numbers. For example if

$d = D = 2$, then $\mathfrak{M}_{2,2} = (x_1, x_2, x_1x_2, x_1^2, x_2^2)$. For any $(x_1, \ldots, x_d) \in \mathbb{R}^d$, define $f_{\mathfrak{M}}(x_1, \ldots, x_d)$ so that the i^{th} coordinate of $f_{\mathfrak{M}}(x_1, \ldots, x_d)$ is the i^{th} monomial in $\mathfrak{M}_{d,D}$. (If $d = D = 2$, then $f_{\mathfrak{M}}(x_1, x_2) = (x_1, x_2, x_1x_2, x_1^2, x_2^2)$.) Then $f_{\mathfrak{M}} : \mathbb{R}^d \to \mathbb{R}^{|\mathfrak{M}|}$ is an injective mapping, called the Veronese mapping.

Proposition 2.4.1. *Let $\mathcal{M}, \mathcal{N}$ be L-structures, with $\mathcal{M} \subseteq \mathcal{N}$. Suppose $R(\bar{x}, \tilde{y})$ is some quantifier free formula. Then if R is VC linear in $\mathcal{N}$, it is also VC linear in $\mathcal{M}$.*

Proof. Since $\mathcal{M} \subseteq \mathcal{N}$, $\mathcal{N}_M \models \Delta(\mathcal{M})$, the basic diagram of $\mathcal{M}$. $\Delta(\mathcal{M})$ includes any quantifier free combination of R, with constants from $\mathcal{M}$. By way of contradiction, assume that for any $K \in \omega$, there is some $m_K \in \omega$ such that

$$\mathcal{M} \models \mathrm{VC}(\bigvee_{i=1}^{m_K} R(x, y_i)) > Km_K.$$

For each K, let $A_K \subseteq \mathcal{M}$ be a set of size more than Km_K which is shattered by $\bigvee_{i=1}^{m_K} R(x, y_i)$ with respect to $Th(\mathcal{M})$. Define $\theta_K(x, y_1, y_2, \ldots, y_{m_K}) = \bigvee_{i=1}^{m_K} R(x, y_i)$.

Let $B_K \subseteq \mathcal{M}^{m_K}$ be the parameters used in the shattering of A_K. Suppose $B_K = \{\tilde{b}_w : w \subseteq A_K\}$, where for any $a \in A_K$, $\mathcal{M} \models \theta_K(a, \tilde{b}_w)$ if and only if $a \in w$.

Then

$$\psi_K := \bigwedge_{w \subseteq A_K} \bigwedge_{a \in w} \theta_K(a, \tilde{b}_w) \wedge \bigwedge_{a \in A_k \setminus w} \neg\theta_K(a, \tilde{b}_w)$$

is a sentence in the basic diagram of $\mathcal{M}$. Thus $\mathcal{N}_{\mathcal{M}} \models \psi_K$. But this means that R is not VC linear in $Th(\mathcal{N})$. For if it were, with linearity constant K^*, this would contradict ψ_{K^*}.

Since R was assumed to be VC linear in $Th(\mathcal{N})$, this is a contradiction. □

Definition Given a formula $\varphi(\bar{x}, \tilde{y})$, and model $\mathcal{M}$, define the bipartite graph associated with $\varphi(\bar{x}, \tilde{y})$ on $\mathcal{M}$, denoted $\mathfrak{P}_\varphi(\mathcal{M})$, as a 2-sorted bipartite graph structure (X, Y, R), where $X = \mathcal{M}^{|\bar{x}|}$, $Y = \mathcal{M}^{|\tilde{y}|}$, and for all $a \in X$ and $b \in Y$,

$$\mathfrak{P}_\varphi(\mathcal{M}) \models R(a, b) \iff \mathcal{M} \models \varphi(a, b).$$

The only difference, "platonically" between a bipartite graph and a concept class is that a bipartite graph may not be extensional, in the sense that two distinct elements of, say, X, may be connected to the same elements of Y, whereas if two elements of a concept class contain the same points, they must be identical. This will not be relevant for our purposes; we introduce graphs only because they provide a convenient way to define embeddings between definable families. This was implicitly done in [4], and the present discussion owes much to that paper.

Definition Let $\varphi(\bar{x}, \tilde{y})$ a L_1 formula, $\psi(\bar{z}, \tilde{w})$ a L_2 formula, $\mathcal{M}$ a L_1 structure and $\mathcal{N}$ a L_2 structure. Say that φ is embeddable, as a definable family, into ψ, denoted $\varphi(\bar{x}, \tilde{y}) \subseteq_{\mathcal{M},\mathcal{N}} \psi(\bar{z}, \tilde{w})$, if $\mathfrak{P}_\varphi(\mathcal{M})$ is isomorphically embeddable into $\mathfrak{P}_\psi(\mathcal{N})$. That is, there is a function $f : X \biguplus Y \to Z \biguplus W$, such that

$$\mathcal{M} \models \varphi(\bar{x}, \tilde{y}) \iff \mathcal{N} \models \psi(f(\bar{x}), f(\tilde{y})).$$

Lemma 2.4.2. *Let $\varphi(\bar{x}, \tilde{y})$, $\psi(\bar{z}, \tilde{w})$ be formulas, and suppose $\varphi(\bar{x}, \tilde{y}) \subseteq_{\mathcal{M},\mathcal{N}} \psi(\bar{z}, \tilde{w})$. Then if ψ is VC linear, then φ is VC linear.*

Proof. Proposition 2.4.1. □

Proposition 2.4.3. *The following are equivalent:*

1. *All concept classes of the form $\mathfrak{C}_{\varphi(\bar{x},\tilde{y})}(\mathbb{R})$ for a polynomial inequality $\varphi(\bar{x}, \tilde{y})$ in variables $\bar{x}$ with coefficient parameter variables $\tilde{y}$ are VC linear.*

2. *All concept classes of the form* $\mathfrak{C}_{\varphi(\bar{x},\tilde{y})}(\mathbb{R})$ *for a linear inequality* $\varphi(\bar{x},\tilde{y})$ *in variables* $\bar{x}$ *with scalar parameter variables* $\tilde{y}$ *are VC linear.*

Proof. The implication (1) $\Longrightarrow$ (2) is obvious. We show the converse.

Let $k = |\mathfrak{M}_{d,D}|$, and $p(x_1, \ldots, x_d; y_0, \ldots, y_k) \in R[x_1, \ldots, x_d]$ be a polynomial of degree at most D, with coefficient parameters $y_0, \ldots, y_k$, and take

$$\varphi(\bar{x}, \tilde{y}) = p(x_1, \ldots, x_d; y_0, \ldots, y_k) \geq 0.$$

We must demonstrate that $\varphi(\bar{x}, \tilde{y})$ is VC linear. Define

$$\psi(z_1, \ldots, z_k, y_0, \ldots, y_k) = y_k z_k + y_{k-1} z_{k-1} + \cdots + y_1 z_1 + y_0 \geq 0.$$

There exists $f : \mathbb{R}^d \to \mathbb{R}^k$, a Veronese mapping which orders the monomials of $(x_1, \ldots, x_d)$ in such a way that it induces an isomorphic embedding of φ into ψ. Thus if ψ is VC linear, then so is φ. Since ψ is a linear inequality, we are done. $\square$

2.5 Euclidean half spaces

While it is unknown whether half spaces in $\mathbb{R}^d$ are VC linear if $d > 3$, it is known that half spaces are VC linear under conjunction for $d = 1, 2, 3$ [9]. A short argument, using Tychonov closures, shows that this implies VC linearity under disjunctions as well. Where the following definition uses $2 = \{0{,}1\}$ with the discrete topology, it would be permissible to substitute an arbitrary compact space, though we will not need so much generality. See [36] for more on Tychonov closures. Recall from Chapter 1 that $[\mathfrak{C}_\varphi(A)] := \{f_c : c \in \mathfrak{C}_\varphi(A)\}$, where $f_c : A \to 2$ is the characteristic function of c.

Definition Let A a set, and $\mathfrak{F} \subseteq {}^{A}2$ a collection of functions from A to $2 = \{0, 1\}$. Then $\mathfrak{F}$ has a closure, with respect to the Tychonov topology on ${}^{A}2$. This is the Tychonov closure of $\mathfrak{F}$, denoted $cl(\mathfrak{F})$.

By abuse of notation, we identify $cl(\mathfrak{C})$ and $cl([\mathfrak{C}])$. Item (1) in Proposition 2.5 was observed in [11].

Proposition 2.5.1. *Let X a set, and $\mathfrak{C} \subseteq \mathcal{P}(X)$ of finite VC dimension, and let $\bar{\mathfrak{C}}$ be any $\mathcal{D}$-maximal extension of $\mathfrak{C}$.*[7]

1. $cl(\bar{\mathfrak{C}}) = \bar{\mathfrak{C}}$

2. $VC(\mathfrak{C}) = VC(cl(\mathfrak{C}))$

3. *for all $m \in \omega$*

$$VC(\sqcap_{i=1}^{m}\mathfrak{C}) = VC(\sqcap_{i=1}^{m} cl(\mathfrak{C})). \qquad (*)$$

Proof of 1. We must show that $[\bar{\mathfrak{C}}]$ is closed as a subset of ${}^{X}2$. Let $f \in {}^{X}2 \setminus [\bar{\mathfrak{C}}]$. Suppose $\mathrm{VC}(\mathfrak{C}) = d$. Then $[\bar{\mathfrak{C}}] \cup \{f\}$ shatters a set $B \subseteq X$, $|B| = d + 1$. Let $U_{f\restriction_B} = \{g \in {}^{X}2 : g \restriction_B = f \restriction_B\}$. Clearly $U_{f\restriction_B}$ is an open set separating $\{f\}$ and $[\bar{\mathfrak{C}}]$. Thus the complement of $[\bar{\mathfrak{C}}]$ is open, and $[\bar{\mathfrak{C}}]$ is closed. □

Proof of 2. Since $\mathrm{VC}(\mathfrak{C}) = \mathrm{VC}(\bar{\mathfrak{C}})$, and $\mathfrak{C} \subseteq cl(\mathfrak{C}) \subseteq \bar{\mathfrak{C}}$, this is clear by (1). □

Proof of 3. The "$\leq$" direction is clear. Fix $m \in \omega$. We must show $\mathrm{VC}(\sqcap_{i=1}^{m}\mathfrak{C}) \geq \mathrm{VC}(\sqcap_{i=1}^{m} cl(\mathfrak{C}))$. Let $A \subseteq X$ be a finite set, and suppose $B \subseteq A$ is cut out by $f_1 \wedge \cdots \wedge f_m$ in $\sqcap_{i=1}^{m} cl(\mathfrak{C}))$, *ie* $B = \{a \in A : f_1(a) = 1 \wedge \cdots \wedge f_m(a) = 1\}$. For

[7] This exists by Proposition 4.1.2.

each $i = 1, \ldots, m$, $f_i \in [cl(\mathfrak{C})]$. However, every open set containing f_i intersects $[\mathfrak{C}]$, and so we can pick $f'_1, \ldots, f'_m$ with $f'_i \in [\mathfrak{C}]$ such that $f'_i \restriction_A = f_i \restriction_A$ for all i. Thus $B \subseteq A$ is cut out by $f'_1 \wedge \cdots \wedge f'_m$. Then if $\sqcap_{i=1}^m cl(\mathfrak{C})$ shatters A, $\sqcap_{i=1}^m \mathfrak{C}$ does also, and so $\mathrm{VC}(\sqcap_{i=1}^m \mathfrak{C}) \geq \mathrm{VC}(\sqcap_{i=1}^m cl(\mathfrak{C}))$. □

Note that the proof of (3) makes it clear that it is immaterial what boolean combination of $\mathfrak{C}$ is used in (*). A corollary of Proposition 2.5 is the following.

Corollary 2.5.2. *Fix $d \in \omega$, and let $\mathfrak{H}$ be the collection of all half spaces in $\mathbb{R}^d$, open and closed. Let $\mathfrak{C}_\Downarrow$ be the collection of all closed downward facing half spaces in $\mathbb{R}^d$. Then $\mathfrak{H}$ is VC linear iff $\mathfrak{C}_\Downarrow$ is VC linear under conjunctions.*

Proof. The "$\Longrightarrow$" direction is clear.

Suppose $\mathfrak{C}_\Downarrow$ is VC linear under conjunctions. Let $\mathfrak{C}_\Uparrow$ denote the set of all closed upward facing half spaces in $\mathbb{R}^d$, and $\mathfrak{C}^o_\Uparrow$ denote the set of all open upward facing half spaces. First we show $\mathfrak{C}_\Downarrow$ is VC linear (with respect to unions as well as intersections).

$$\mathrm{VC}(\sqcup_{i=1}^m \mathfrak{C}_\Downarrow) = \mathrm{VC}(\neg \sqcap_{i=1}^m \mathfrak{C}^o_\Uparrow) = \mathrm{VC}(\sqcap_{i=1}^m \mathfrak{C}^o_\Uparrow)$$

By Proposition 2.5, (3), this last quantity is equal to $\mathrm{VC}(\sqcap_{i=1}^m \mathfrak{C}_\Uparrow)$, since $\mathfrak{C}_\Uparrow$ and $\mathfrak{C}^o_\Uparrow$ have the same Tychonov closure. But, by symmetry,

$$\mathrm{VC}(\sqcap_{i=1}^m \mathfrak{C}_\Uparrow) = \mathrm{VC}(\sqcap_{i=1}^m \mathfrak{C}_\Downarrow).$$

This shows that $\mathfrak{C}_\Downarrow$ is VC linear. Then clearly $\mathfrak{C}_\Downarrow \cup \neg \mathfrak{C}_\Downarrow$ is VC linear by Proposition 2.2.11, noting that $\neg \mathfrak{C}_\Downarrow$ is VC linear, by Lemma 2.2.8. Since $\mathfrak{H} \subseteq cl(\mathfrak{C}_\Downarrow \cup \neg \mathfrak{C}_\Downarrow)$, and $cl(\mathfrak{C}_\Downarrow \cup \neg \mathfrak{C}_\Downarrow)$ is VC linear by Proposition 2.5, $\mathfrak{H}$ is VC linear. □

2.5.1 Low dimensions

We will use the above corollary to give a proof that half spaces are VC linear in dimension 2. The inspiration for the argument came from [9],[8] though it is different. In particular it throws out a general position requirement on the shattered points (though the general position assumption in the original paper is claimed to be unnecessary) and has an inductive quality.

Define the formula $\varphi(x_1, x_2; m, b) :=$ "$mx_1 + b \leq x_2$,", and let $\mathfrak{H}_2$ denote the set of all open and closed half spaces in $\mathbb{R}^2$.

To show that $\mathfrak{H}_2$ is VC linear, it suffices to show that φ is VC linear, by Corollary 2.5.2. Note that $\mathfrak{C}_\varphi(\mathbb{R})$ is the set of all downward facing closed half spaces in two dimensional Euclidean space.

Proposition 2.5.3. *There exists a $K \in \omega$ such that for all $m \in \omega$,*

$$VC(\bigwedge_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i) \leq Km.$$

Proof. Fix $m \in \omega$. Clearly the intersection of m half spaces in $\mathbb{R}^2$ is a convex polytope. Let X, $|X| \geq 3$, be a set of points shattered by $\bigwedge_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)$. We claim that every point in X must lie on the boundary of the convex hull of X, $\texttt{conv}(X)$. Suppose not, and let $p \in X$ be a witness. Then it is clear that no element of $\mathfrak{C}_{\bigwedge_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)}(\mathbb{R})$ can realize the subset $X \setminus p$ of X, contradicting the assumption that X is shattered.

Since all elements of $\mathfrak{C}_{\bigwedge_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)}(\mathbb{R})$ are unbounded below, no two points of X are on the same vertical line. In fact we can say more.

[8] The paper shows that the VC dimension of the intersection of m half planes is $2m + 1$, crediting Welzl.

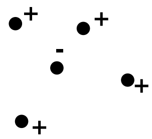

Figure 2.3: An unrealizable non-convex configuration.

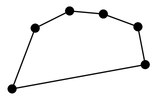

Figure 2.4: The 1-dimensional complex G.

Let G denote the boundary of $\texttt{conv}(X)$, viewed as a 1 dimensional complex, with points of X acting as vertices. Then, since X is convex, G is either a loop or else a tree with branching factor 1 (in the event that X is colinear). Suppose that G is a loop, and let L denote the lower envelope of G. That is, L consists of all the points p of G such that a vertical ray emanating downwards from p does not intersect G. We argue that L contains exactly two vertices, or equivalently, two elements of X.

Let v_l and v_r denote the leftmost and rightmost points of L; these must obviously be vertices of G. L may be thought of as a connected path, and so we can speak of elements between v_l and v_r on the path L. Suppose, by way of contradiction, that there is a vertex v' between v_l and v_r in L. Then because every element of $\mathfrak{C}_{\bigwedge_{i=1}^m \varphi(\bar{x},\tilde{y}_i)}(\mathbb{R})$ is convex and unbounded below, any element of $\mathfrak{C}_{\bigwedge_{i=1}^m \varphi(\bar{x},\tilde{y}_i)}(\mathbb{R})$ which contains v_l and v_r must also contain v', and consequently X is not shattered. This is a contradiction.

Thus, without loss of generality, tossing out the edge between v_l and v_r if necessary, G can be regarded as a path with no vertical edges. Let $\pi : \mathbb{R}^2 \to \mathbb{R}$ be downward projection, and consider the set of points $\pi(X)$ on the real line.

Whenever a downward half-space h is intersected with the path G, the complement of the intersection is connected in G; whatever set of vertices is cut out by h in G, the complementary set of points can be cut out of $\pi(X)$ by an interval I. Moreover, the set of vertices cut out of G by the intersection of some k downward half-spaces $h_1, \ldots, h_k$, the complementary set of points can be cut out of $\pi(X)$ by a union of the corresponding intervals $I_1, \ldots, I_k$. Thus if X is shattered by the conjunction of m downward facing half spaces, $\pi(X)$ is shattered by the disjunction of m intervals. Therefore the linearity of intervals implies the linearity of downward facing half spaces in $\mathbb{R}^2$.

$\square$

A less serpentine argument for the above can be found in [9]. The authors use a general position assumption on X, which they claim can be obviated by a perturbation argument; at any rate, the assumption does not seem necessary, provided the reader is willing to do extra work. They also prove the analogous fact for dimension 3, using a beautiful application of the 4-color theorem, the idea for which they attribute to E. Welzl.

2.6 A stable family which is not VC linear

We will define a structure $\mathfrak{M}$ in the language of a single 2-sorted relation $L = \{R(x, y)\}$ such that for any $K \in \omega$ there exists some $m \in \omega$, with

$$\mathfrak{M} \models \mathrm{VC}(\bigvee_{i=0}^{m} \varphi(x, y_i)) \geq Km \qquad (*)$$

where $\varphi = R^{\mathfrak{M}}$, and φ is stable with respect to $T = Th(\mathfrak{M})$. Since $\mathfrak{M}$ is 2-sorted, we will think of the universe M as $X \uplus Y$, the disjoint union of the x and y variable sorts, respectively.

The existence of such a structure establishes:

Theorem 2.6.1. *There exists a stable definable family which is not VC linear.*

2.6.1 A finite case of non-linearity in VC dimension

The argument will use a construction from [12].

Definition An *Eisenstat-Angluin random L-structure* $\mathfrak{M}_n$ is constructed as follows.

1. The X component of the universe of $\mathfrak{M}_n$ is $[n2^n] = \{1, 2, \dots, n2^n\}$.

2. For each singleton $i \in [n2^n]$, define a new element $y_{\{i\}}$ of the Y sort by $\varphi(\mathfrak{M}_n, y_{\{i\}}) = \{i\}$. Additionally, define $y_\emptyset$ by $\varphi(\mathfrak{M}_n, y_\emptyset) = \emptyset$.

3. For each $b = 1, \dots, \ulcorner \log n \urcorner$, define $Y_b = \{y_b^1, \dots, y_b^t\}$, where $t = 2^{(2+o(1))n}$, such that $\varphi(\mathfrak{M}_n, y_b^1), \dots, \varphi(\mathfrak{M}_n, y_b^t)$ are independently selected random subsets (of X) of size $\ulcorner n/b \urcorner$.

4. Take $Y := \{y_\emptyset\} \cup \{y_{\{i\}} : i \in [n2^n]\} \cup \bigcup_{b=0}^{\ulcorner \log n \urcorner} Y_b$.

In their paper, Eisenstat and Angluin establish that an E-A random structure has the following properties.

Theorem 2.6.2. *Let $\mathfrak{M}_n$ denote an E-A random L-structure. Then*

1. *For some $\alpha \in (0, 1]$ and all sufficiently large $n \in \omega$, $\mathfrak{M}_n$ models the sentence expressing*
$$VC(\bigvee_{i=0}^{n} \varphi(x, y_i)) \geq \alpha n \log n$$
with probability at least $1 - \ulcorner \log n \urcorner e^{-n}$.

2. *The probability that the VC dimension of φ with respect to $\mathfrak{M}_n$ is more than 5 is at most $2^{-(1+o(1))n}$.*

The second part of the above theorem follows from:

Lemma 2.6.3. *Let $\mathfrak{M}_n$ denote an E-A random L-structure. Then the probability that $\mathfrak{M}_n$ models the sentence expressing $\exists c \exists d\, [|\varphi(\mathfrak{M}_n, c) \wedge \varphi(\mathfrak{M}_n, d)| \geq 5]$ is at most $2^{-(1+o(1))n}$.*

This establishes that for sufficiently large n, we can actually find structures with the properties listed in Theorem 2.6.2. This is used in the theorem below.

Theorem 2.6.4. *For some $\alpha \in (0, 1]$, there is a strictly increasing $f : \omega \to \omega$, and a sequence of models $\langle \mathfrak{M}_{f(i)} : i \in \omega \rangle$ such that for all $i \in \omega$,*

$$\mathfrak{M}_{f(i)} \models \left[VC(\bigvee_{j=0}^{f(i)} \varphi(x, y_j)) \geq \alpha f(i) \log f(i) \right]$$

and

$$\mathfrak{M}_{f(i)} \models \neg \exists c \exists d \left[|\varphi(\mathfrak{M}_{f(i)}, c) \wedge \varphi(\mathfrak{M}_{f(i)}, d)| \geq 5 \right]$$

.

2.6.2 Application

We now give an infinite model and a stable formula which is not VC linear in the model.

Let $\langle \mathfrak{M}_{f(i)} : i \in \omega \rangle$ be the sequence from Theorem 2.6.4. Let $X^{\mathfrak{M}} := \biguplus X^{\mathfrak{M}_{f(i)}}$, and $Y^{\mathfrak{M}} := \biguplus Y^{\mathfrak{M}_{f(i)}}$. Fix α as in Theorem 2.6.4.

Definition The model $\mathfrak{M}$ will be the structure with universe $X^{\mathfrak{M}} \biguplus Y^{\mathfrak{M}}$, and $\mathfrak{M} \models \varphi(x, y) \iff \exists i \in \omega \left[x \in X^{\mathfrak{M}_{f(i)}} \wedge y \in Y^{\mathfrak{M}_{f(i)}} \wedge \mathfrak{M}_{f(i)} \models \varphi(x, y) \right]$.

We now show that (*) holds for $\mathfrak{M}$, and that φ is stable with respect to $Th(\mathfrak{M})$.

Claim 1 For arbitrarily large $m \in \omega$,

$$\mathfrak{M} \models \left[VC(\bigvee_{i=0}^{m} \varphi(x, y_i)) \geq \alpha m \log m \right].$$

Proof. Fix $N \in \omega$, and choose $i \in \omega$ with $f(i) > N$. It is clear from the definition that $\mathfrak{M} \models \left[VC(\bigvee_{j=0}^{f(i)} \varphi(x, y_j)) \geq \alpha f(i) \log f(i) \right]$. □ □

Claim 2

$$\mathfrak{M} \models \neg \exists c \exists d \left[|\varphi(\mathfrak{M}, c) \wedge \varphi(\mathfrak{M}, d)| \geq 5 \right]$$

Proof. Suppose not. Let $c, d \in Y^{\mathfrak{M}}$ with $|\varphi(\mathfrak{M}, c) \wedge \varphi(\mathfrak{M}, d)| \geq 5$. Then there must be some $i \in \omega$ such that $c, d \in Y^{\mathfrak{M}_{f(i)}}$, and thus $|\varphi(\mathfrak{M}_{f(i)}, c) \wedge \varphi(\mathfrak{M}_{f(i)}, d)| \geq 5$, contradicting the choice of $\mathfrak{M}_{f(i)}$. □ □

Proposition 2.6.5. *$\varphi^{\mathfrak{M}}$ is stable.*

Proof. Let $\mathbb{C}$ be a large saturated model of $Th(\mathfrak{M})$. We will show that for any $A \subseteq \mathbb{C}$, every $C \in \mathfrak{C}_{\varphi}(A)$ is definable over A by a formula in $L \restriction_{\varphi}$.

Fix A and C as above.

Case 1: $|C|$ is finite. Suppose $C = \{a_1, \ldots, a_n\}$ for some $n \in \omega$. Let

$$\delta(x) = \bigvee_{i=1}^{n} x = a_i.$$

Case 2: $|C|$ is infinite. In this case, let $\{a_1, \ldots, a_6\} \subseteq C$, all distinct. Let

$$\delta(x) = \exists y \left(\varphi(x, y) \wedge \bigwedge_{i=1}^{6} \varphi(a_i, y) \right).$$

Now if there is some $a \in \delta(A) \setminus C$, let $b \in \mathbb{C}$ be such that

$$\models \left(\varphi(a, b) \wedge \bigwedge_{i=1}^{6} \varphi(a_i, b) \right).$$

Then $|\varphi(\mathfrak{M}, b) \cap \varphi(\mathfrak{M}, c)| \geq 6$ where $C = \varphi(\mathfrak{M}, c)$, a contradiction. Any $a \in C \setminus \delta(A)$ gives a similar contradiction.

$\square$

Claim 1 establishes that (*) holds for $\mathfrak{M}$. This completes the argument.

2.7 Linearity of independence dimension

As remarked in the first chapter, independence dimension intuitively corresponds to the maximal number of elements from a family which can be arranged in a Venn diagram.[9] In this section we show that while VC dimension is not always linear under disjunctions (conjunctions), there is a sense in which linearity does always hold for independence dimension.

Note that in the statement of the theorem below, the parameter variables are increased with each disjunction, but not the object variables. As usual, everything is with respect to some fixed complete theory T.

Theorem 2.7.1. *For any partitioned formula $\varphi(\bar{x}, \tilde{y})$, of finite independence dimension (equivalently finite VC dimension), there is a $K \in \omega$ such that for all $m \in \omega$,*

$$IN(\bigvee_{i=1}^{m} \varphi(\bar{x}, \tilde{y}_i)) \leq Km,$$

where $IN(\cdot)$ denotes the independence dimension of a formula.

We illustrate the geometric meaning of this statement by considering a class $\mathfrak{C}$ of all closed discs in $\mathbb{R}^2$. The independence dimension of $\mathfrak{C}$ is 3, because at most 3

[9] We mean, a little vaguely, that e elements can be arranged to divide the space into 2^e regions. See [17].

discs can be arranged in such a way that any inclusion/exclusion condition has a witness. This corresponds to $m = 1$. For $m = 2$, we consider how many matched pairs of discs can be arranged in such a way that any inclusion/exclusion condition *on the pairs* has a witness. Higher values of m involve matched collections of m discs. A similar question is asked in [17], Theorem 2.

The theorem will be proved with the aid of a series of lemmas. As we go along, the reader may wish to observe that while the theorem is stated for a simple disjunction, the statement holds for any boolean combination of φ in which the parameters are increased and the object variables fixed.

Lemma 2.7.2. *Let X a set, and $A \subseteq X^m$, finite, with $m \in \omega$. For any $i = 1, \ldots, m$, let A_i denote the projection of A onto the i^{th} coordinate. Put $A^* = \bigcup_{i=1}^m A_i$. Then $|A^*| \leq m|A|$.*

Proof. Each A_i has size at most $|A|$. Thus the disjoint union, and hence the union, of the A_i has size bounded by $m|A|$. □

For the rest of the section we take $\theta_m(\bar{x}, \tilde{y}_1, \ldots, \tilde{y}_m) = \bigvee_{i=1}^m \varphi(\bar{x}, \tilde{y}_i)$. The dual of $\theta_m(\bar{x}, \tilde{y}_1, \ldots, \tilde{y}_m)$, denoted $\theta_m^*(\tilde{y}_1, \ldots, \tilde{y}_m, \bar{x})$, has object variables $\tilde{y}_1, \ldots, \tilde{y}_m$ and parameter variables $\bar{x}$. We fix $e = IN(\varphi)$ for the remainder of the section.

Lemma 2.7.3. *Let $\mathcal{M} \models T$, and $A \subseteq \mathcal{M}^{m|\tilde{y}|}$, $n = |A|$, $k = |\tilde{y}|$. Then*

$$|\mathfrak{C}_{\theta_m^*(\tilde{y}_1, \ldots, \tilde{y}_m, \bar{x})}(A)| \leq \sum_{i=0}^{e} \binom{(mkn)^k}{i}.$$

Proof. **Claim 1:** $|\mathfrak{C}_{\theta_m^*(\tilde{y}_1, \ldots, \tilde{y}_m, \bar{x})}(A)| \leq |S_{\theta_m(\bar{x}, \tilde{y}_1, \ldots, \tilde{y}_m)}(A^*)|$.

This is true because there is an injection given by $\theta_m^*(A, \bar{a}) \longmapsto tp_{\theta_m}(\bar{a}/A^*)$, where

$$tp_{\theta_m}(\bar{a}/A^*) := \{\theta_m(\bar{x}, \tilde{a_1}, \ldots, \tilde{a_m}) : \tilde{a_i} \in A^{*|\tilde{y}|}, \models \theta_m(\bar{a}, \tilde{a_1}, \ldots, \tilde{a_m})\} \cup$$

$$\{\neg\theta_m(\bar{x},\tilde{a_1},\dots,\tilde{a_m}) : \tilde{a_i} \in A^{*|\tilde{y}|}, \models \neg\theta_m(\bar{a},\tilde{a_1},\dots,\tilde{a_m})\}.$$

Claim 2: $|S_{\theta_m(\bar{x},\tilde{y}_1,\dots,\tilde{y}_m)}(A^*)| \leq |S_{\varphi(\bar{x},\tilde{y})}(A^*)|$

Because if $tp_\varphi(\bar{a}/A^*) = tp_\varphi(\bar{b}/A^*)$, then $tp_{\theta_m}(\bar{a}/A^*) = tp_{\theta_m}(\bar{b}/A^*)$.

Claim 3: $|S_{\varphi(\bar{x},\tilde{y})}(A^*)| \leq |\mathfrak{C}_{\varphi^*(\tilde{y},\bar{x})}((A^*)^k)|$

This is Proposition 1.1.6.

Claim 4: $|\mathfrak{C}_{\varphi^*(\tilde{y},\bar{x})}((A^*)^k)| \leq \sum_{i=0}^{e} \binom{(mkn)^k}{i}$

By the Sauer-Shelah lemma, $|\mathfrak{C}_{\varphi^*(\tilde{y},\bar{x})}((A^*)^k)| \leq \sum_{i=0}^{d} \binom{|A^*|^k}{i}$, where $d = \mathrm{VC}(\mathfrak{C}_{\varphi^*(\tilde{y},\bar{x})}((A^*)^k))$. But $d \leq e$ by Proposition 1.1.7. By Lemma 2.7.2, $|A^*|^k \leq (mkn)^k$. □

Let $k = |\tilde{y}|$ as in the above.

Lemma 2.7.4. *There is $K_0 \in \omega$ such that, for all $m \in \omega$, and all $A \subseteq \mathcal{M}^{m|\tilde{y}|}$ with $|A| = n$, $|\mathfrak{C}_{\theta^*_m(\tilde{y}_1,\dots,\tilde{y}_m,\bar{x})}(A)| \leq K_0(mn)^{ke}$.*

Proof. It is well known that, for fixed d, $\sum_{i=0}^{d} \binom{n}{i}$ is $\mathcal{O}(n^d)$. Thus by Lemma 2.7.3 there is $K_0' \in \omega$ such that, for all $m \in \omega$, and all $A \subseteq \mathcal{M}^{m|\tilde{y}|}$ with $|A| = n$,

$$|\mathfrak{C}_{\theta^*_m(\tilde{y}_1,\dots,\tilde{y}_m,\bar{x})}(A)| \leq \sum_{i=0}^{e} \binom{(mkn)^k}{i} \leq K_0'(mkn)^{ke}.$$

Now take $K_0 := K_0' k^{ke}$. □

Fix K_0 as in the above lemma.

Proof of Theorem 2.7.1. Let $N \in \omega$ be least such that $2^N > K_0 N^{2ke}$. Define $K = N$. Suppose, by way of contradiction, that $\mathrm{IN}(\theta_m(\bar{x},\tilde{y}_1,\dots,\tilde{y}_m)) \geq Km$. Let $A \subseteq \mathcal{M}^{m|\tilde{y}|}$ be a set of Km parameters witnessing the independence dimension. Then $|\mathfrak{C}_{\theta^*_m(\tilde{y}_1,\dots,\tilde{y}_m,\bar{x})}(A)| = 2^{Km}$. Then by Lemma 2.7.4,

$$2^{Km} \leq K_0(m(Km))^{ke} = K_0(m^2K)^{ke} \leq K_0(mK)^{2ke}.$$

On the other hand, since $Km > N$,

$$2^{Km} > K_0(mK)^{2ke},$$

a contradiction. □

2.8 Remaining questions

Many interesting questions remain to be asked about VC linearity of formulas and theories. Foremost, while we have shown that there is a stable family which is superlinear in VC dimension, we have not shown there is a stable *theory* which is not VC linear. In fact the theory given in our (essentially unique) counterexample may not even be dependent. Thus it is natural to ask for an example of a non VC linear theory which is dependent.

In particular, it is interesting to ask whether an ordered structure, such as RCOF, could be VC linear. The question should first be resolved for the theory of dense linear order. At the time of this writing, the author knows of no argument showing the VC linearity even of axis parallel rectangles; this would clearly follow from the VC linearity of DLO.

Regarding the VC linearity of half spaces in high dimensions, it would be sufficient to show (following the proof for the VC linearity of half spaces in dimension 3) that, for any euclidean dimension d, there is a uniform finite bound on chromatic number of any graph corresponding to the surface of a polytope in $\mathbb{R}^d$. For $d = 3$, this is the 4-color theorem.

Finally, there is a question of whether linearity is likely to hold in a broad class of families (such as all families in all dependent theories). There are several phenomena in discrete geometry that have easy log-linear upper bounds on complexity, which in fact grow at a rate strictly between linear and log-linear. Davenport-Schinzel sequences [26] provide an example. These can be made to correspond to the complexity (number of alternations) of the lower envelope of n line segments in the plane. It has been shown that the rate of growth in this situation is $n\alpha(n)$, where

$$\alpha(n) = min\{k \geq 1 : A(k) \geq n\},$$

and $A(k)$ is the Ackermann function. This gives an example of a geometric situation in which the log-linear rate is much too high, but the true rate is more subtle than simple linearity.

Chapter 3

Compression Schemes

3.1 Introduction

In this chapter we discuss a notion of sample compression schemes from the literature on computational learning theory. The use of compression schemes, and some classical examples, are given in [13, 14, 22]. Additionally, there is a short appendix, on page 92. We focus solely on model-theoretic aspects of compression schemes in this section.

In section 3.1.1 we give basic definitions, and a few interesting properties. The main results of this chapter are stated in sections 3.2 and 3.3. In section 3.2 it is shown that any stable definable family has a compression scheme of finite size. Section 3.3 gives a slightly more informative proof for the special case in which the theory does not have the finite cover property (is NFCP). A topological characterization of what we call consistent compressibility is given in section 3.4, along with a few properties. There is a brief discussion of how compressibility behaves over boolean combinations of formulas in section 3.5, and the final section lists possible avenues for further exploration.

We state for clarity that neither of the bounds given for the size of a compression scheme in this chapter (one arising from compactness, the other from NFCP) relate to the VC dimension of the family considered.

3.1.1 Basic definitions

We begin with a few definitions. Let $\varphi(x, y)$ a formula.[1] For $A \subseteq \mathbb{C}$,

$$tp_{\varphi^*}(b/A) := \{\psi(y) :\models \psi(b), \text{ and } \psi(y) \in \{\varphi^*(y, a), \neg\varphi^*(y, a)\}, \text{ some } a \in A\}.$$

Definition For any complete theory T, with monster model $\mathbb{C}$, any formula $\varphi(x; y)$, and any $A, B \subseteq \mathbb{C}$,

$$S^B_{\varphi^*}(A) := \{tp_{\varphi^*}(y/A) :\models tp_{\varphi^*}(b/A) \text{ for some } b \in B\}.$$

Let $A, B \subseteq \mathbb{C}$, and $e \in \omega$.

Let

$$S^B_{\varphi^*}(A)^{<\omega} := \bigcup\{S^B_{\varphi^*}(C) : C \subseteq_\omega A\},$$

and for $e \in \omega$,

$$S^B_{\varphi^*}(A)^{\leq e} := \bigcup\{S^B_{\varphi^*}(C) : C \subseteq_{e+1} A\}.$$

Define, for any $\eta \in {}^A 2$,

$$r_\eta(y) = \{\varphi^*(y, a)^{\eta(a)} : a \in A\}.$$

Note that r_η is a possibly inconsistent φ^*-type over A. (Later we will draw some extra conclusions assuming r_η is consistent.)

[1] We use object and parameter variables of arity 1 for simplicity; all results are true for all finite arities of variables.

Definition A partitioned formula $\varphi(x;y)$ is called *e-compressible* over A with realizers from B if there are two functions

$$comp : S^B_{\varphi^*}(A)^{<\omega} \to S^B_{\varphi^*}(A)^{\leq e}$$

and

$$ext : S^B_{\varphi^*}(A)^{\leq e} \to \{r_\eta : \eta \in {}^A 2\},$$

such that for any $p \in S^B_{\varphi^*}(A)^{<\omega}$, $comp(p) \subseteq p \subseteq ext \circ comp(p)$. If no B is specified, the realizers are assumed to come from $\mathbb{C}$.

Additionally, $\varphi(x;y)$ is *consistently e-compressible* if there is such an *ext* whose range is $S_{\varphi^*}(A)$.

Thus φ compresses over A just in case every complete φ^*-type over a finite subset of A is essentially determined by some subtype of size at most e. The above definition captures the intuitive idea of compression and expansion that was part of the original conception of Littlestone and Warmuth. However, it sometimes simplifies things to suppress the *comp* function. This observation is from [4].

Proposition 3.1.1. *Suppose there is a function* $ext : S^B_{\varphi^*}(A)^{\leq e} \to \{r_\eta : \eta \in {}^A 2\}$ *such that for every finite* $C \subseteq A$*, and every* $p \in S^B_{\varphi*}(C)$*, there is some* $C_0 \subseteq_{e+1} C$ *such that* $ext(p \restriction C_0) \supseteq p$*. Then* φ *e-compresses over* A *with realizers from* B*.*

Proof. Clear from the definitions. □

Proposition 3.1.2. *If* $A' \subseteq A$ *and* $B' \subseteq B$*, then if* $S^B_{\varphi^*}(A)$ *is e-compressible, then so is* $S^{B'}_{\varphi^*}(A')$*.*

Proof. From the definitions. □

From [4], we also have:

Proposition 3.1.3. *$\varphi(x;y)$ e-compresses over A if and only if $\varphi(x;y)$ e-compresses over every finite $B \subseteq A$.*

Proof. "$\Longrightarrow$" Trivial.

"$\Longleftarrow$": Let $\mathcal{M} \models T$, $A \subseteq \mathcal{M}$. We may assume without loss that the language is $L = \{X, Y, \varphi(x;y)\}$, and that $X^{\mathcal{M}} = A$, and $Y^{\mathcal{M}} = M$.[2]

Add to the language of $\mathcal{M}$ a set of new constant symbols for A, and a new relation symbol $H(x_1, \ldots, x_e, z_1, ..., z_e, x)$.

For $B \subseteq_\omega A$, let θ_B be the following formula. For readability we write $\bar{a} \frown \bar{b}$ for $(a_1, ..., a_e, b_1, ..., b_e)$.

$$\forall y \in Y \Bigg(\bigvee_{\bar{a}\frown\bar{b} \in B^{2e}} \bigwedge_{i=1}^{e} (a_i = b_i \iff \varphi^*(y, a_i)) \wedge$$

$$\bigwedge_{c \in B} [H(a_1, \ldots, a_e, b_1, ..., b_e, c) \iff \varphi^*(y, c)] \Bigg)$$

This says that for every realizer of a φ^*-type over B, there is some labeled example set of size e, such that H, instantiated with the labeled example, defines the type of the realizer over B. By "labeled" we mean that the b_i variables code the truth value of φ on the respective a_i parameters.

Let $Perm(e) = \{\sigma \in {}^{[e]}[e] : \sigma \text{ is a permutation}\}$, and Let ψ_B be the formula

$$\bigwedge_{\bar{a}\frown\bar{b} \in B^{2e}} \bigwedge_{\sigma \in Perm(e)} (H(a_1, \ldots, a_e, b_1, ..., b_e, x) \iff$$

$$H(a_{\sigma(1)}, \ldots, a_{\sigma(e)}, b_{\sigma(1)}, ..., b_{\sigma(e)}, x))$$

This says that the order of the labeled examples does not influence H.

[2]If this is not the case, we can perform the subsequent argument in a new language and structure where the assumptions hold, as is done in 1.2.4.

Let $\Sigma = \{\theta_B : B \subseteq_\omega A\} \cup \{\psi_B : B \subseteq_\omega A\} \cup \Delta_{A,M}(\mathcal{M})$. By the hypothesis (and the compactness theorem), Σ is consistent, and there is some $\mathcal{M}'$ with $\mathcal{M} \subseteq \mathcal{M}'$ such that $\mathcal{M}' \models \Sigma$. Then $H^{\mathcal{M}'}$ defines an e compression on $S_{\varphi^*}^{\mathcal{M}'}(A)$. Since $\varphi(x; y)$ is atomic, $S_{\varphi^*}^{\mathcal{M}}(A) \subseteq S_{\varphi^*}^{\mathcal{M}'}(A)$, and thus $\varphi(x; y)$ e-compresses over A, by 3.1.2. □

This shows that e-compressibility is a "property of the theory." In other words,

Corollary 3.1.4. *For any complete theory T, formula $\varphi(x, y)$, and $\mathcal{M}, \mathcal{N} \models T$, $\varphi(x, y)$ is e-compressible over $\mathcal{M}$ iff $\varphi(x, y)$ is e-compressible over $\mathcal{N}$.*

Proof. Suppose $\varphi(x, y)$ is not e-compressible over $\mathcal{M}$. Then there must be some finite $C \subseteq \mathcal{M}$ such that $\varphi(x, y)$ is not e-compressible over C. Since C is finite, it is possible to pick a finite subset $B \subseteq \mathcal{M}$ such that $S_{\varphi^*}^{B}(C) = S_{\varphi^*}(C)$.

By elementary equivalence, there are $C', B' \subseteq \mathcal{N}$ such that $S_{\varphi^*}^{B}(C)$ and $S_{\varphi^*}^{B'}(C')$ are isomorphic as set systems.[3] Thus $S_{\varphi^*}^{B'}(C')$ is not e-compressible, and the result follows. □

3.2 Compressibility of Stable Families

The compressibility of stable definable families is essentially a consequence of the existence of "uniform definitions" for elements in the Stone space of a stable formula. The existence of such definitions is a standard result in stability theory. For convenience, we give a short exposition of definability (which can be found also in also [32], and [3]).

We give these definitions in terms of an arbitrary formula, $\delta(x, y)$, to avoid confusion concerning φ and φ^*.

[3]Specifically, we can write them as structures in the language $(X, Y, R(x, y))$, given in 1.2.4.

Definition Given a type $p \in S_{\delta(x;y)}(A)$, we say that p is δ-definable over B if there is a formula $\psi(y)$, with parameters from B, such that:

$$\forall a \in A\,(\models \psi(a) \iff \delta(x;a) \in p)$$

Definition We say that $\psi(y,\bar{z})$ is a uniform definition for $S_{\delta(x;y)}(A)$ over B if for every $p \in S_{\delta(x;y)}(A)$ there is a tuple $\bar{a}_{p,A} \in B^{|\bar{z}|}$ such that $\psi(y,\bar{a}_{p,A})$ is a δ-definition of p.

Shelah established in [32] that either kind of definability of types is equivalent (where A is arbitrary) to the stability of δ. It will be made clear below that the uniform defining formula $\psi(y,\bar{z})$ depends only on $\delta(x;y)$, not on A or p.

The type definitions we will be concerned with are based around important finite subsets of the types they define. Loosely speaking, these subtypes restrict the possible extensions of p just as strongly as p itself does.

We make this precise by introducing the notion of a φ-tree [18].

A φ-tree over p is a binary tree of consistent but mutually exclusive extensions of the type p using instances of φ. The following formal definition is rather compact; a more verbose statement may be found in [18, 3, 32].

Definition Given a formula $\varphi(x;y)$ and a type p, a full φ-tree on p of height n is a collection of formulas $\bigcup\{r_\eta(x_\eta) : \eta \in {}^n2\}$ over a set of parameters $\{a_\mu : \mu \in {}^{<n}2\}$, where for all μ, $a_{\mu\frown 0} = a_{\mu\frown 1}$, and where $r_\eta(x_\eta) = \{\varphi(x_\eta, a_{\eta\restriction k})^{\eta(k)} : k = 0,1,2,\ldots,n-1\}$, and $p(x_\eta) \cup r_\eta(x_\eta)$ is consistent for all $\eta \in {}^n2$.

If T_φ is a full φ-tree on p, let $ht(T_\varphi)$ represent its height.

Definition Given any type p, the *φ-depth* of p, denoted φ-dp, is *sup* $\{ht(T_\varphi) : T_\varphi$ is a full φ-tree on $p\}$. If no p is specified, φ-dp is assumed to mean the *φ-depth* of $\{x = x\}$.

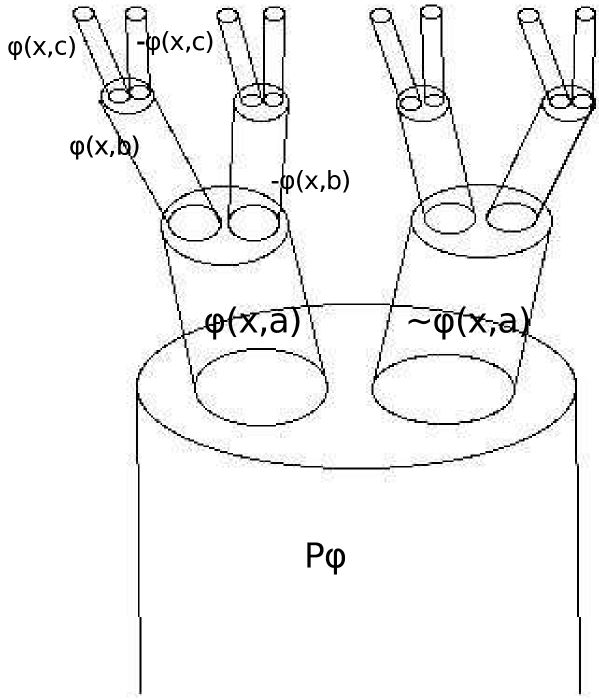

Figure 3.1: φ-depth

It turns out that finite φ-*depth* is one of the many characterizations of stability.

Proposition 3.2.1. *$\varphi(x; y)$ is stable iff φ-dp is finite.*

Proof. Assume $\varphi(x; y)$ is stable, and by way of contradiction, that φ-dp is infinite. Then by compactness, there exists a countably infinite $B \subseteq \mathbb{C}$ such that there is a complete binary φ-tree on B of height ω. Each of the $2^{\aleph_0}$ paths through this tree gives a consistent φ type on some subset of B. As types, each of the paths are pairwise inconsistent. Extending each of the types given by a path to a complete φ type over B shows that $|S_\varphi(B)| > |B|$.

By Lemma 1.3.3, it follows that there is some infinite $A \subseteq \mathbb{C}$ such that $|\mathfrak{C}_\varphi(A)| > |A|$. But, by stability of φ, every element of $\mathfrak{C}_\varphi(A)$ is definable by a formula in $L \restriction_\varphi$ with parameters from $|A|$. Thus we get $|\mathfrak{C}_\varphi(A)| \leq |A|$. This is a contradiction, and so φ-dp is finite.

The reverse direction is shown in Lemma 3.2.3.

□

Furthermore, for any type p, φ-dp p is at most φ-dp. Thus if φ is stable, then every φ type has finite φ-*depth*.

We need the following to build a defining formula for a type in the Stone space of a stable φ.

Proposition 3.2.2. *For any formula $\varphi(x;y)$ and type p, there is a finite $p_0 \subseteq p$ such that φ-dp $p = \varphi$-dp p_0.*

Proof. If φ-dp p is infinite, this is obvious by compactness. Assume that there is $n \in \omega$ such that φ-dp $p = n$.

By way of contradiction, assume that every finite $p_0 \subseteq p$ has φ-dp $p_0 \geq n+1$.

By adding constants $A = \{a_\mu : \mu \in {}^{<n+1}2\} \cup \{b_\eta : \eta \in {}^{n+1}2\}$ to the language, we can construct a set of sentences whose consistency expresses that φ-dp p is at least $n+1$. Namely, let $\Sigma = \{p(b_\eta) : \eta \in {}^{n+1}2\} \cup \{r_\eta(b_\eta) : \eta \in {}^{n+1}2\}$, where $r_\eta(b_\eta) = \{\varphi(b_\eta, a_{\eta\restriction i})^{\eta(i)} : i < n+1\}$.

By hypothesis, Σ is finitely consistent, and hence consistent by compactness. Thus any model of $T \cup \Sigma$, reducted to L, shows φ-dp $p > n$. This is a contradiction. □

The crucial observations for building a compression (type definition) can be listed as fairly self-evident facts. The second assertion is justified by noting that its negation would add another level to the maximal φ-tree on p. In order for this to be a contradiction, we need the assumption that φ-depth of p is finite.

1. The finiteness of $p_0 \subseteq p$, φ-dp $p_0 = \varphi$-dp p, allows the statement "φ-dp $p = e$" to be written in first order form.

2. If p is over A, then for all $a \in A$ exactly one of $p_0 \cup \{\varphi(x,a)\}$ and $p_0 \cup \{\neg\varphi(x,a)\}$ has *φ-depth* equal to φ-dp p.

We can now construct a δ-definition for an arbitrary $p \in S_{\delta(x,y)}(A)$, provided δ is stable.

Definition Let $\delta(x;y)$ a formula. For any $n \in \omega$, $d \in \omega$, and $\eta \in {}^n2$, a (n,d,η) *depth expansion formula for* δ is a formula

$$\psi_{n,d,\eta}(y;y_1,\dots,y_n) = \theta_{\eta,d}(y_1,\dots,y_n) \wedge \neg\theta_{\eta,d+1}(y_1,\dots,y_n) \wedge \gamma_{\eta,d}(y_1,\dots,y_n,y),$$

where for any $m \in \omega$,

$$\theta_{\eta,m}(y_1,\dots,y_n) = \exists_{(\epsilon\in{}^m2)}x_\epsilon\exists_{(\mu\in{}^{<m}2)}z_\mu \bigwedge_{\mu\in{}^{<m-1}2} z_{\mu\frown 0} = z_{\mu\frown 1}\wedge$$

$$\bigwedge_{\epsilon\in{}^m2}\left(\bigwedge_{i=1}^{n}\delta(x_\epsilon,y_i)^{\eta(i)}\right)\wedge\bigwedge_{\epsilon\in{}^m2}\bigwedge_{i=1}^{m}\delta(x_\epsilon,z_{\epsilon\restriction i})^{\epsilon(i)},$$

and

$$\gamma_{\eta,m}(y_1,\dots,y_n,y) = \theta_{\eta\frown 1,m}(y_1,\dots,y_n,y)$$

That is, $\theta_{\eta,d}(y_1,\dots,y_n)$ says that the boolean combination of δ-formulas corresponding to η, instantiated appropriately with the y_i parameters, has δ-depth at least d. The formula $\gamma_{\eta,d}(y_1,\dots,y_n,y)$ says that the boolean combination of δ-formulas corresponding to η, conjoined with the new instance $\delta(x,y)$, still has δ-depth d. The entire formula $\psi_{n,d,\eta}(y;y_1,\dots,y_n)$ then defines the unique extension (possibly inconsistent) of the δ-type determined by η and the parameters $y_1,\dots,y_n$, which has the same δ-depth.

A formula is simply a *depth expansion formula* if it is a depth expansion formula for some choice of (n,d,η).

Lemma 3.2.3. *Suppose $\delta(x,y)$ and $p \in S_{\delta(x,y)}(A)$ are such that δ-dp p is finite. Then p has a δ-definition over A, which is a depth expansion formula for δ.*

Proof. Let $d = \delta$-dp p, $p_0 \subseteq_\omega p$ of δ-depth d. Let $\rho_0(x, a_1, a_2, \ldots, a_n)$ be the conjunction of the finitely many δ-formulas in p_0. In particular,

$$\rho_0(x, a_1, a_2, \ldots, a_n) = \bigwedge_{i=0}^{n} \delta(x, a_i)^{\eta(i)},$$

where $\eta : [n] \to \{0, 1\}$ by $i \longmapsto 1$ if $\delta(x, a_i) \in p_0$, and $i \longmapsto 0$ if $\neg\delta(x, a_i) \in p_0$.

It is clear from definition 3.2 that we can construct $\psi(y, z_1, z_2, \ldots, z_n)$, a first order formula, such that $\forall y \forall z_1 \ldots \forall z_n [\models \psi(y, z_1, \ldots, z_n) \iff \rho_0(x, z_1, \ldots, z_n) \wedge \delta(x; y)$ has depth $d]$. And clearly $\rho_0(x, a_1, \ldots, a_n) \wedge \delta(x, b)$ has depth d if and only if $p_0 \cup \{\delta(x; b)\}$ has δ-depth equal to δ-dp p. Thus

$$\forall b \in A \, [\models \psi(b, a_1, \ldots, a_n) \iff \delta(x; b) \in p] \, .$$

In other words, $\psi(y, a_1, \ldots, a_n)$ is a δ-definition of p over A. Moreover, ψ is a (n, d, η) depth expansion formula for δ. □

Definition A δ-definition $\psi(y, a_1, \ldots, a_l)$, with parameters from A, of a type $p \in S_{\delta(x,y)}(A)$ is said to be *set-parameterized modulo δ* if for any permutation $\sigma \in {}^{[l]}[l]$ such that for all $i = 1, \ldots, l$, $\delta(x, a_i) \in p \iff \delta(x, a_{\sigma(i)}) \in p$,

$$\models \psi(y, a_1, \ldots, a_l) \iff \psi(y, a_{\sigma(1)}, \ldots, a_{\sigma(l)}).$$

Proposition 3.2.4. *For any A, any type $p \in S_{\delta(x,y)}(A)$, any δ depth expansion formula which serves as a δ-definition for p is set-parameterized modulo δ.*

Proof. Let $\psi_{n,d,\eta}(y; y_1, \ldots, y_n)$ be a (n, d, η) depth expansion formula for $\delta(x; y)$, and suppose for $\bar{a} \in A^n$, $\psi_{n,d,\eta}(y; a_1, \ldots, a_n)$ defines p_δ. Let $\sigma \in {}^{[n]}[n]$ be a

δ-preserving permutation (with respect to p). Then since

$$\bigwedge_{i=0}^{n} \delta(x, a_i)^{\eta(i)} \equiv \bigwedge_{i=0}^{n} \delta(x, a_{\sigma(i)})^{\eta(\sigma(i))},$$

it is clear from the definition that $\psi_{n,d,\eta}(y; a_1, \ldots, a_n) \equiv \psi_{n,d,\eta\circ\sigma}(y; a_{\sigma(1)}, \ldots, a_{\sigma(n)})$.

□

Lemma 3.2.5 (Shelah)**.** *For any stable $\delta(x, y)$, there is a finite set Δ of depth expansion formulas for δ such that for any set A, $|A| \geq 2$, and any $p \in S_{\delta(x,y)}(A)$, p is δ-defined by some instance of a Δ-formula with parameters from A.*

Proof. We show that there is a finite set of formulas $\Delta(y) = \{\psi_i(y, \bar{z}_i) : i = 1, 2, \ldots, m\}$, all depth expansion formulas for δ, such that every $p \in S_{\delta(x;y)}(A)$ is defined by some $\psi_i(y, \bar{a}_p)$, $\bar{a}_p \in A^{|\bar{z}_i|}$.

Suppose not. We will use compactness to produce a set B and a $q \in S_{\delta(x;y)}(B)$ which is not defined by a depth expansion formula for δ, contradicting Lemma 3.2.3.

Let c a new constant symbol and P a new one-place predicate. For any not necessarily finite set Δ of depth expansion formulas for δ, $\psi(y, \bar{z}_\psi)$, let

$$T_\Delta = T \cup$$

$$\left\{ \neg\exists z_1, \ldots, \exists z_{|\bar{z}_\psi|} \left[\bigwedge_{l=1}^{|\bar{z}_\psi|} P(z_l) \wedge \forall y \left[P(y) \to \Big(\delta(c, y) \equiv \psi(y, z_1, \ldots, z_{|\bar{z}_\psi|}) \Big) \right] \right] : \psi \in \Delta \right\}.$$

Now T_Δ must be consistent. Thus if Δ_* is the set of all depth expansion formulas for δ, T_{Δ_*} is consistent. Let $\mathcal{M} \models T_{\Delta_*}$, $B = P^{\mathcal{M}}$. Put $\mathcal{M}'$ as the reduct of $\mathcal{M}$ to L, and let $q = tp_\delta^{\mathcal{M}'}(c^{\mathcal{M}}/B)$.[4] Then q is not δ-defined by any

[4]This is the δ-type of the $\mathcal{M}$ interpretation of c, with parameters in the instances of δ coming from B.

depth expansion formula for δ. This contradiction establishes the existence of the desired finite Δ.

$\square$

The finite Δ produced in Lemma 3.2.5 may be encoded into a single uniform defining formula, as illustrated in the proof of Proposition 3.3.2. Since this will not simplify the argument in the following theorem, we work with Δ.

Theorem 3.2.6. *For any formula $\varphi(x; y)$, if φ is stable for T, then there exists $e \in \omega$ such that φ is e-compressible over all sets A.*

Proof. Let Δ as in Lemma 3.2.5. Let $e = max\{n : \psi \in \Delta \text{ is a } (n, d, \eta) \text{ depth expansion formula for } \varphi^*\}$. Fix a set A, any $p \in S_{\varphi^*}(A)$, and let $B \subseteq A$, finite. Then there is $\bar{b}_{p\restriction B} \in B^n$, $n \leq e$, such that for some $\psi_i \in \Delta$, $\psi_i(x, \bar{b}_{p\restriction B})$ defines $p \restriction B$. Let $comp(p \restriction B)$ be $p_0 = p \restriction dom(\bar{b}_{p\restriction B})$. This defines *comp*. It is important for the expansion to note that $\psi_i(x, \bar{b}_{p\restriction B})$, assuming it is a (n, d, η) depth expansion formula, actually asserts that φ^*-dp $p_0 = d$.

Now let $p \restriction B_0 = comp(p \restriction B)$ for $B_0 \subseteq_{e+1} B$ be given. We must define $ext(p \restriction B_0)$. Suppose $B_0 = \{b_1, \ldots, b_n\}$. Then $n = |B_0|$. Let $d = \varphi^*$-dp $p \restriction B_0$, and let $\eta \in {}^n2$ be defined by $\eta(i) = 1$ iff $\varphi^*(y, b_i) \in p \restriction B_0$. By the definition of *comp*, there must be some[5] (n, d, η) depth expansion formula for φ^* in Δ, say $\psi_{(n,d,\eta)}(x, x_1, \ldots, x_n)$. (Possibly we need to apply a φ^*-preserving permutation to η, and the ordering on B_0, but we have shown this doesn't matter in Proposition 3.2.4.) Then let $ext(p \restriction B_0)$ be the type defined by $\psi_{(n,d,\eta)}(x, b_1, \ldots, b_n)$. $\square$

[5]This formula will be unique up to equivalence.

3.3 The Finite Cover Property

Here we improve the result from the previous section. In the event that the stable theory under consideration is also NFCP, the size of the compression scheme can be given by a naturally occurring constant.

The reader may recall an interesting result from combinatorial geometry, due to Eduard Helly [26].

Theorem 3.3.1. [6] *[Helly's Theorem] If $\{X_i : i \in \kappa\}$ is a family of convex subsets of $\mathbb{R}^d$, $d \in \omega$, and $\omega > \kappa > d$, and for every subfamily $X_{i_1}, \ldots, X_{i_{d+1}}$,*

$$\bigcap_{j=1}^{d+1} X_{i_j} \neq \varnothing,$$

then $\bigcap\{X_i : i \in \kappa\} \neq \varnothing$.

Theories without the finite cover property are characterized by an analogous behavior in their definable families.

Definition Say that a formula $\varphi(x; y)$ *does not have the finite cover property* (is NFCP) if there exists some $k \in \omega$ such that for any set Γ of positive instances of φ, Γ is consistent if and only if every subset of size k is consistent. If $\varphi(x; y)$ is NFCP, let NFCP(φ) denote the least $k \in \omega$ which suffices in the definition of NFCP.

A theory T is said to not have the finite cover property (to be NFCP) if every formula is NFCP with respect to T. It is a classical result of Shelah that all NFCP theories are stable. Any $\aleph_1$-categorical theory is NFCP [21].

[6] The assumption that κ be finite is not necessary if the X_i are all compact, or if one is willing to have it be merely *consistent* that the intersection exists.

Proposition 3.3.2 (Shelah)**.** *Suppose T is NFCP and*

$$\Delta(x_1, \ldots, x_n) = \{\chi_i(x_1, \ldots, x_n, \bar{y}_i) : i = 1, \ldots, l\}$$

is a finite set of formulas. Then there is a $k \in \omega$ such that any set $\Gamma(x_1, \ldots, x_n)$ of positive Δ-formulas is consistent if and only if every subset of size k is consistent.

Proof. Define a formula φ by

$$\varphi(x_1, \ldots, x_n; \bar{y}_1, \ldots, \bar{y}_l, z_*, z_1, \ldots, z_l) = \bigwedge_{i=1}^{l} z_* = z_i \rightarrow \chi_i(x_1, \ldots, x_n, \bar{y}_i).$$

Since T is NFCP there is $k \in \omega$ such that any set $\Gamma'(x_1, \ldots, x_n)$ of positive instances of φ is consistent iff any subset of size k is consistent. Clearly every positive Δ-formula is equivalent to a positive instance of φ.

□

Note that in Proposition 3.3.2, the requirement that the instances of the Δ-formulas be "positive" can easily be removed by also coding the negations of the χ_i formulas into φ.

If $\Delta(x_1, \ldots, x_n)$ is a finite set of formulas as above, define NFCP(Δ) to be NFCP(φ), where φ is the associated coding formula given above.

We will show that in an NFCP theory, every definable family given by some $\varphi(x; y)$ has a compression scheme of finite size, and that in fact this size can be given in terms of NFCP($\{\varphi(x; y), \neg\varphi(x; y)\}$).

Proposition 3.3.3 (Shelah)**.** *If T is NFCP, then for any $\delta(x, y)$ there is an $e \in \omega$ such that for any A, any $p \in S_{\delta(x,y)}(A)$, there is some $p_0 \subseteq p$, of size at most e, such that δ-dp p_0 = δ-dp p. Moreover, we can choose e =NFCP($\{\delta, \neg\delta\}$).*

Proof. Suppose not. Then there is some δ such that for all $k \in \omega$ there exists a set A_k and a type $p_k \in S_{\delta(x,y)}(A_k)$ such that p_k has no subtype of cardinality k with the same δ-depth as p_k.

Choose $K \in \omega$ such that any set Γ of instances of δ is consistent if and only if it is K-consistent. We can do this by choosing $\Delta = \{\delta(x, y_1), \neg\delta(x, y_0)\}$, as in Proposition 3.3.2.

Let $p \in S_{\delta(x,y)}(A)$ be such that p has no subtype of size K with identical δ-depth.

Let $n \in \omega$ be δ-dp p. Take Σ to be the set of formulas:

$$\Sigma = \{p(x_\eta) : \eta \in {}^{n+1}2\} \cup \{r_\eta(x_\eta) : \eta \in {}^{n+1}2\},$$

Where $r_\eta(x_\eta) = \{\delta(x_\eta, a_{\eta\restriction i})^{\eta(i)} : i < n+1\}$, and $\{a_\mu : \mu \in {}^{<n+1}2\}$ is a set of new constant symbols, with $a_{\mu\frown 0} = a_{\mu\frown 1}$ for all $\mu \in {}^{<n}2$.

Then Σ is a set of instances of δ, and is consistent if and only if it is K-consistent. But Σ is K consistent, because every size K subset of p has δ-depth at least $n+1$. This is a contradiction, because Σ implies that δ-dp $p \neq n$. $\square$

Now we are able to prove the main theorem for this section:

Theorem 3.3.4. *For any formula $\varphi(x, y)$, if T is NFCP, then the definable family associated with φ has a compression scheme of finite size. In particular, it has a compression of size $e \in \omega$, where e is minimal such that any set of $\Delta = \{\varphi^*(y, x_1), \neg\varphi^*(y, x_0)\}$ formulas is consistent iff it is e-consistent.*

Proof. For concreteness, work in $S_{\varphi^*}(B)$, for some set B. Let e be such that any $\varphi^*(y; x)$-type over B has a size e subtype with the same φ^*-depth. Define the function *comp* by, for any $p \in S_{\varphi^*}(B)$, and $A \subseteq_\omega B$, $p \restriction A \longmapsto p \restriction A_0$, where A_0 is any set such that $A_0 \subseteq_{e+1} A$, and φ^*-dp $p \restriction A_0 = \varphi^*$-dp $p \restriction A$.

Define the *ext* function as follows. Let $p \restriction A_0$ be a type of size at most e, and a any parameter not in A_0. If $(p \restriction A_0) \cup \{\varphi^*(y,a)\}$ has the same φ^*-depth as $p \restriction A_0$, then put $\varphi^*(y,a)$ in $ext(p \restriction A_0)$. Otherwise, put $\neg\varphi^*(y,a)$ in $ext(p \restriction A_0)$. This defines a complete (possibly inconsistent) φ^*-type over B.

It is clear that $ext \circ comp(p \restriction A)$ extends $p \restriction A$. □

3.4 Consistent Compressions

We now make a few more remarks about consistent compressions (definition 3.1.1.) The requirement that a compression scheme be consistent does not seem to be especially restrictive. While the range of a consistent expansion function may be a strict superset of the original concept class, it will be contained in the Tychonov closure of the original class.

The next few propositions show the generality of the notion.

Proposition 3.4.1. *Let φ any formula, A any set, and consider a compression scheme on φ over A determined by the functions comp and ext. Let $p_0 \in dom(ext)$. Then if, for any finite $B \supseteq dom(p_0)$, there is some $q \in S_{\varphi^*}(A)$ such that $comp(q \restriction B) = p_0$, then $ext(p_0)$ is consistent.*

Proof. The definition then shows that $ext(p_0)$ is finitely consistent, and hence consistent. □

The compression schemes referred to by the following propositions are given in the appendix.

Proposition 3.4.2. *The compression scheme on axis-parallel rectangles is consistent.*

Proof. By Proposition 3.4.1. □

Proposition 3.4.3. *The compression scheme on intervals is consistent.*

Proof. By Proposition 3.4.1. □

In this section we make use of the topology on Stone spaces. We give a topological characterization of the existence of a consistent compression scheme.

Definition For any formula $\delta(x,y)$, and any set A, an open set U in $S_{\delta(x,y)}(A)$ is said to be $\leq n$*-open* if $U = U_{\theta(x)} = \{p \in S_{\delta(x,y)}(A) : \theta(x) \in p\}$, where θ is a quantifier free boolean combination of δ-formulas of length at most n.

Proposition 3.4.4. *Fix $\varphi(x;y)$, A, and $S_{\varphi^*}(A)$, and, for any $n \in \omega$, let $\tau_{\leq n} = \{U \subseteq S_{\varphi^*}(A) : U$ is $\leq n$-open$\}$. Then $\varphi(x;y)$ has a consistent n-compression on A if and only if there is a function $f : \tau_{\leq n} \to S_{\varphi^*}(A)$, such that*

1. *For all $U \in$dom(f), $f(U) \in U$, and*

2. *ran(f) is dense in $S_{\varphi^*}(A)$.*

Proof. Suppose there is such a function f. Let $B \subseteq A$ finite, $p \in S_{\varphi^*}(A)$. Let $J_{p\restriction B} = \{q \in S_{\varphi^*}(A) : p \restriction B \subseteq q\}$. Then $J_{p\restriction B}$ is an open set. Since the range of f is dense, there is some $\leq n$-open set $U = U_{r_0}$, r_0 a φ^*-type of size at most n, such that $f(U) \in J_{p\restriction B}$. That is, $f(U)$ extends $p \restriction B$, which implies (by 1) $r_0 = p \restriction B_0$ for some appropriate $B_0 \subseteq_{n+1} B$. If we take, for any size $\leq n$ type s, $ext(s) = f(U_s)$, then this shows that ext is as in Proposition 3.1.1. Thus this yields a consistent compression scheme of size n.

Now suppose φ has a consistent compression scheme of size n. For any φ^*-type s_0 (over A) of size at most n, take $f(U_{s_0}) = ext(s_0)$. Then since $ext(s_0)$ is a consistent type extending s_0, $f(U_{s_0}) \in U_{s_0}$. We now show $ran(f)$ is dense. Let U an open set. Then $U = \{q \in S_{\varphi^*}(A) : q \supseteq p \restriction B\}$ for some finite $B \subseteq A$, and $p \in S_{\varphi^*}(A)$. Take $p_0 = comp(p \restriction B)$, and $r = f(U_{p_0}) = ext \circ comp(p \restriction B)$. By the definition of a compression scheme, $r \supseteq p \restriction B$, and so $f(U_{p_0}) \in U$. Thus $ran(f)$ is dense. □

Thus a formula consistently n-compresses just if there is a choice function on $\leq n$-open subsets of the Stone space with a dense image.

3.5 Compressions Over Boolean Combinations

One of the interesting characteristics of a compression scheme, say of size e, on a certain family, is that it bounds the VC dimension of that family above by $5e$ [14]. Since we are concerned elsewhere with the question of how VC dimension changes over boolean combinations of families, it is worthwhile to ask an analogous question for compression schemes. We would like to know, if $\varphi(x; y)$ has a compression of finite size, whether there is some $K \in \omega$ such that $\bigvee_{i=1}^{m} \varphi(x; y_i)$ has a compression scheme of size at most Km, for all $m \in \omega$.

In general the answer to this question must be no, because

Proposition 3.5.1. *There is a stable formula $\varphi(x; y)$ such that for all $k \in \omega$ there exists some $m \in \omega$ such that $\bigvee_{i=1}^{m} \varphi(x; y_i)$ has no compression of size mk.*

Proof. Suppose not. Let $\varphi(x; y)$ be a stable formula such that for all $l \in \omega$ there is some $m \in \omega$ such that $VC(\bigvee_{i=1}^{m} \varphi(x; y_i)) > lm$. (Such a formula exists by the argument in section 2.6.) Let $k \in \omega$ be such that for all $m \in \omega$, $\bigvee_{i=1}^{m} \varphi(x; y_i)$ has a compression of size at most mk. Then for all $m \in \omega$, we get $VC(\bigvee_{i=1}^{m} \varphi(x; y_i)) \leq 5mk$, a contradiction. □

3.6 Remaining questions

A well-known conjecture of Manfred Warmuth asserts that every formula $\varphi(x; y)$ with VC dimension d has a compression of order d [14]. Since a compression scheme of finite size implies finite VC dimension, a slight weakening of this conjecture poses a model theoretically interesting question: Is it the case that dependent theories, *i.e.* theories in which every formula has finite VC dimension, are characterized by the existence of finite compression schemes for all formulas?

Similar questions present themselves for other nice classes of model theoretic structure. For example, one might ask whether all formulas in an *o*-minimal structure are finitely compressible. Note that in dimension one this is clearly the case. Is one variable sufficient?

Another interesting question involves a special type of "label-free" compression [22]. A compression is label-free if the range of the *comp* function (and the domain of *ext*) is not a collection of types, but a collection of tuples. That is, the compression forgets the truth value of $\delta(x, y)$ on the compression sets. There are currently no known counter-examples to the existence of label-free compressions of order d for all VC-classes of dimension d. However, in the algorithm for compressing stable families, knowing the truth value of the formula on the saved parameter sets is highly essential. Thus one may ask whether, in particular, stable families have label-free compression schemes.

Chapter 4

Definability and $\mathcal{D}$-maximal classes

4.1 Introduction

In this chapter, we describe the relationship between $\mathcal{S}$-maximality and $\mathcal{D}$-maximality, relating the two in Proposition 4.2.2. We show that if a definable family is $\mathcal{D}$-maximal in a structure $\mathcal{M}$, then $\mathcal{M}$ has unusually strong saturation properties. This implies that many natural definable families are not $\mathcal{D}$-maximal in any structure. On the other hand, we see that there are definable families which are $\mathcal{D}$-maximal with respect to some models, but not to others (Theorem 4.2.6). In other words, $\mathcal{D}$-maximality is not first-order.

We emphasize that $\mathcal{S}$-maximality *is* first-order.

Proposition 4.1.1. *Let T a complete theory, $\varphi(\bar{x}, \tilde{y})$ a formula of VC dimension $d \in \omega$, and $\mathcal{M}, \mathcal{N} \models T$. Then $\mathfrak{C}_\varphi(\mathcal{M})$ is $\mathcal{S}$-maximal of dimension d iff $\mathfrak{C}_\varphi(\mathcal{N})$ is $\mathcal{S}$-maximal of dimension d.*

Proof. As in Chapter 1, we use the abbreviation $\Phi_d(n) := \sum_{i=0}^{d} \binom{n}{i}$. For $n \in \omega$, let θ_n be a sentence stating that for any distinct $\bar{x}_1, \ldots, \bar{x}_n$, there exist $\tilde{y}_1, \ldots, \tilde{y}_{\Phi_d(n)}$, such that if $i \neq j$, then there exists some k such that $\not\models \varphi(\bar{x}_k, \tilde{y}_i) \iff \varphi(\bar{x}_k, \tilde{y}_j)$. Then since $\mathcal{N}$ and $\mathcal{M}$ agree on every θ_n, the statement holds. □

Recall the following definition from Chapter 1.

Definition Let X a set, and $\mathfrak{C} \subseteq \mathcal{P}(X)$, of VC dimension $d \in \omega$. We say that $\mathfrak{C}$ is *Dudley maximal*, or $\mathcal{D}$-maximal, if for any set $A \subseteq X$, $A \in \mathcal{P}(X) \setminus \mathfrak{C} \implies \mathrm{VC}(\mathfrak{C} \cup \{A\}) > d$.

Thus a definable family is $\mathcal{D}$-maximal just in case it is impossible to add any new sets to the family without increasing the VC dimension. A simple Zorn's lemma argument shows that every definable family has a (not necessarily definable) $\mathcal{D}$-maximal extension.

Proposition 4.1.2 (Dudley)**.** *Let X a set and $\mathfrak{C} \subseteq \mathcal{P}(X)$, with VC dimension d. Then there exists a $\mathcal{D}$-maximal $\mathfrak{C}' \subseteq \mathcal{P}(X)$ with $\mathfrak{C} \subseteq \mathfrak{C}'$ such that VC($\mathfrak{C}$) = VC($\mathfrak{C}'$).*

Proof. Let $\{\mathfrak{C}_\alpha\}_{\alpha\in\delta}$, for some ordinal δ, be a collection of of concept classes on X, totally ordered by inclusion, and such that

1. $\mathrm{VC}(\mathfrak{C}_\alpha) = d$ for all $\alpha \in \delta$, and
2. $\mathfrak{C}_0 = \mathfrak{C}$.

By Zorn's lemma[1] it suffices to show that if $\mathfrak{C}_\delta := \bigcup_{\alpha<\delta} \mathfrak{C}_\alpha$, then $\mathrm{VC}(\mathfrak{C}_\delta) = d$. By way of contradiction suppose that $\mathfrak{C}_\delta$ shatters $A \subseteq X$, $|A| = d + 1$. Then there exist $c_1, \ldots, c_{2^{d+1}}$ in $\mathfrak{C}_\delta$ which realize the shattering. There must then be ordinals $\iota_1, \ldots, \iota_{2^{d+1}}$, all less than δ, such that for all $i = 1, \ldots, 2^{d+1}$, $c_i \in \mathfrak{C}_{\iota_i}$. Thus A is shattered by $\mathfrak{C}_\mu$, where $\mu = sup\{\iota_i : i = 1, \ldots, 2^{d+1}\}$. By point (1), this is a contradiction.

□

[1]Zorn's lemma states that any partial order in which every chain has an upper bound has at least one maximal element.

4.2 Infinite $\mathcal{D}$-maximal and $\mathcal{S}$-maximal classes

It is clear that every finite $\mathcal{S}$-maximal class is also $\mathcal{D}$-maximal. However it is easy to have an infinite system which is $\mathcal{S}$-maximal and not $\mathcal{D}$-maximal. The first example below was discovered by Welzl and Woeginger [14].

Example 4.2.0.1. *Let X be an infinite set, and take $\mathfrak{C} = \{c \subseteq X : 0 < |c| < d+1\}$.*

The above $\mathfrak{C}$ is $\mathcal{S}$-maximal of VC dimension d, because for any finite set $X_0 \subseteq X$, $\mathfrak{C} \restriction_{X_0} = [X_0]^{\leq d}$, the set of all subsets of X_0 of size at most d. The latter class clearly has the maximum cardinality allowable by Sauer's lemma. Nevertheless, $\mathrm{VC}(\mathfrak{C}) = \mathrm{VC}(\mathfrak{C} \cup \{\emptyset\})$, and so $\mathfrak{C}$ is not $\mathcal{D}$-maximal.

Example 4.2.0.2. *Let $X = \mathbb{R}^d$ and let $\mathfrak{C}$ be the set of all positive half spaces in $\mathbb{R}^d$.*

This is a $\mathcal{S}$-maximal family (of the same VC dimension) when restricted to a certain dense $X' \subseteq X$ [13]. However, $VC(\mathfrak{C} \restriction_{X'}) = VC(\mathfrak{C}) = VC(\mathfrak{C} \cup \{X\} \cup \{\emptyset\})$. Thus $\mathfrak{C} \restriction_{X'}$ is not $\mathcal{D}$-maximal.

To see why $VC(\mathfrak{C}) = VC(\mathfrak{C} \cup \{X\} \cup \{\emptyset\})$, we must invoke the Tychonov closure proposition from page 49. By Proposition 2.5, it suffices to show that $cl(\mathfrak{C}) \supseteq (\mathfrak{C} \cup \{X\} \cup \{\emptyset\})$. We show first that $X \in cl(\mathfrak{C})$. We must show that any open set (in the Tychonov space) containing X (thought of as a characteristic function $f_X : \mathbb{R}^d \to 2$) intersects $\mathfrak{C}$, the class of all positive half spaces in $\mathbb{R}^d$. Fix a finite set $A \subseteq \mathbb{R}^d$. This gives a corresponding open set:

$$U_A^{f_X} = \{g \in {}^{\mathbb{R}^d}2 : g \restriction A = f_X \restriction A\}$$

To show that this intersects $\mathfrak{C}$ we must find a positive half space in $\mathfrak{C}$ which contains A. But it is clear that such a half space exists. Therefore $X \in cl(\mathfrak{C})$. We leave the fact that $\emptyset \in cl(\mathfrak{C})$, and the following example, to the reader.

Other examples of this type can be given by classes resulting from inequalities of polynomials.

Example 4.2.0.3. *Let $X = \mathbb{R}^d$ and let $\mathfrak{C}$ be the set of all balls in X.*

Then $\mathfrak{C}$ is a $\mathcal{S}$-maximal class [13] when restricted to a dense $X' \subseteq X$. However, if $\mathfrak{C}'$ is the set of all half spaces in $\mathbb{R}^d$, then $\mathrm{VC}(\mathfrak{C}) = \mathrm{VC}(\mathfrak{C} \cup \mathfrak{C}')$. Thus $\mathfrak{C} \restriction_{X'}$ is not $\mathcal{D}$-maximal.

The following was discovered by Dudley.

Lemma 4.2.1. *Let $\mathcal{M} \models DLO$, and $\mathfrak{C} = \{A \subseteq \mathcal{M} : A$ is convex$\}$. Then $\mathfrak{C}$ is $\mathcal{D}$-maximal.*

Proof. First note that $\mathrm{VC}(\mathfrak{C}) = 2$. Let $B \subseteq \mathcal{M}$, not convex. Then there are points $a_0, a_1, a_2 \in \mathcal{M}$ such that $a_1 < a_2 < a_3$, and $a_0, a_2 \in B$, $a_1 \notin B$. Then $\mathfrak{C} \cup \{B\}$ shatters $\{a_0, a_1, a_2\}$. □

The next proposition illustrates a general technique; we find a $\mathcal{D}$-maximal class by taking the Tychonov closure of an $\mathcal{S}$-maximal class.

Proposition 4.2.2. *The Tychonov closure of an $\mathcal{S}$-maximal class is $\mathcal{D}$-maximal.*

Proof. Suppose not. Let $\mathfrak{C}$ be $\mathcal{S}$-maximal of VC dimension d, $\bar{\mathfrak{C}}$ its Tychonov closure, and suppose there is $A \in \mathcal{P}(X) \setminus \bar{\mathfrak{C}}$, such that $\bar{\mathfrak{C}} \cup \{A\}$ has the same VC dimension as $\bar{\mathfrak{C}}$. Define f_A to be the characteristic function of A. That is, for all $x \in X$,

$$f_A(x) := \begin{cases} 1 & \text{if } x \in A; \\ 0 & \text{if } x \notin A. \end{cases}$$

Since $f_A \notin [\bar{\mathfrak{C}}]$, there must be an open set in ${}^X2$ which contains f_A, but does not intersect $\mathfrak{C}$. Let this open set be, for some $g \in {}^X2$, and some $B \subseteq X$, finite,

$$U_{g,B} = \{h \in {}^X2 : h \restriction_B = g \restriction_B\}.$$

Without loss, $|B| \geq d$. Then $f_A \restriction_B = g \restriction_B$, and there is no element $c \in \mathfrak{C}$ such that $c \cap B = A \cap B$. Since $\mathfrak{C}$ is $\mathcal{S}$-maximal, $|\mathfrak{C} \restriction_B|$ is the maximum possible to allow $\mathrm{VC}(\mathfrak{C} \restriction_B) = d$. Yet $\mathrm{VC}(\{A \cap B\} \cup \mathfrak{C} \restriction_B) = \mathrm{VC}(\mathfrak{C} \restriction_B)$. This is a contradiction. □

A number of authors have remarked [13, 14, 4] that there are no 'natural' examples of classes which are $\mathcal{D}$-maximal but not $\mathcal{S}$-maximal. For instance, the class of all convex subsets in $\mathbb{R}$ is $\mathcal{D}$-maximal, but also $\mathcal{S}$-maximal.

We submit that the scarcity of $\mathcal{D}$-maximal not $\mathcal{S}$-maximal classes is partly caused by the difficulty of finding 'natural' (*i.e.* definable) families which can be $\mathcal{D}$-maximal. These are rare for an important theoretical reason, namely that $\mathcal{D}$-maximality requires a strong saturation[2] condition on the model in which a definable family appears. Sometimes, depending on the relation giving the definable family, the condition will not be satisfied in any model (of the theory).

The main lemma, which follows, was also observed in [36].

Lemma 4.2.3. *Let $\varphi(\bar{x}, \tilde{y})$ a formula. Then, for any model $\mathcal{M}$ of T, $\mathfrak{C}_\varphi^{\mathcal{M}}(\mathcal{M}) = cl(\mathfrak{C}_\varphi^{\mathcal{M}}(\mathcal{M}))$ iff $\mathcal{M}$ realizes every type in $S_{\varphi^*}(\mathcal{M})$.*

Proof. Suppose $\mathfrak{C}_\varphi^{\mathcal{M}}(\mathcal{M}) = cl(\mathfrak{C}_\varphi^{\mathcal{M}}(\mathcal{M}))$, and let $p(\tilde{y}) \in S_{\varphi^*}(\mathcal{M})$. Pick[3] $f_p \in [S_{\varphi^*}(\mathcal{M})]$ such that

$$f_p(\bar{a}) = 1 \iff \varphi^*(\tilde{y}, \bar{a}) \in p(\tilde{y}).$$

[2]The word *saturation* in model theory refers to various situations in which a model realizes all types over (some subset of) its parameters.

[3]The existence of such an indicator function is just the definition of $[S_{\varphi^*}(\mathcal{M})]$, given in Chapter 1.

We claim that $f_p \in [\mathfrak{C}_\varphi^\mathcal{M}(\mathcal{M})]$. Since $\mathfrak{C}_\varphi^\mathcal{M}(\mathcal{M})$ is Tychonov closed, it suffices to show that for any finite $A \subseteq \mathcal{M}^{|\bar{x}|}$, $f_p \restriction_A \in [\mathfrak{C}_\varphi^\mathcal{M}(A)]$.

Fix such a finite $A = \{\bar{a_1}, \ldots, \bar{a_n}\}$. Since p is consistent,

$$\mathcal{M} \models \exists \tilde{y} \bigwedge_{i=1}^{n} \varphi^*(\tilde{y}, \bar{a}_i).$$

Let $\tilde{b} \in \mathcal{M}$ be a witness to the existential quantifier. Let $f_{\tilde{b}} : \mathcal{M} \to 2$ be defined as

$$f_{\tilde{b}}(\bar{a}) = 1 \iff \mathcal{M} \models \varphi^*(\tilde{b}, \bar{a}).$$

Then $f_p \restriction_A = f_{\tilde{b}} \restriction_A$, and so $f_p \restriction_A \in [\mathfrak{C}_\varphi^\mathcal{M}(A)]$. Therefore $f_p \in [\mathfrak{C}_\varphi^\mathcal{M}(\mathcal{M})]$. Since this is true, there must be some $\tilde{b}^* \in \mathcal{M}$ such that $f_p = f_{\tilde{b}^*}$. But then $\tilde{b}^*$ realizes p.

Conversely, suppose $\mathcal{M}$ realizes every type in $S_{\varphi^*}(\mathcal{M})$. Let $f \in {}^{\mathcal{M}^{|x|}}2$ be in $cl(\mathfrak{C}_\varphi^\mathcal{M}(\mathcal{M}))$. We must show that $f \in \mathfrak{C}_\varphi^\mathcal{M}(\mathcal{M})$. Let $p_f(\tilde{y})$ be a set of φ^* formulas over $\mathcal{M}$, defined by

$$\varphi^*(\tilde{y}, \bar{a}) \in p_f(\tilde{y}) \iff f(\bar{a}) = 1,$$

and

$$\neg\varphi^*(\tilde{y}, \bar{a}) \in p_f(\tilde{y}) \iff f(\bar{a}) = 0.$$

We must show that $p_f(\tilde{y})$ is (finitely) consistent. Let $A \subseteq \mathcal{M}^{|\bar{x}|}$, $A = \{\bar{a}_1, \ldots, \bar{a}_n\}$. Let $\eta : [n] \to 2$ be defined by

$$\eta(i) = 1 \iff \varphi^*(\tilde{y}, \bar{a}_i) \in p_f.$$

We must show

$$\mathcal{M} \models \exists \tilde{y} \bigwedge_{i=1}^{n} \varphi^*(\tilde{y}, \bar{a}_i)^{\eta(i)}.$$

By the choice of f, there is some $g \in [\mathfrak{C}_{\varphi}^{\mathcal{M}}(\mathcal{M})]$ such that $g \restriction_A = f \restriction_A$. Let $c \in \mathfrak{C}_{\varphi}^{\mathcal{M}}(\mathcal{M})$ be such that $c = g^{-1}(1)$, and $\tilde{b}$ such that $c = \varphi(\mathcal{M}, \tilde{b})$. Then

$$\mathcal{M} \models \bigwedge_{i=1}^{n} \varphi^*(\tilde{b}, \bar{a}_i)^{\eta(i)}.$$

Therefore p_f is consistent, and so, by assumption, there is a realizer of p_f, $\tilde{b}^* \in \mathcal{M}$. Then $f = f_{\tilde{b}^*}$, and so $f \in \mathfrak{C}_{\varphi}^{\mathcal{M}}(\mathcal{M})$. This shows $\mathfrak{C}_{\varphi}^{\mathcal{M}}(\mathcal{M})$ is closed, and completes the argument.

□

Above, we make reference to the Tychonov closure of $\mathfrak{C}_{\varphi}^{\mathcal{M}}(\mathcal{M})$. It is important to distinguish this from the closure of $\mathfrak{C}_{\varphi}(\mathcal{M})$. For example, let $\mathcal{M}$ a model of T, and $\mathcal{N}$ a $|\mathcal{M}|^+$-saturated[4] elementary extension of $\mathcal{M}$. Then $\mathfrak{C}_{\varphi}^{\mathcal{N}}(\mathcal{M})$ is closed in the Tychonov topology on ${}^{\mathcal{M}^{|\bar{x}|}}2$, but $\mathfrak{C}_{\varphi}^{\mathcal{M}}(\mathcal{M})$ may or may not be closed. By choice of the monster, $\mathfrak{C}_{\varphi}(\mathcal{M})$ is always closed.

Proposition 2.5 shows that any $\mathcal{D}$-maximal class is closed with respect to the Tychonov topology. Thus if a definable family in a model is not closed, it is not $\mathcal{D}$-maximal. There are many natural relations, however, which can never satisfy the saturation condition in Lemma 4.2.3.

Proposition 4.2.4. *Let $E(x, y)$ a binary relation symbol, and let T a theory which says that E is an equivalence relation with infinitely many equivalence classes. Then for any $\mathcal{M} \models T$, $\mathfrak{C}_{E(x,y)}^{\mathcal{M}}(\mathcal{M})$ is not $\mathcal{D}$-maximal.*

Proof. It suffices to show that $\mathfrak{C}_{E(x,y)}^{\mathcal{M}}(\mathcal{M})$ is not closed with respect to the Tychonov topology, or, by Lemma 4.2.3, that $\mathcal{M}$ does not realize every type in

[4]This means that $\mathcal{N}$ realizes every type in $S_{\varphi}(A)$ for every $A \subseteq_{|\mathcal{M}|^+} \mathcal{N}$, for every formula $\varphi(\bar{x}, \tilde{y})$. Thus in particular, $\mathcal{N}$ realizes every type in $S_{\varphi}(\mathcal{M})$.

$S_{\varphi^*}(\mathcal{M})$. Let

$$p(y) = \{\neg E^*(y, a) : a \in \mathcal{M}\}.$$

This is a complete E^* type over $\mathcal{M}$, consistent because T has ∞-many equivalence classes. But clearly it is not realized by any element of $\mathcal{M}$. Therefore $\mathfrak{C}^{\mathcal{M}}_{E(x,y)}(\mathcal{M})$ is not closed in ${}^{\mathcal{M}}2$, and so is not $\mathcal{D}$-maximal. $\square$

Proposition 4.2.5. *Let $T = DLO$, the theory of dense linear orders, in the language $L = \{<\}$. Let $\varphi(x, y) = y < x$. Then for any model $\mathcal{M} \models T$, $\mathfrak{C}^{\mathcal{M}}_{\varphi}(\mathcal{M})$ is not $\mathcal{D}$-maximal.*

Proof. Fix some element $a^* \in \mathcal{M}$. Let p be the type defined by

$$p(y) = \{y < a : a \geq a*\} \cup \{\neg(y < a) : a < a*\}.$$

Then $p(y)$ is consistent but not realized in $\mathcal{M}$, and so $\mathfrak{C}^{\mathcal{M}}_{\varphi}(\mathcal{M})$ is not $\mathcal{D}$-maximal. $\square$

We can ask whether a partitioned formula (with respect to some complete theory) being $\mathcal{D}$-maximal is a property of the theory, or the model. That is, if a partitioned formula is $\mathcal{D}$-maximal with respect to some model of a complete theory T, is it necessarily $\mathcal{D}$-maximal with respect to every model of T? The answer turns out to be negative.

Theorem 4.2.6. *There is a complete theory T, models $\mathcal{N}$, $\mathcal{M}$ of T, and a partitioned formula $\varphi(\bar{x}, \tilde{y})$ such that $\mathfrak{C}^{\mathcal{M}}_{\varphi}(\mathcal{M})$ is $\mathcal{D}$-maximal, but $\mathfrak{C}^{\mathcal{N}}_{\varphi}(\mathcal{N})$ is not.*

Proof. Let $CONV(\mathbb{Q})$ denote all convex subsets of the rationals. Define a language $L = \{X, Y, R(x, y)\}$, where X and Y are unary predicates, and R is a binary relation. Let T be the complete L-theory of the 2-sorted structure

$\mathcal{Q} = (\mathbb{Q}, CONV(\mathbb{Q}), R(x, y))$, where x is of the first sort, y of the second, and for any x, y,

$$\mathcal{Q} \models R(x, y) \iff x \in y.$$

Then T has one model where $R(x, y)$ is $\mathcal{D}$-maximal, namely $\mathcal{Q}$. On the other hand, by Löwenheim-Skolem, there must be some countable model $\mathcal{M}$ of T. We claim that $\mathfrak{C}_R^{\mathcal{M}}(\mathcal{M})$ is not $\mathcal{D}$-maximal.

Note that R is unstable, because it has the descending chain condition. Therefore, $|S_{R^*}(\mathcal{M})| > |\mathcal{M}|$, and so $\mathcal{M}$ can not realize every type in $S_{R^*}(\mathcal{M})$. By Lemma 4.2.3, the theorem is proved. □

4.2.1 Remaining questions

It would be interesting to characterize the relations which are "unclosable" in the sense of Lemma 4.2.3. The examples show that such families can be stable (as in Proposition 4.2.4) or unstable (Proposition 4.2.5). Any parameterized p.p formula in a pure injective R-module[5] has a type space in which all types are realized [30]. This therefore provides an example of a non-trivial family which can be closed. Pure injective R-modules seem to be a promising hunting ground for natural examples of $\mathcal{D}$-maximal but not $\mathcal{S}$-maximal classes.

[5] Any direct sum of finite cyclic groups (considered as a $\mathbb{Z}$-module) is an example of a pure injective module, as is any injective module.

Chapter 5

Conclusion

Here we will briefly summarize the results of the previous chapters, and give some open problems.

In Chapter 2, the notion of VC linearity is defined, and a series of families are shown to be VC linear, including all families definable in an R-module. We note that linearity holds in dimension 1 of o-minimal and strongly minimal structures, as well as on sets of indiscernibles, and show that there is no 1-dimensional definable cell which connects all members of a super-linear concept domain in an o-minimal structure. We show that the Veronese embedding of geometric set systems preserves linearity. Tychonov closures are introduced, and used to prove a sufficient condition for VC linearity of euclidean half spaces. A new proof is given for linearity of half planes, and a stable family is exhibited which is super-linear in VC dimension. Finally, this chapter shows that while VC dimension may increase superlinearly as the number of parameters is increased, independence dimension does not.

Chapter 3 translates the idea of compression schemes into the language of types, and shows that all set systems which are stable (in the model theoretic sense) have compression schemes of finite size. A slightly more informative argument is given in the case that the theory is NFCP. We characterize consistent compression schemes in topological terms, and discuss compressions of boolean combinations of formulas.

Chapter 4 discusses the relationship between $\mathcal{S}$-maximal and $\mathcal{D}$-maximal definable families. We show that while $\mathcal{S}$-maximality is naturally first-order, $\mathcal{D}$-maximality is not.

5.1 Open Problems

It remains to be shown whether superlinear growth in VC dimension is possible in a dependent theory. In particular, it is unknown whether superlinear growth in VC dimension is possible in DLO. The question of VC linearity in intersections of half spaces remains unsolved as well.

We also leave open the "weak Warmuth conjecture," which states that every set system of finite VC dimension has a compression scheme of finite size. If true, this would give a new characterization of dependent theories somewhat similar to the "definable types" characterization of stable theories. Less generally, we wonder whether any family in an o-minimal structure has a compression of finite size.

Appendix A

Appendix: Some Geometric Compression Schemes

In this appendix, we give several geometric examples of compression schemes. The material is taken from [13, 14].

Let X any set, and $\mathfrak{C} \subseteq \mathcal{P}(X)$. We think elements of $\mathfrak{C}$ as indicator (or characteristic) functions.

That is, for $c \in \mathfrak{C}$ and $x \in X$, $c(x) = \begin{cases} 1 & \text{if } x \in c, \\ 0 & \text{if } x \notin c. \end{cases}$

Define $\mathfrak{C}^{<\infty} = \{c \restriction_A : A \subseteq X, \text{ finite, and } c \in \mathfrak{C}\}$, and $\mathfrak{C}^{\leq d} = \{c \restriction_{A_0} : A_0 \subseteq X, |A_0| \leq d \text{ and } c \in \mathfrak{C}\}$.

Definition A family $\mathfrak{C} \subseteq \mathcal{P}(X)$, is said to have a *compression scheme of size d* if there are are two functions

$$\text{comp} : \mathfrak{C}^{<\infty} \to \mathfrak{C}^{\leq d}$$

and

$$\exp : \mathfrak{C}^{\leq d} \to \mathcal{P}(X)$$

such that for every finite $A \subseteq X$, and every $c \in \mathfrak{C}$, there exists $B \subseteq A$ such that $comp(c \restriction_A) = c \restriction_B$, and $\exp \circ \text{comp}(c \restriction_A)$ is consistent with $c \restriction_A$.

A.0.1 Axis Parallel Rectangles

A simple example of a compression scheme is found on the class of all axis parallel rectangles, illustrated in Figure A.1. For any finite set of points in $\mathbb{R}^2$, labeled consistently with some rectangle, the smallest rectangle containing the positively labeled points has the same trace on the finite set as the original rectangle. This allows for the following compression scheme:

1. The compression function saves the (at most) four points representing the topmost, leftmost, rightmost, and bottommost points enclosed in the original rectangle.

2. The expansion function returns the smallest rectangle containing the saved points.

It is then easy to see that the expansion rectangle is consistent with the original rectangle on the given finite set.

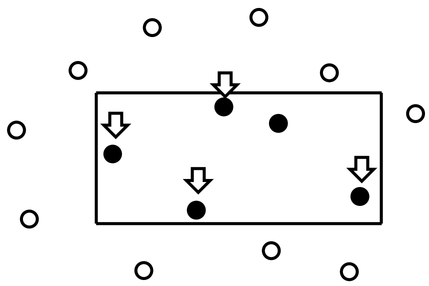

Figure A.1: Select the leftmost, rightmost, lowest and highest points included in the rectangle.

A.0.2 Intervals

There is a compression scheme of size 2 definable on the class of all intervals on a dense linear order. We may assume that the particular model in question is, say $(\mathbb{R}, <)$, and that

$$\mathfrak{C}_{Int} = \{(a, b) : a, b \in \mathbb{R}\}.$$

Now consider an element $(a, b) \in \mathfrak{C}_{Int}$ on a finite $A \subseteq \mathbb{R}$. For a compression, save the rightmost point (if one exists) in $(a, b) \cap A$, and the rightmost point (if one exists) in $\{c \in A : c \leq a\}$, both labeled accordingly.[1] Non-degenerate compression sets are then of the form $\{(c, 0), (d, 1)\}$ (see figure A.2), and can be expanded as $(c, d]$. Note that though the expansion concept is not in $\mathfrak{C}_{Int}$, it is in the Tychonov closure, $cl(\mathfrak{C}_{Int})$. The degenerate compressions expand in the obvious way. For example, $\{(d, 1)\}$ expands to $(-\infty, d]$.

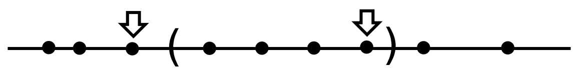

Figure A.2: Select the rightmost negative and rightmost positive examples.

Interestingly, this same scheme generalizes to a union of intervals. Let

$$\sqcup_{i=1}^{m} \mathfrak{C}_{Int} = \{(a_1, b_1) \cup \cdots \cup (a_m, b_m) : a_k, b_k \in \mathbb{R}\}.$$

Figure A.3 illustrates a compression scheme for this class. An expansion can be given by ordering the compression set, and using half open intervals in a way similar to the first example. This expansion also happens to be in the Tychonov closure of the original class.

[1] That is, the expansion function knows whether each saved point was inside or outside (a, b).

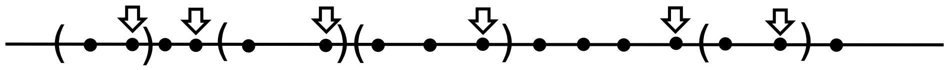

Figure A.3: Select the rightmost negative and rightmost positive examples, where they exist, for each interval.

In her thesis, Sally Floyd gives a large number of compression algorithms, and shows many concept classes to be compressible. Warmuth and Floyd demonstrate, among other things, that any $\mathcal{S}$-maximal class $\mathfrak{C}$ has a compression scheme of size at most $\mathrm{VC}(\mathfrak{C})$.

BIBLIOGRAPHY

[1] Adler, H.: Introduction to theories without the independence property, Submitted, July 2007.

[2] Assouad,P.: Densité et dimension, Ann. Inst. Fourier (Grenoble), 33 (1983), (3), 233-282.

[3] Baldwin, J. T.: Fundamentals of Stability Theory, Springer-Verlag, New York, 1988.

[4] Ben-David, S., Litman, A.: Combinatorial variability of Vapnik-Chervonenkis classes with applications to sample compression schemes, Discrete Applied Mathematics 86, (1997) 3-25.

[5] Blumer, A., Ehrenfeucht, A., Haussler, D., Warmuth, M.: Learnability and the Vapnik-Chervonenkis dimension, J. Assoc. Comput. Mach. 36(4), (1989) 929-965.

[6] Bollobás, B., Leader, I., Radcliffe, A.J., Reverse Kleitman Inequalities, Proc. London Math. Soc. (3) 58 (1989), 153-168.

[7] Buechler, S.: Essential Stability Theory, Springer, Berlin, 1991.

[8] Chang, C.C., Keisler, H.J.: Model Theory, Elsevier, Amsterdam, 1990.

[9] Dobkin, D., Gunopulos, D., Maas, W.: Concept learning with geometric hypotheses, Proceedings of the 8th Ann. Conf. on Computational Learning Theory, ACM Press, New York, NY, (1995) 329-336.

[10] Dudley, R.M.: A Course on Empirical Processes, Lecture Notes in Math., vol 1997, Springer, New York, 1984.

[11] Dudley, R.M.: Uniform Central Limit Theorems, Cambridge University Press, New York, 1999.

[12] Eisenstat, D., Angluin, D.: The VC dimension of k-fold union, Information Processing Letters, 101(5), (2007) 181-184.

[13] Floyd, Sally.: Space-bounded learning and the Vapnik-Chervonenkis dimension, Proceedings of the second annual workshop on Computational learning theory, (1989) 349-364.

[14] Floyd, S., Warmuth, M.: Sample compression, learnability, and the Vapnik-Chervonenkis dimension, Mach. Learning, 21, (1995) 1-36.

[15] Gärtner, B.: Set Systems of Bounded Vapnik-Chervonenkis Dimension and a Relation to Arrangements, PhD. Thesis, Freie Universität Berlin, 1991.

[16] Goldberg, P., Jerrum, M.: Bounding the Vapnik-Chervonenkis dimension of concept classes parameterized by real numbers, Mach. Learning, 18, (1995) 131-148.

[17] Grunbaum, B.: Venn Diagrams and Independent Families of Sets, Mathematics Magazine, 48 (1), (Jan., 1975), 12-23.

[18] Harnik, V., Harrington, L.: Fundamentals of forking, Ann. Pure Appl. Logic, 26 (1984), (3), 245-286.

[19] Karpinski M., Macintyre, A.: Polynomial bounds for VC dimension of sigmoidal and general Pfaffian neural networks, J. Comput. System Sci. 54 (1997), no. 1, part 2, 169–176

[20] Kearns, M., Vazirani U.V.: An Introduction to Computational Learning Theory, MIT Press, Boston, 1994.

[21] Keisler, H. J.: Ultraproducts which are not saturated, The Journal of Symbolic Logic 32 (1967), pp. 23-46.

[22] Kuzmin, D., Warmuth, M.: Unlabeled Compression Schemes for Maximum Classes, Lecture Notes in Comput. Sci., 3559, Springer, Berlin, 2005.

[23] Laskowski, M.C.: Vapnik-Chervonenkis classes of definable sets, J. London Math. Soc. 45 (2), (1992) 377-384.

[24] Littlestone, N., Warmuth, M., Relating Data Compression and Learnability, Unpublished Manuscript, University of California Santa Cruz, 1986.

[25] Marker, D.: Model Theory: An Introduction, Springer, New York, 2002.

[26] Matoušek, Jiří : Lectures on Discrete Geometry, Springer. New York, 2002.

[27] Pillay, A.: Geometric Stability Theory, Oxford Science Publications, Oxford, 2002.

[28] Poizat, B.: A Course in Model Theory: An Introduction to Contemporary Mathematical Logic, Springer, New York, 2000.

[29] Pollard, D.: Lecture notes at Institut Henri Poincare, Paris, 2001, http://www.stat.yale.edu/ pollard/Paris2001/lectures.html.

[30] Prest, M.: Model theory and modules, London Mathematical Society lecture note series no. 130. Cambridge University Press, Cambridge, 1988.

[31] Sauer, N.: On the density of families of sets, J. Combinatorial Theory Ser. A 13 (1972), 145–147.

[32] Shelah, S.: Classification Theory, North Holland (2nd Ed.), Amsterdam, 1990.

[33] Shelah, S.: A combinatorial problem: stability and order for models and theories in infinitary languages, Pacific J. Math. 41 (1972), 247–261.

[34] Sontag, D.: Shattering all sets of k points in 'general position' requires $(k-1)/2$ parameters, Neural Comput., 9 (1997) (2), 337-348.

[35] Stengle, G., Yukich, J.E.: Some new Vapnik-Chervonenkis classes, Ann. Statist. 17 (1989), no. 4, 1441–1446.

[36] Van den Dries, L. Limit sets in o-minimal structures, Proceedings of the RAAG Summer School Lisbon 2003: O-minimal structures, Eds M. Edmundo, D. Richardson, A.J. Wilkie (2005), 159-212.

[37] Van den Dries, L. Tame Topology and O-minimal Structures, Cambridge University Press, London, 1997.

[38] Vapnik, V., Chervonenkis, A.: On the uniform convergence of relative frequencies of events to their probabilities, Theory of Probability and its Applications, 16(2):264–280, 1971.

[39] Welzl, E.: Complete range spaces, Unpublished Notes (1987).

[40] Ziegler, M.: Model theory of modules, Ann. Pure and App. Logic, 26 (2), (1984), 149-213.