NIP Theories and Computational Learning Theory

Vincent Guingona

UNIVERSITY OF NOTRE DAME, DEPARTMENT OF MATHEMATICS, 255 HURLEY, NOTRE DAME, IN 46556
E-mail address: guingona.1@nd.edu

Contents

Forward

I hope you enjoy these lecture notes! These notes use the ideas in Hunter Johnson's graduate thesis (see [**14**]) to apply ideas from NIP theories to computational learning theory. On the NIP theory side, I mostly follow Pierre Simon's book [**24**] together with various source material ([**1**, **2**, **5**, **6**, **8**, **11**–**13**, **15**, **22**]). On the computational learning theory side, I use concepts from Jiři Matoušek's book [**17**] along with source material ([**4**, **7**, **18**]).

I attempted to write these notes to require very little background in both Model Theory and Computational Learning Theory. I wrote Chapter 1 specifically to review the material I need for the bulk of the notes. Chapter 2 is a quick overview on Computational Learning Theory, culimating in the final section which outlines the relationship to Model Theory. Chapter 3 goes through the basics of our modern understanding of NIP theories. Chapter 4 is an aside about different means of measuring the complexity of formulas in NIP theories, which relate back to measuring the complexity of concept classes.

Special thanks to Hunter Johnson for taking the time to read over these notes and giving me helpful feedback.

CHAPTER 1

Introduction

1. Notation

For these notes, we will assume knowledge of basic model theory and basic set theory. We will use the standard definition of ordinals. In particular, $0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$, ..., $\omega = \{0, 1, 2, ...\}$, etc.. We will use $<$ for the ordinal ordering, so $<=\in$. In particular, "$k < \omega$" will denote $k \in \omega$, so k is a natural number. For two sets A and B, we will let ${}^{A}B$ denote the set of all functions from A to B. For a set A, we will let $\mathcal{P}(A)$ denote the set of all subsets of A (the powerset of A). Notice that there is a natural bijection from ${}^{A}2$ to $\mathcal{P}(A)$, namely for each $f \in {}^{A}2$, associate the support $\text{supp}(f) = \{a \in A : f(a) = 1\} \in \mathcal{P}(A)$. We prefer to consider ${}^{A}2$, since every element implicitly codes A (i.e., A is the domain of each element).

If X is a set and κ is any cardinal (usually we will consider κ to be finite), then let $\binom{X}{\kappa}$ denote the set of all subsets of X of size κ. We choose this notation because, if X is an n-element set and $k < \omega$, then $\binom{X}{k}$ has cardinality

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

2. Counting

Throughout this book, we will use counting estimates. For $0 \leq k \leq n \leq \omega$, let

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

This is the number of subsets of n of size k. Since the number of subsets of n is 2^n, we get that $2^n = \sum_{k=0}^{n} \binom{n}{k}$.

Because we are concerned with $\binom{n}{k}$, we will frequently need to estimate $n!$. The following is an approximation to Stirling's Formula.

LEMMA 1.2.1. *For any integer $n > 0$,*

$$n^n e^{-n} < n! < (n+1)^{n+1} e^{-n}.$$

PROOF. Taking the logarithm, it suffices to show

$$n\log(n) - n < \log(n!) < (n+1)\log(n+1) - n.$$

For this you use integral estimates. Notice that

$$\log(1) + ... + \log(n) \leq \int_1^{n+1} \log(x)dx = (n+1)\log(n+1) - n$$

and

$$\log(1) + ... + \log(n) \geq \int_0^{n} \log(x)dx = n\log(n) - n.$$

□

This immediately gives us the following lemma:

LEMMA 1.2.2. *For all* $0 < k \leq n < \omega$,

$$\binom{n}{k} \leq \frac{n^k}{k!} \leq \left(\frac{en}{k}\right)^k.$$

PROOF. To see the first inequality, note that

$$\binom{n}{k} = \frac{n(n-1)...(n-1+k)}{k!} \leq \frac{n^k}{k!}.$$

For the second inequality, use Lemma 1.2.1 to get

$$\frac{n^k}{k!} \leq \frac{n^k}{k^k e^{-k}} = \left(\frac{en}{k}\right)^k.$$

□

For $0 < k \leq n \leq m$,

$$\frac{\binom{m}{k}}{\binom{n}{k}} = \frac{m(m-1)...(m-k+1)/k(k-1)...1}{n(n-1)...(n-k+1)/k(k-1)...1} = \left(\frac{m}{n}\right)...\left(\frac{m-k+1}{n-k+1}\right).$$

However, if $0 < i < n \leq m$, then $\frac{m}{n} \geq \frac{m-i}{n-i}$. Therefore,

$$\frac{\binom{m}{k}}{\binom{n}{k}} \leq \left(\frac{m}{n}\right)^k. \tag{1.1}$$

For $0 \leq k \leq n < \omega$, write

$$\Phi_k(n) = \sum_{i=0}^{k} \binom{n}{i}.$$

This is the number of subsets of n of size $\leq k$. Thus, $\Phi_n(n) = 2^n$ and $\Phi_0(n) = 1$.

LEMMA 1.2.3. *For all* $0 = k \leq n < \omega$,

$$\Phi_k(n) \leq n^k$$

Proof. This is a simple induction proof using the first inequality of Lemma 1.2.2. □

Another useful bound is the following. For all $x \in \mathbb{R}$,

$$1 + x \leq e^x. \tag{1.2}$$

To see why (1.2) holds, first notice that, when $x = 0$, $1 = e^0$. Now, consider $f(x) = e^x - x - 1$, hence $f(0) = 0$. Then, $f'(x) = e^x - 1$. For all $x > 0$, $f'(x) > 0$, hence f is strictly increasing, so $f(x) > 0$ for all $x > 0$. On the other hand, $f'(x) < 0$ when $x < 0$, so $f(x) > 0$ for all $x < 0$ as well.

3. Probability

Some arguments in this book require basic probability theory. We introduce that in this section. Let X be any set. A set $\mathcal{B} \subseteq \mathcal{P}(X)$ is called a *σ-algebra* if the following three properties hold

(1) $\mathcal{B} \neq \emptyset$,
(2) if $A \in \mathcal{B}$, then $(X \setminus A) \in \mathcal{B}$, and
(3) if $A_i \in \mathcal{B}$ for $i < \omega$, then $\bigcup\{A_i : i < \omega\} \in \mathcal{B}$.

Consider X with σ-algebra $\mathcal{B}$ so that $\emptyset \in \mathcal{B}$. A function $\mu : \mathcal{B} \to \mathbb{R}$ is called a *measure* if

(1) for all $A \in \mathcal{B}$, $\mu(A) \geq 0$,
(2) $\mu(\emptyset) = 0$, and
(3) if $A_i \in \mathcal{B}$ for $i < \omega$ are pairwise disjoint, then

$$\mu\left(\bigcup\{A_i : i < \omega\}\right) = \sum\{\mu(A_i) : i < \omega\}.$$

If, in addition, $\mu(X) = 1$, then we say that μ is a *probability measure* and we call $(X, \mathcal{B}, \mu)$ a *probability space*.

If $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \gamma)$ are two probability spaces, then we can consider the product probability space $(X \times Y, \mathcal{D}, \eta)$ where $\mathcal{D}$ is the σ-algebra generated by $\mathcal{A}$ and $\mathcal{B}$, and η is the unique probability measure $\eta : \mathcal{D} \to [0, 1]$ such that, for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$,

$$\eta(A \times B) = \mu(A)\gamma(B).$$

In particular, for $n < \omega$, we can consider the power probability measure on X^n and we will denote μ^n.

Fix a probability space $(X, \mathcal{B}, \mu)$. An "event" is just an element $A \in \mathcal{B}$ and the "probability" of the event A occuring is just $\mu(A)$. A *random variable* is a measurable function $f : X \to \mathbb{R}$. That is, a function f such that, for each Lebesgue measurable subset $A \subseteq \mathbb{R}$, $f^{-1}(A) \in \mathcal{B}$. The *expected value* of f is

$$\mathrm{E}(f) = \int_X f d\mu.$$

The *variance* of f is

$$\mathrm{Var}(f) = E((f - E(f))^2) = E(f^2) - E(f)^2.$$

We say that random variables $f, g : X \to \mathbb{R}$ are *uncorrelated* if $E(fg) = E(f)E(g)$. We say that random variables $f_1, ..., f_n : X \to \mathbb{R}$ are *independent* if, for all (Lebesgue) measurable sets $B_1, ..., B_n \subseteq \mathbb{R}$,

$$\mu\left(\bigcap_i \{a \in X : f_i(a) \in B_i\}\right) = \prod_i \mu\big(\{a \in X : f_i(a) \in B_i\}\big).$$

In particular, two independent random variables are uncorrelated.

LEMMA 1.3.1. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \gamma)$ be two probability spaces, $f : (X \times Y) \to \mathbb{R}$ a random variable on the product space. Then, there exists $a \in X$ so that the restricted function $f|_a : Y \to \mathbb{R}$, where $f|_a(b) = f(a, b)$, such that*

$$E(f|_a) \geq E(f).$$

PROOF. We have

$$\alpha := E(f) = \int_{X \times Y} f(\mu \times \gamma).$$

However, if, for all $a \in X$, we had $E(f|_a) < \alpha$, then

$$\int_{X \times Y} f(\mu \times \gamma) < \alpha\gamma(Y) = \alpha.$$

□

For $f \in {}^X 2$ measurable, let $\mu(f) = \mu(\{a \in X : f(a) = 1\})$.

LEMMA 1.3.2. *If $f : X \to \mathbb{R}_+$ be a random variable. Then, for all $r > 0$,*

$$\mu\big(\{a \in X : f(a) \geq r\}\big) \leq \frac{E(f)}{r}.$$

PROOF. Let $g : X \to 2$ be given by

$$g(a) = 1 \text{ if and only if } f(a) \geq r.$$

Then, $E(f) = \int_X f d\mu \geq \int_X fg d\mu \geq r\mu(g)$. The result follows. □

PROPOSITION 1.3.3 (Chebyshev's Inequality). *If $f : X \to \mathbb{R}$ is a random variable and $\epsilon > 0$, then*

$$\mu\big(\{a \in X : |f(a) - E(f)| \geq \epsilon\}\big) \leq \frac{\mathrm{Var}(f)}{\epsilon^2}.$$

PROOF. Apply Lemma 1.3.2 to the random variable $x \mapsto (f(x) - E(f))^2$ and $r = \epsilon^2$. □

4. Ultrafilters

DEFINITION 1.4.1. Let X be a set and $\mathcal{F} \subseteq \mathcal{P}(X)$. We say that $\mathcal{F}$ is an *ultrafilter* on X if

(1) $\emptyset \notin \mathcal{F}$,
(2) If $A \subseteq B \subseteq X$ and $A \in \mathcal{F}$, then $B \in \mathcal{F}$,
(3) If $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$, and
(4) For all $A \subseteq X$, either A or $X \setminus A$ is an element of $\mathcal{F}$.

An ultrafilter $\mathcal{F}$ on X is called *principal* if there exists $a \in X$ such that $A \in \mathcal{F}$ if and only if $a \in A$. Otherwise, $\mathcal{F}$ is called *non-principal*. If X is infinite, then non-principal ultrafilters exist by the axiom of choice. We show this in the proof of Proposition 1.4.2 below.

Let $\mathcal{A} \subseteq \mathcal{P}(X)$. We say that $\mathcal{A}$ has the *finite intersection property* if, for all finite $\mathcal{A}_0 \subseteq \mathcal{A}$, $\bigcap \mathcal{A}_0 \neq \emptyset$. In particular, any ultrafilter has the finite intersection property, by conditions (1), (3), and (4). For another example, if X is infinite, then the set

$$\mathrm{Cof}(X) = \{A \in \mathcal{P}(X) : (X \setminus A) \text{ is finite}\}$$

has the finite intersection property. Notice that if $\mathcal{F} \supseteq \mathrm{Cof}(X)$ is an ultrafilter on X, then $\mathcal{F}$ cannot be principal. This is because, for all $a \in X$, $(X \setminus \{a\}) \in \mathrm{Cof}(X)$.

PROPOSITION 1.4.2. *If X is infinite and $\mathcal{A} \subseteq \mathcal{P}(X)$ has the finite intersection property, then there exists $\mathcal{F} \in \mathcal{P}(X)$ such that $\mathcal{F} \supseteq \mathcal{A}$ and $\mathcal{F}$ is an ultrafilter on X.*

PROOF. We use Zorn's Lemma. Consider the partial order (under inclusion) of all subsets of $\mathcal{P}(X)$ extending $\mathcal{A}$ with the finite intersection property.

$$\mathbf{P} = \{\mathcal{B} \subseteq \mathcal{P}(X) : \mathcal{A} \subseteq \mathcal{B}, \mathcal{B} \text{ has the finite intersection property}\}.$$

For any chain $\mathbf{C} \subseteq \mathbf{P}$, consider $\mathcal{C} = \bigcup \mathbf{C}$. Clearly $\mathcal{A} \subseteq \mathcal{C}$. Moreover, for all $B_1, ..., B_n \in \mathcal{C}$, $B_1, ..., B_n \in \mathcal{D}$ for some $\mathcal{D} \in \mathbf{C}$. Therefore, since $\mathcal{D}$ has the finite intersection property, $\bigcap_{i=1}^{n} B_i \neq \emptyset$. Therefore, $\mathcal{C}$ has the finite intersection property. Hence, $\mathcal{C} \in \mathbf{P}$ is an upper bound for the chain $\mathbf{C}$.

By Zorn's Lemma, there exists $\mathcal{F}$ a maximal element of $\mathbf{P}$. We claim that $\mathcal{F}$ is an ultrafilter extending $\mathcal{A}$. First, since $\mathcal{F} \in \mathbf{P}$, $\mathcal{A} \subseteq \mathcal{F}$ and $\emptyset \notin \mathcal{F}$ (1). For (2), suppose $A \in \mathcal{F}$ and $B \supseteq A$. Then, $\mathcal{F} \cup \{B\}$ certainly extends $\mathcal{A}$ and has the finite intersection property. By maximality, $B \in \mathcal{F}$. For (3), consider $A, B \in \mathcal{F}$. Again consider $\mathcal{F} \cup \{A \cap B\}$ and, as before, $(A \cap B) \in \mathcal{F}$. For (4), consider any $A \subseteq X$. If $\mathcal{F} \cup \{A\}$ has the finite intersection property, then $A \in \mathcal{F}$, so suppose not. Hence, there exists $B \in \mathcal{F}$ so that $A \cap B = \emptyset$ (recall by (3) $\mathcal{F}$ is closed under intersection, so "finite intersection property" reduces to "two intersection property"). That is, $B \subseteq (X \setminus A)$. For any other $C \in \mathcal{F}$, if $C \cap (X \setminus A) = \emptyset$, then $C \subseteq A$, hence $B \cap C = \emptyset$, contrary to $\mathcal{F} \in \mathbf{P}$. Hence, $\mathcal{F} \cup \{X \setminus A\}$ has the finite intersection property, so $(X \setminus A) \in \mathcal{F}$. This concludes the proof. □

Corollary 1.4.3. *If X is infinite, then X has a non-principal ultrafilter $\mathcal{F}$.*

Proof. Fix $\mathcal{A} = \mathrm{Cof}(X)$ and, by Proposition 1.4.2, there is $\mathcal{F} \supseteq \mathcal{A}$ an ultrafilter on X. □

5. Types

On the model theory side, we will work in a fixed lauguage L. We will write x, y, z for tuples of variables (as opposed to the cumbersome $\overline{x}$ or $\vec{x}$) and refer to these as simply "variables." Suppose that M is a fixed L-structure, which will have universe M (by abuse of notation). Recall that an L-structure is a set, M, together with an interpretation for each symbol in the language L. For example, if $L = \{+\}$ is the language of pure abelian groups, then $M = (\mathbb{Z}; +)$ is the L-structure of the abelian group $\mathbb{Z}$ (with usual addition). For a variable x, let $|x|$ denote the length of x. For $A \subseteq M$, let A_x denote all tuples of elements from A of length $|x|$, so $A_x = A^{|x|}$. If $|x| = 1$, we say that x is of the *home sort*.

Remark 1.5.1. Note that we can also consider multi-sorted languages L, in which case A_x will denote all elements of A of the same sort as x. However, some of the definitions and arguments in these lecture notes require a single "home sort," regardless.

If $\varphi(x; y)$ is an L-formula, M is an L-structure, $a \in M_x$, and $B \subseteq M_y$, let

$$\varphi(a; B) = \{b \in B : M \models \varphi(a; b)\}.$$

Moreover, for $A \subseteq M_x$, define

$$\varphi(A; B) = \{\varphi(a; B) : a \in A\}$$

Notice that this is a subset of $\mathcal{P}(B)$. We can identify this with the corresponding subset of ${}^{B}2$.

Fix a formula $\theta(x)$. We set

$$\theta(x)^1 = \theta(x) \text{ and } \theta(x)^0 = \neg\theta(x).$$

For a formula $\varphi(x;y)$ and $B \subseteq M_y$, a (*complete*) φ*-type* $p(x)$ over B is a maximally consistent subset of

$$\{\varphi(x;b)^t : b \in B, t = 0,1\}.$$

A *partial* φ*-type* is a subset of a φ-type. The set of all complete φ-types over B is denoted $S_\varphi(B)$. If B is smaller than the saturation of M (e.g., if B is finite), then there is a natural bijection between $S_\varphi(B)$ and $\varphi(M_x;B)$. If $a \in M_x$, define

$$\operatorname{tp}_\varphi(a/B) = \{\varphi(x;b)^t : b \in B, t < 2 \text{ so that } M \models \varphi(a;b)^t\}.$$

Clearly $\operatorname{tp}_\varphi(a/B) \in S_\varphi(B)$. In fact, we see that $\varphi(a;B)$ is exactly the positive part of the φ-type $\operatorname{tp}_\varphi(a/B)$, showcasing the bijection between $S_\varphi(B)$ and $\varphi(M_x;B)$.

If $C \subseteq M$ and x is a variable, a *(complete)* x*-type over* C is a maximally consistent subset of

$$\{\varphi(x;c) : y \text{ a variable}, \varphi(x;y) \text{ a formula}, c \in C_y\}.$$

The set of all such types is denoted $S_x(C)$. If $a \in M_x$, define

$$\operatorname{tp}(a/C) = \{\varphi(x;c) : \varphi(x;c) \text{ a formula over } C, M \models \varphi(a;c)\}.$$

This is the type of a over C. For each formula $\varphi(x;c)$, consider the subset of $S_x(C)$

$$[\varphi(x;c)] = \{p \in S_x(C) : \varphi(x;c) \in p(x)\}.$$

This forms a clopen basis for a topology on $S_x(C)$, and we call this the stone space. This space is compact, hausdorff, and totally disconnected.

Therefore, any closed subset $X \subseteq S_x(C)$ is compact. If we consider the open cover

$$\{[\varphi(x;c)] \cap X : \varphi(x;c) \in p(x) \text{ for some } p(x) \in X\}$$

of X, then there is a finite subcover. Therefore, we get the following proposition:

PROPOSITION 1.5.2. *Fix* $C \subseteq M$, x *a variable, and* $X \subseteq S_x(C)$ *a closed subset. Then, there exists* $\varphi_1(x;c_1), ..., \varphi_N(x;c_N)$ *so that:*

(1) *For each* $i = 1, ..., N$, *there exists* $p(x) \in X$ *so that* $\varphi_i(x;c_i) \in p(x)$; *and*

(2) *For each* $p(x) \in X$*, there exists* $i = 1, ..., N$ *so that* $\varphi_i(x; c_i) \in p(x)$.

If $p(x) \in S_x(C)$ and $\{p(x)\}$ is open, then there exists a single formula $\varphi(x; c) \in p(x)$ such that, for all other $\psi(x; d) \in p(x)$,

$$M \models \forall x(\varphi(x; c) \to \psi(x; d)).$$

In this case, we say that p is *isolated by* $\varphi(x; c)$ (this is the same notion of isolated as in the topological sense).

A *partial type* $p(x)$ is any consistent collection of formulas. If $\varphi(x; y)$ is a formula, $B \subseteq M_y$, and $p(x)$ is a partial type, then let

$$S_\varphi(B) \cap [p(x)] = \{q(x) \in S_\varphi(B) : p(x) \cup q(x) \text{ is consistent}\}. \tag{1.3}$$

In particular, if $p(x) = [x = x]$, then $S_\varphi(B) \cap [p(x)] = S_\varphi(B)$. We may also use this notation on complete type spaces. For $C \subseteq M$, let

$$S_x(C) \cap [p(x)] = \{q(x) \in S_x(C) : q(x) \cup p(x) \text{ is consistent}\}.$$

CHAPTER 2

Concept Classes

1. Introduction

Let X be a set and let $\mathcal{C} \subseteq {}^X 2$ (so $\mathcal{C}$ is a set of functions from X to $\{0,1\}$). We will call $\mathcal{C}$ a *concept class* on X.

This is used to model concepts in machine learning. For example, imagine that $X = \mathbb{R}^2$ and so an element of ${}^X 2$ can be thought of as a black-and-white image, and $\mathcal{C}$ a collection of images. In pattern recognition, we want to build an algorithm that determine which image from $\mathcal{C}$ the viewer is looking at. The usual method is through sampling points. That is, we look at some small subset $Y \subseteq X$ with a function $g \in {}^Y 2$ (a labeled sample) and try to recover a concept $f \in \mathcal{C}$ extending g. This is also related to compressibility (though it is different from how compression is understood in information theory). We imagine that X is the harddrive and $\mathcal{C}$ is a collection of possible states of the harddrive (assignments of 0 or 1 to the bits). Then, can we determine which state we are in by sampling (choosing an appropriate $Y \subseteq X$)?

This question depends on $\mathcal{C}$. If $\mathcal{C} = {}^X 2$, for example, then no information can be gained by sampling. For example, if we take $Y \subsetneq X$, $g \in {}^Y 2$, and ask which concepts extend g to a function on X, then all possible extensions lie in $\mathcal{C}$. So, for sampling to work, we need notions of measuring the complexity of a given concept class, $\mathcal{C}$.

1.1. VC-dimension and VC-density. If $\mathcal{C}$ is a concept class on X and $Y \subseteq X$, then define

$$\mathcal{C}|_Y = \{f|_Y : f \in \mathcal{C}\}.$$

We say that $\mathcal{C}$ *shatters* a set $Y \subseteq X$ if $\mathcal{C}|_Y = {}^Y 2$. That is, $\mathcal{C}$ shatters Y if all possible functions from Y to 2 can be extended to an element of $\mathcal{C}$.

Definition 2.1.1. The *VC-dimension* of a concept class $\mathcal{C}$ on X is the size of the largest finite subset of X that can be shattered by $\mathcal{C}$. If arbitrarily large finite subsets of X can be shattered by $\mathcal{C}$, then we say that $\mathcal{C}$ has infinite VC-dimension.

EXAMPLE 2.1.2. Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the set of all open intervals in $\mathbb{R}$ (technically, the characteristic functions of sets of the form $\{x : a < x < b\}$). We claim that this has VC-dimension 2. It is clear that any two-element subset can be shattered by $\mathcal{C}$. However, if we take $a < b < c$ from $\mathbb{R}$, then $\{a, b, c\}$ cannot be shattered by $\mathcal{C}$, since the subset $\{a, c\}$ cannot be realized (any interval containing a and c must contain b).

EXERCISE 2.1.3. Let $X = \mathbb{R}^2$ and let $\mathcal{C}$ be the set of all open axes-parallel rectangles in $\mathbb{R}^2$ (i.e. of the form $(a, b) \times (c, d)$). Compute the VC-dimension of $\mathcal{C}$ (Answer: 4; Hint: Smallest rectangle containing a finite set of points). Generalize to $\mathbb{R}^n$.

EXERCISE 2.1.4. Let $X = \mathbb{R}^2$ and let $\mathcal{C}$ be the set of all convex subsets in $\mathbb{R}^2$. Show that $\mathcal{C}$ has VC-dimension ∞. (Hint: Consider n points along a circle.)

EXERCISE 2.1.5. Suppose X is a set and $\mathcal{C}_1$ and $\mathcal{C}_2$ are concept classes. Suppose $\mathcal{C}_t$ has VC-dimension $n_t < \omega$ for both $t = 1, 2$. Compute a bound for the VC-dimension of the following concept classes:

(1) $\mathcal{C}_\cap = \{f_1 \cdot f_2 : f_1 \in \mathcal{C}_1 \text{ and } f_2 \in \mathcal{C}_2\}$.
(2) $\mathcal{C}_{1,\neg} = \{(1 - f_1) : f_1 \in \mathcal{C}_1\}$.
(3) $\mathcal{C}_\cup = \{(1 - (f_1 \cdot f_2)) : f_1 \in \mathcal{C}_{1,\neg} \text{ and } f_2 \in \mathcal{C}_{2,\neg}\}$.

(Hint: For (1), using the notation below, $\pi_{\mathcal{C}}(m) \leq \pi_{\mathcal{C}_1}(m) \cdot \pi_{\mathcal{C}_2}(m)$.)

So VC-dimension gives a measure of the complexity of a concept class $\mathcal{C}$. To get a measure of complexity for each size of a subset of X, we define the *shatter function* of a concept class $\mathcal{C}$ on X as follows. First, note that $\binom{X}{m}$ is the set of all subsets of X of size m.

$$\pi_{\mathcal{C}}(m) = \max\left\{ \left|\mathcal{C}|_Y\right| : Y \in \binom{X}{m} \right\}.$$

Of course $\pi_{\mathcal{C}}(m) \leq 2^m$. Moreover, if $\mathcal{C}$ has VC-dimension n, then $\pi_{\mathcal{C}}(m) = 2^m$ for $m \leq n$ and $\pi_{\mathcal{C}}(m) < 2^m$ for $m > n$. Can we get a tighter bound on the shatter function for $m > n$? Recall that

$$\Phi_n(m) = \sum_{i=0}^{n} \binom{m}{i}.$$

It is not hard to show that $\Phi_n(m) = \Phi_{n-1}(m-1) + \Phi_n(m-1)$.

THEOREM 2.1.6 (Sauer's Lemma). *If $\mathcal{C}$ has VC-dimension n and $m > n$, then $\pi_{\mathcal{C}}(m) \leq \Phi_n(m)$.*

Proof. Notice that VC-dimension does not increase by taking subsystems, so it suffices to show that $|\mathcal{C}| \leq \Phi_n(m)$ for any concept class $\mathcal{C}$ of VC-dimension n on X with $|X| = m \geq n$. We proceed by induction on m and n. If $m = n$, this is immediate (as $\Phi_n(n) = 2^n$).

Take $x \in X$ arbitrarily, let $X_0 = X \setminus \{x\}$, and define two concept classes on X_0, $\mathcal{C}_1$ and $\mathcal{C}_2$. Let $\mathcal{C}_2$ be the collection of all $f \in {}^{X_0}2$ so that both extensions of f to a function in ${}^{X}2$ are contained in $\mathcal{C}$. Similarly, let $\mathcal{C}_1$ be the collection of all $f \in {}^{X_0}2$ so that exactly one extension of f to a function in ${}^{X}2$ is contained in $\mathcal{C}$. By simple counting,

$$|\mathcal{C}| = |\mathcal{C}_1| + 2|\mathcal{C}_2|.$$

Notice also that $\mathcal{C}_2$ has VC-dimension $\leq n-1$. This is because, if $Y \subseteq X$ with $|Y| = n$ but $\mathcal{C}_2$ shatters Y, then $\mathcal{C}$ shatters $Y \cup \{x\}$, showing that $\mathcal{C}$ has VC-dimension $\geq n+1$. Therefore, by induction,

$$|\mathcal{C}_2| \leq \Phi_{n-1}(m-1).$$

On the other hand, $\mathcal{C}_1 \cup \mathcal{C}_2$ is a concept class in X_0 with VC-dimension n. By induction,

$$|\mathcal{C}_1| + |\mathcal{C}_2| \leq \Phi_n(m-1).$$

Therefore,

$$|\mathcal{C}| \leq \Phi_{n-1}(m-1) + \Phi_n(m-1) = \Phi_n(m).$$

□

Therefore, knowing the VC-dimension of a concept class gives information about the entire shatter function. If $\mathcal{C}$ has VC-dimension n,

$$\pi_{\mathcal{C}}(m) \begin{cases} = 2^m & \text{if } m \leq n. \\ \leq \Phi_n(m) & \text{if } m > n. \end{cases}$$

If $\mathcal{C}$ has VC-dimension n and $|\mathcal{C}|_Y| = \Phi_n(|Y|)$ for all $Y \subseteq X$ with $n \leq |Y| < \omega$, then we say that $\mathcal{C}$ is *maximum* of dimension n. In Example 2.1.2, $\mathcal{C}$ is maximum of dimension 2.

Exercise 2.1.7. Fix $n \geq 1$ and let $X = \omega$. Take $\mathcal{C}$ to be all functions $f : \omega \to 2$ so that $\mathrm{supp}(f) = \{n \in \omega : f(n) = 1\}$ has size at most n. Show that $\mathcal{C}$ is maximum of dimension n.

Exercise 2.1.8. Using the proof of Theorem 2.1.6 above as an outline, prove that: For X a finite set and $\mathcal{C}$ a concept class on X of VC-dimension n, $\mathcal{C}$ is maximum of dimension n if and only if $|\mathcal{C}| = \Phi_n(|X|)$.

Along these lines, we define the VC-density of a concept class $\mathcal{C}$.

Definition 2.1.9. Let $\mathcal{C}$ be a concept class on X. The *VC-density* of $\mathcal{C}$ is the infimum of all $\ell \in \mathbb{R}$ so that there exists K such that, for all m, $\pi_{\mathcal{C}}(m) \leq Km^{\ell}$. That is, VC-density is infimum of the power of a polynomial bound on the shatter function.

Theorem 2.1.6 tells us that if $\mathcal{C}$ has VC-dimension n, then it has VC-density $\leq n$. Moreover, it is easy to see that if the VC-density of $\mathcal{C}$ is finite, then so is the VC-dimension. This leads to another definition of a special kind of concept class.

Definition 2.1.10. We say that $\mathcal{C}$ is a *VC-class* if it has finite VC-dimension (equivalently, it has finite VC-density).

Given any concept class $\mathcal{C}$ on X, we can always consider the dual concept class $\mathcal{C}^*$ on $\mathcal{C}$ defined as follows:

$$\mathcal{C}^* = \{\sigma \in {}^{\mathcal{C}}2 : (\exists x)(\forall f \in \mathcal{C})(\sigma(f) = f(x))\}.$$

We say that $\mathcal{C}$ has *VC-codimension* n if $\mathcal{C}^*$ has VC-dimension n and we define VC-codensity similarly. How do these relate?

Theorem 2.1.11 (Proposition 1.3 of [**16**]). *A concept class $\mathcal{C}$ is a VC-class if and only if its dual $\mathcal{C}^*$ is a VC-class.*

Proof. To prove this, we establish that, if the VC-dimension of $\mathcal{C}$ is $n-1$, then the VC-dimension of $\mathcal{C}^*$ is bounded by $2^n - 1$.

Suppose the VC-dimension of $\mathcal{C}^*$ is $\geq 2^n$. Therefore, there exists $\mathcal{C}_0 \in \binom{\mathcal{C}}{2^n}$ that is shattered by $\mathcal{C}^*$. That is, for each $\mathcal{C}_1 \subseteq \mathcal{C}_0$, there exists $a_{\mathcal{C}_1} \in X$ so that

$$(\forall f \in \mathcal{C}_0)(f(a_{\mathcal{C}_1}) = 1 \text{ iff. } f \in \mathcal{C}_1).$$

Since $|\mathcal{C}_0| = 2^n$, index $\mathcal{C}_0$ as elements of $\mathcal{P}(n)$:

$$\mathcal{C}_0 = \{f_I : I \in \mathcal{P}(n)\}.$$

For each $i < n$, let $S_i = \{f_I : I \in \mathcal{P}(n), i \in I\}$. Therefore,

$$(\forall I \in \mathcal{P}(n))(f_I(a_{S_i}) = 1 \text{ iff. } i \in I).$$

Therefore, we see that $\{a_{S_i} : i < n\}$ is shattered by $\mathcal{C}_0$, hence by $\mathcal{C}$. So the VC-dimension of $\mathcal{C}$ is $\geq n$. □

Exercise 2.1.12. Let $X = \mathbb{R}^2$ and let $\mathcal{C}$ be the concept class of axes-parallel rectagles (as in Exercise 2.1.3). Show that the VC-codimension of $\mathcal{C}$ is 4. Show the VC-codensity is 2. This relates to Venn-Diagrams. For example, this says we cannot draw a Venn-Diagram using axes-parallel rectangles with 5 sets.

EXERCISE 2.1.13. Show the bound in the proof of Theorem 2.1.11 is tight. That is, exhibit a concept class $\mathcal{C}$ so that the VC-dimension is $n-1$ and the VC-codimension is $2^n - 1$.

The dual shatter function, denoted $\pi^*_{\mathcal{C}}(m)$, is the shatter function of the dual, $\mathcal{C}^*$. That is,

$$\pi^*_{\mathcal{C}}(m) = \pi_{\mathcal{C}^*}(m).$$

In particular, $\pi^*_{\mathcal{C}}(m) \geq \ell$ if there exists $\mathcal{C}_0 \in \binom{\mathcal{C}}{m}$ such that

$$|\{\mathcal{C}_1 \subseteq \mathcal{C}_0 : (\exists a \in X)(\forall f \in \mathcal{C}_1)(f(a) = 1)\}| \geq \ell.$$

Therefore, the VC-codimension of $\mathcal{C}$ is $\geq d$ if $\pi^*_{\mathcal{C}}(d) = 2^d$.

2. The Fractional Helly Property

2.1. Transversals and Packing. Given a concept class $\mathcal{C}$ on X, a subset $T \subseteq X$ is called a *transversal* of $\mathcal{C}$ if, for all $f \in \mathcal{C}$, there exists $x \in T$ so that $f(x) = 1$. That is, T "touches" each concept in $\mathcal{C}$. The *transveral number* of $\mathcal{C}$, denoted $\tau(\mathcal{C})$, is the smallest cardinality of a transversal of $\mathcal{C}$.

Dually, we define a *packing* of $\mathcal{C}$ to be some $\mathcal{C}_0 \subseteq \mathcal{C}$ so that, for each $f_1, f_2 \in \mathcal{C}_0$, $\operatorname{supp}(f_1) \cap \operatorname{supp}(f_2) = \emptyset$. The *packing number* of $\mathcal{C}$, denoted $\nu(\mathcal{C})$, is the maximum cardinality of a packing of $\mathcal{C}$.

LEMMA 2.2.1. *For any concept class $\mathcal{C}$,*

$$\nu(\mathcal{C}) \leq \tau(\mathcal{C}).$$

PROOF. Fix $T \subseteq X$ a transversal for $\mathcal{C}$ and let $\mathcal{C}_0 \subseteq \mathcal{C}$ be a packing for $\mathcal{C}$. Then, since each $x \in T$ is contained in at most one element of $\mathcal{C}_0$, $|\mathcal{C}_0| \leq |T|$. □

Now we work over a finite set X. A *fractional transversal* for a concept class $\mathcal{C}$ on X is a map $\sigma : X \to [0, 1]$ such that, for each $f \in \mathcal{C}$,

$$\sum\{\sigma(x) : x \in X, f(x) = 1\} \geq 1.$$

The *fractional transversal number*, denoted $\tau^*(\mathcal{C})$, is the infimum of $\sum\{\sigma(x) : x \in X\}$ over all fractional transversals σ. So the idea is that we can put parts of a point into the support of each $f \in \mathcal{C}$, but to total amount of points must be at least one.

We consider the dual notion. A *fractional packing* for a finite concept class $\mathcal{C}$ is a function $\sigma : \mathcal{C} \to [0, 1]$ such that, for each $x \in X$,

$$\sum\{\sigma(f) : f \in \mathcal{C}, f(x) = 1\} \leq 1.$$

The *fractional packing number* of $\mathcal{C}$, denoted $\nu^*(\mathcal{C})$, is the supremum of $\sum\{\sigma(f) : f \in \mathcal{C}\}$ over all fractional packings σ. Again, the idea is that we include parts of functions $f \in \mathcal{C}$ so that the intersection of the supports is no more than one.

REMARK 2.2.2. For any concept class $\mathcal{C}$ over a finite set X, it is easy to check that

(1) $\nu(\mathcal{C}) \leq \nu^*(\mathcal{C})$ and
(2) $\tau^*(\mathcal{C}) \leq \tau(\mathcal{C})$.

The remarkable fact is that $\nu^* = \tau^*$.

THEOREM 2.2.3. *For every concept class $\mathcal{C}$ on a finite set X, we have*

$$\nu^*(\mathcal{C}) = \tau^*(\mathcal{C}).$$

Moreover, there is an optimal fractional packing and this takes on rational values (and similarly for fractional transversals).

This theorem follows from the duality of linear programming.

PROPOSITION 2.2.4. *Let A be an $m \times n$ real matrix, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. Let*

$$P = \{x \in \mathbb{R}^n : x \geq 0, Ax \geq b\}$$

and

$$D = \{y \in \mathbb{R}^m : y \geq 0, y^T A \leq c^T\}.$$

If P and D are non-empty, then

$$\min\{c^T x : x \in P\} = \max\{y^T b : y \in D\}.$$

PROOF OF THEOREM 2.2.3. Let A be the $m \times n$ matrix

$$A = (f(a))_{a \in X, f \in \mathcal{C}}$$

Then we have that

$$\tau^*(\mathcal{C}) = \min\{\mathbf{1}_X^T x : x \geq 0, Ax \geq \mathbf{1}_\mathcal{C}\}$$

and

$$\nu^*(\mathcal{C}) = \max\{y^T \mathbf{1}_\mathcal{C} : y \geq 0, y^T A \leq \mathbf{1}_X^T\},$$

where $\mathbf{1}_X$ is the column vector of all 1's indexed by X. Therefore, by Proposition 2.2.4, $\tau^*(\mathcal{C}) = \nu^*(\mathcal{C})$.

Since the linear program is integer-valued, any optimal fractional packing takes on rational values. ◻

2.2. Epsilon Nets. For this subsection, let $(X, \mathcal{B}, \mu)$ be a probability space. Let $\mathcal{C}$ be a concept class on X so that each $f \in \mathcal{C}$ has μ-measurable support (i.e., for all $f \in \mathcal{C}$, $\mathrm{supp}(f) \in \mathcal{B}$). For each $f \in \mathcal{C}$, let $\mu(f) = \mu(\mathrm{supp}(f))$ (so $\mu(f) = \int_X f d\mu$). That is, $\mu(f)$ is the μ-probability that, given $a \in X$, $f(a) = 1$. For any $n < \omega$, we consider the product measure of μ on X^n, which, by abuse of notation, we will denote by μ.

LEMMA 2.2.5. *Fix $p \in [0,1]$ and $n < \omega$ with $n \geq 8/p$. Let $f_0, ..., f_{n-1} \in {}^X2$ such that $\mu(f_i) = p$ for all $i < n$. Then,*

$$\mu\left(\left\{\langle a_1, ..., a_n\rangle \in X^n : \sum_{i=1}^{n} f_i(a_i) \leq \frac{1}{2}np\right\}\right) \leq \frac{1}{2}.$$

PROOF. Define $f : X^n \to \mathbb{R}$ as

$$f(a_1, ..., a_n) = \sum_{i=1}^{n} f_i(a_i).$$

The expected value of f is $E(f) = np$ and the variance of f is $\mathrm{Var}(f) = np(1-p)$. By Proposition 1.3.3 (Chebyshev's Inequality), setting $\epsilon = np/2$,

$$\mu\big(\{\overline{a} \in X^n : |f(\overline{a}) - np| \geq np/2\}\big) \leq \frac{np(1-p)}{(np/2)^2} \leq \frac{4}{pn}.$$

However, by assumption, $4/pn \leq 1/2$. The conclusion follows. ☐

DEFINITION 2.2.6. Fix $\epsilon \in [0,1]$. A subset $N \subseteq X$ is called an *ϵ-net* for $\mathcal{C}$ if, for all $f \in \mathcal{C}$ with $\mu(f) \geq \epsilon$, there exists $a \in N$ such that $f(a) = 1$.

In other words, N is a transversal for the concept class

$$\mathcal{C}_\epsilon = \{f \in \mathcal{C} : \mu(f) \geq \epsilon\}.$$

We now show the main result of this subsection:

THEOREM 2.2.7 (VC-Theorem). *Fix $(X, \mathcal{B}, \mu)$ a probability space, $\mathcal{C}$ a concept class on X with each concept having μ-measurable support, $d, n < \omega$, and $\epsilon > 0$. If the VC-dimension of $\mathcal{C}$ is $\leq d$, then*

$$\mu\big(\{\overline{a} \in X^n : \{a_1, ..., a_n\} \textit{ is not an } \epsilon\textit{-net for } \mathcal{C}\}\big) \leq 2(2n)^d 2^{-\epsilon n/2}.$$

PROOF. Without loss of generality, suppose that $\mathcal{C} = \mathcal{C}_\epsilon$, so each $f \in \mathcal{C}$ is such that $\mu(f) \geq \epsilon$. Define

$$Y_0 = \{\overline{a} \in X^n : (\exists f \in \mathcal{C})(\forall i < n)(f(a_i) = 0)\}.$$

That is, Y_0 is the set we are trying to measure. In other words, $\mu(Y_0)$ is the probability that a randomly chosen n-element sequence of X is not a transversal of $\mathcal{C}$. For each $f \in \mathcal{C}$, define

$$Y_f = \left\{ \overline{a} \in X^{2n} : (\forall i < n)(f(a_i) = 0) \wedge \sum_{i=n}^{2n-1} f(a_i) \geq \frac{\epsilon n}{2} \right\}.$$

and let $Y_1 = \bigcup_{f \in \mathcal{C}} Y_f$. Fix $\overline{a} \in Y_0$ and some $f \in \mathcal{C}$ witnessing this (so $f(a_i) = 0$ for all $i < n$). Consider

$$Y_1|_{\overline{a}} = \{\overline{b} \in X^n : \overline{a}^\frown \overline{b} \in Y_1\}.$$

By Lemma 2.2.5,

$$\mu\left(Y_1|_{\overline{a}}\right) \geq 1/2.$$

Moreover, if $\overline{a} \in (X^n \setminus Y_0)$, then $\mu(Y_1|_{\overline{a}}) = 0$, as the first condition fails. Therefore,

$$\mu(Y_0) \leq 2\mu(Y_1).$$

So it suffices to compute a bound for $\mu(Y_1)$.

We put the uniform probability measure on $\binom{2n}{n}$ and consider the product space $X^{2n} \times \binom{2n}{n}$. By abuse of notation, we will let μ denote the measure on this space. For each $I \in \binom{2n}{n}$, let σ_I be a permutation of $2n$ that sends I to n (and sends $(2n \setminus I)$ to $(2n \setminus n)$). For each $\overline{a} \in X^{2n}$, define

$$\overline{a}_I = \langle a_{\sigma_I(0)}, ..., a_{\sigma_I(2n-1)} \rangle.$$

Fix $\overline{a} \in X^{2n}$ and $f \in \mathcal{C}$. We compute the probability that $\overline{a}_I \in Y_f$ for some $I \in \binom{2n}{n}$. If $\sum_{i<2n} f(a_i) < \epsilon n/2$, then $\overline{a}_I \notin Y_f$ for all I (since any permutation will still fail the second condition of being in Y_f). Otherwise, $\overline{a}_I \in Y_f$ if and only if $(\forall i \in I)(f(a_i) = 0)$. Let $k = \lceil \frac{\epsilon n}{2} \rceil$, so I has to avoid the subset of $2n$ where $f(a_i) = 1$, which is of size at least k. Thus, the probability that $\overline{a}_I \in Y_f$, i.e.,

$$\mu\left(\left\{ I \in \binom{2n}{n} : \overline{a}_I \in Y_f \right\}\right)$$

is bounded by

$$\frac{\binom{2n-k}{n}}{\binom{2n}{n}} \leq \frac{n(n-1)...(n-k+1)}{2n(2n-1)...(2n-k+1)} \leq \left(\frac{1}{2}\right)^k \leq 2^{-\epsilon n/2}.$$

Recall that $\mathcal{C}$ has VC-dimension $\leq d$. By Theorem 2.1.6, $\pi_{\mathcal{C}}(2n) \leq \Phi_d(2n)$. Thereby, for fixed $\overline{a} \in X^{2n}$,

$$\left|\{I \subseteq (2n) : (\exists f \in \mathcal{C})(f(a_i) = 1 \text{ iff. } i \in I)\}\right| \leq \Phi_d(2n).$$

So the probability that $\overline{a}_I \in Y_1$ (i.e., there exists some $f \in \mathcal{C}$ so that $\overline{a}_I \in Y_f$) is

$$\mu\left(\left\{I \in \binom{2n}{n} : \overline{a}_I \in Y_1\right\}\right) \leq \Phi_d(2n)2^{-\epsilon n/2}.$$

However, by Lemma 1.2.3, $\Phi_d(2n) \leq (2n)^d$, so this is bounded by $(2n)^d 2^{-\epsilon n/2}$. That is, for any choice of $\overline{a} \in X^{2n}$,

$$\mu\left(\left\{I \in \binom{2n}{n} : \overline{a}_I \in Y_1\right\}\right) < (2n)^d 2^{-\epsilon n/2}.$$

Therefore, $\mu(Y_1) < (2n)^d 2^{-\epsilon n/2}$. Hence,

$$\mu(Y_0) < 2(2n)^d 2^{-\epsilon n/2}.$$

□

Choose a function $g : \omega \times \mathbb{R}_+ \to \mathbb{R}$ such that, for $d < \omega$ and $\delta > 0$, for all $n \geq g(d, \delta)$,

$$2n^d < 2^{\delta n}.$$

Since $x \mapsto 2x^d$ is a polynomial function and $x \mapsto 2^{\delta x}$ is exponential function, such a function g exists.

Exercise 2.2.8. Show that

$$g(d, \delta) = \max\left\{\delta, 2\delta d \log_2(2\delta d)\right\}$$

works, for example.

Corollary 2.2.9 (Bounded ϵ-nets). *Fix $(X, \mathcal{B}, \mu)$ a probability space, $\mathcal{C}$ a concept class on X with each concept having μ-measurable support, $d < \omega$, and $\epsilon > 0$. If the VC-dimension of $\mathcal{C}$ is $\leq d$, then there exists $N \subseteq X$ a ϵ-net for $\mathcal{C}$ with*

$$|N| \leq \frac{1}{2} g\left(d, \frac{\epsilon}{4}\right).$$

Proof. Choose $n \geq \frac{1}{2} g(d, \epsilon/4)$. By our choice of n (and function g),

$$2(2n)^d 2^{-\epsilon n/2} < 1.$$

By Theorem 2.2.7,

$$\mu\big(\{\overline{a} \in X^n : \{a_1, ..., a_n\} \text{ is not an } \epsilon\text{-net for } \mathcal{C}\}\big) < 1.$$

Hence, there is an ϵ-net of size n. □

The existence of an ϵ-net has an interesting corollary.

COROLLARY 2.2.10. *Fix $d < \omega$. If $\mathcal{C}$ is a finite concept class on a set X with VC-dimension $\leq d$, then*

$$\tau(\mathcal{C}) \leq \frac{1}{2} g\left(d, \frac{1}{4\nu^*(\mathcal{C})}\right).$$

PROOF. By Theorem 2.2.3, $\nu^*(\mathcal{C}) = \tau^*(\mathcal{C})$. Let $\epsilon = 1/\tau^*(\mathcal{C})$. Since $\mathcal{C}$ is finite, we may assume that X is finite. Let $\sigma : X \to [0,1]$ be a optimal fractional transversal and define a probability measure μ on X such that, for all $a \in X$,

$$\mu(\{a\}) = \epsilon\sigma(a).$$

By definition, for all $f \in \mathcal{C}$,

$$\mu(f) = \epsilon \sum \{\sigma(a) : f(a) = 1\} \geq \epsilon.$$

Therefore, $\mathcal{C} = \mathcal{C}_\epsilon$. By Corollary 2.2.9, there exists an ϵ-net of size at most $\frac{1}{2} g(d, \epsilon/4)$. This is a transversal. □

2.3. Helly's Property and Fractional Helly's Property. Let $\mathcal{C}$ be a concept class on X. Recall that $\binom{X}{k}$ denotes the set of all subsets of X of size k.

DEFINITION 2.2.11. We say that $\mathcal{C}$ has *Helly number k* if, for all $n < \omega$ and all $f_0, ..., f_{n-1} \in \mathcal{C}$, if, for all $I \in \binom{n}{k}$, there exists $a \in X$ so that $(\forall i \in I)(f_i(a) = 1)$, then

$$(\exists a \in X)(\forall i < n)(f_i(a) = 1).$$

EXAMPLE 2.2.12 (Helly's Theorem). If $\mathcal{C}$ the collection of convex subsets of $\mathbb{R}^d$, then $\mathcal{C}$ has the Helly number $d + 1$. For example, if $d = 1$, this says that any finite collection of intervals in $\mathbb{R}$ where every two of which intersect must have a common intersection point.

Sometimes this is not enough, so we will have to consider the fractional version.

DEFINITION 2.2.13. We say that $\mathcal{C}$ has *fractional Helly number k* if, for every $\alpha > 0$, there exists $\beta > 0$ such that, if $n < \omega$ and $f_0, ..., f_{n-1} \in \mathcal{C}$ are such that, for an α proportion of the $I \in \binom{n}{k}$, we have that $\bigcap\{\mathrm{supp}(f_i) : i \in I\}$ is non-empty, then there exists $I^* \subseteq n$ with $|I^*| \geq \beta n$ such that $\bigcap\{\mathrm{supp}(f_i) : i \in I^*\}$ is non-empty.

In other words, for all $n < \omega$ and $f_0, ..., f_{n-1} \in \mathcal{C}$,

$$\left|\left\{I \in \binom{n}{k} : (\exists a \in X)(\forall i \in I)(f_i(a) = 1)\right\}\right| \geq \alpha \binom{n}{k}$$

implies that there exists some $a \in X$ with

$$|\{i < n : f_i(a) = 1\}| \geq \beta n.$$

EXERCISE 2.2.14. If $\mathcal{C}_1$ and $\mathcal{C}_2$ have fractional Helly number k, show that

$$\mathcal{C} = \{f : (\exists f_1 \in \mathcal{C}_1)(\exists f_2 \in \mathcal{C}_2)(f(a) = 1 \text{ iff. } f_1(a) = 1 \text{ or } f_2(a) = 1)\}$$

has fractional Helly number k.

EXERCISE 2.2.15. Let $\mathcal{C}$ be the concept class of convex subsets of $\mathbb{R}^d$ intersected with $\mathbb{Z}^d$, so

$$\mathcal{C} = \{\chi_{(X \cap \mathbb{Z}^d)} : X \subseteq \mathbb{R}^d \text{ convex}\}.$$

Show that the Helly number of $\mathcal{C}$ is not $2^d - 1$. (Hint: Consider $\{0,1\}^d \subseteq \mathbb{Z}^d$.) One can show that $\mathcal{C}$ has Helly number 2^d. On the other hand, $\mathcal{C}$ has fractional Helly number $d+1$, as shown in Theorem 1.1 of [**3**]. This shows that fractional Helly number and Helly number need not coincide.

THEOREM 2.2.16 (Theorem 2 of [**18**]). *If $\mathcal{C}$ is a concept class on X with VC-codensity $< k$, then $\mathcal{C}$ has fractional Helly number k. In particular, any VC-class has a fractional Helly number.*

PROOF. First recall that

$$\lim_{m \to \infty} \left(1 - \frac{1}{m}\right)^m = \frac{1}{e} > \frac{1}{3}.$$

Therefore, for sufficiently large m, $(1 - 1/m)^m \geq 1/3$.

Also recall that $\pi^*_{\mathcal{C}}(m)$ is the dual shatter function. That is, $\pi^*_{\mathcal{C}}(m) \geq \ell$ if and only if there exists $\mathcal{C}_1 \in \binom{\mathcal{C}}{m}$ such that

$$\left|\left\{\mathcal{C}_2 \subseteq \mathcal{C}_1 : (\exists a \in X)(\forall f \in \mathcal{C}_1)(f(a) = 1 \text{ iff. } f \in \mathcal{C}_2)\right\}\right| \geq \ell.$$

Since $\mathcal{C}$ has VC-codensity $< k$, for any constant K, for sufficiently large m, $\pi^*_{\mathcal{C}}(m) < K\binom{m}{k}$.

Suppose $\alpha > 0$, $n < \omega$, and $f_0, ..., f_{n-1} \in \mathcal{C}$ so that at least $\alpha\binom{n}{k}$ of the elements $I \in \binom{n}{k}$ are such that there exists $a \in X$ so that, for all $i \in I$, $f_i(a) = 1$. Choose m large enough so that

$$(1 - 1/m)^m \geq 1/3 \tag{2.1}$$

$$\pi^*_{\mathcal{C}}(m) < \frac{1}{3}\alpha\binom{m}{k}. \tag{2.2}$$

Choose

$$\beta = 1/2m. \tag{2.3}$$

Furthermore, we may assume n is arbitrarily large (otherwise, choose β sufficiently small). Therefore, suppose that n is so large that

$$n \geq 2m. \tag{2.4}$$

By means of contradiction, suppose that there exists no $a \in X$ so that

$$\big|\{i < n : f_i(a) = 1\}\big| \geq \beta n.$$

Let $J \in \binom{n}{m}$ and $I \in \binom{J}{k}$. We will say that (I, J) satisfies (†) if

$$(\dagger) : (\exists a \in X)(\forall i \in J)(f_i(a) = 1 \text{ iff. } i \in I).$$

We now compute a lower bound for the probability that a randomly chosen pair (I, J) satisfies (†) (when we use the uniform distribution).

First choose $I \in \binom{n}{k}$ randomly. The probability that there exists $a \in X$ so that $f_i(a) = 1$ for all $i \in I$ is at least α by assumption. Given such a choice of I, fix such an $a \in X$ (i.e., $f_i(a) = 1$ for all $i \in I$). Now choose $(J \setminus I)$ randomly of size $m - k$ from $(n \setminus I)$. By assumption, fewer than βn of the $i < n$ are such that $f_i(a) = 1$. Therefore, the probability that $f_j(a) = 0$ for all $j \in (J \setminus I)$ is at least

$$\frac{\binom{\lceil (1-\beta)n \rceil}{m-k}}{\binom{n-k}{m-k}} = \prod_{i=0}^{m-k-1} \frac{\lceil (1-\beta)n \rceil - i}{n-i} \geq \prod_{i=0}^{m-k-1} \frac{(1-\beta)n - i}{n-i}.$$

Now if $0 < c < a < b$, then $\frac{a}{b} \geq \frac{a-c}{b-c}$. Therefore, the probability is at least

$$\prod_{i=0}^{m-k-1} \frac{(1-\beta)n - m}{n-m} \geq \prod_{i=0}^{m-1} \frac{(1-\beta)n - m}{n-m} = \left(\frac{(1-\beta)n - m}{n-m}\right)^m$$

(the inequality holds because $\frac{(1-\beta)n-m}{n-m} < 1$). By (2.4),

$$\frac{(1-\beta)n - m}{n-m} = \frac{n - \beta n - m}{n-m} = 1 - \frac{n}{n-m}\beta \geq 1 - 2\beta.$$

Therefore, the probability is at least

$$(1 - 2\beta)^m \geq \left(1 - \frac{1}{m}\right)^m \geq \frac{1}{3}$$

(where the first inequality holds by (2.3) and the second by (2.1)). So the probability that a randomly chosen pair (I, J) satisfies (†) is at least $\frac{1}{3}\alpha$.

Choosing $J \in \binom{n}{m}$ at random, the expected number of $I \in \binom{J}{k}$ so that (I, J) satisfies (†) is at least $\frac{1}{3}\alpha\binom{m}{k}$. Therefore, there exists a

$J \in \binom{n}{m}$ such that

$$\mathcal{I} = \left\{ I \in \binom{J}{k} : (I, J) \text{ satisfies } (\dagger) \right\},$$

has cardinality at least $\frac{1}{3}\alpha\binom{m}{k}$. By definition, for each $I \in \mathcal{I}$, there exists $a_I \in X$ so that, for all $i \in J$, $f_i(a_I) = 1$ if and only if $i \in I$. In particular, if $I_1 \neq I_2$ are from $\mathcal{I}$ and $i \in (I_1 \setminus I_2)$, then $f_i(a_{I_1}) \neq f_i(a_{I_2})$. Therefore, $\pi^*_{\mathcal{C}}(m) \geq \frac{1}{3}\alpha\binom{m}{k}$. This contradicts (2.2). □

2.4. The (p,q)-Theorem. We now use Theorem 2.2.16 to prove a powerful result for VC-classes $\mathcal{C}$.

Definition 2.2.17. Fix $n < \omega$, $\overline{f} = \langle f_0, ..., f_{n-1} \rangle \in ({}^X 2)^n$, and $q \leq p < \omega$. We say that $\overline{f}$ has the *(p,q)-property* if, for all $J \in \binom{n}{p}$, there exists $a \in X$ such that

$$\left| \{ i \in J : f_i(a) = 1 \} \right| \geq q.$$

If $\mathcal{C}$ is a finite concept class, we say that $\mathcal{C}$ has the *(p,q)-property* if any enumeration of $\mathcal{C}$ has the (p,q)-property.

Theorem 2.2.18 (Theorem 4 of [**18**]). *For each $0 < k \leq p < \omega$ and $d < \omega$, there exists $K < \omega$ such that, if $\mathcal{C}$ is a concept class on X with VC-codensity $< k$ and VC-dimension $\leq d$, then, for all finite $\mathcal{C}_0 \subseteq \mathcal{C}$, if $\mathcal{C}_0$ has the (p, k)-property, then $\tau(\mathcal{C}_0) \leq K$.*

That is, if $\mathcal{C}_0$ has the (p,q)-property, then $\mathcal{C}_0$ has a transversal of bounded size. To prove this theorem, we use ϵ-nets together with a bound on $\nu^*(\mathcal{C}_0)$.

Lemma 2.2.19. *Fix $q \leq p < \omega$ and $\mathcal{C}$ a concept class with fractional Helly number q. Then, there exists $\beta > 0$ such that, for any $n < \omega$ and any $f_0, ..., f_{n-1} \in \mathcal{C}$, if $\overline{f}$ has the (p,q)-property, then there exists $a \in X$ so that*

$$|\{ i < n : f_i(a) = 1 \}| \geq \beta n.$$

Proof. Suppose $\mathcal{C}$ has fractional Helly number q. Let

$$\alpha = \frac{1}{\binom{p}{q}}$$

and let β be the corresponding β from the fractional Helly property.

Fix $n < \omega$ and $f_0, ..., f_{n-1} \in \mathcal{C}$. We first compute the number of q-element subsets of n that have non-empty common support. For each $J \in \binom{n}{p}$, there exists $I \in \binom{J}{q}$ and $a \in X$ so that $(\forall i \in I)(f_i(a) = 1)$.

On the other hand, for each $I \in \binom{n}{q}$, there exists $\binom{n-q}{p-q}$ many $J \in \binom{n}{p}$ so that $I \subseteq J$. Therefore, there are at least

$$\frac{\binom{n}{p}}{\binom{n-q}{p-q}} = \frac{(p-q)!n!}{p!(n-q)!} = \alpha \binom{n}{q}$$

many $I \in \binom{n}{q}$ so that $(\exists a \in X)(\forall i \in I)(f_i(a) = 1)$. Since $\mathcal{C}$ has fractional Helly number q, there exists $a \in X$ so that

$$|\{i < n : f_i(a) = 1\}| \geq \beta n.$$

□

LEMMA 2.2.20. *Fix $q \leq p < \omega$ and $\mathcal{C}$ a concept class with fractional Helly number q. Then there exists $K < \omega$ such that, for all finite $\mathcal{C}_0 \subseteq \mathcal{C}$ with the (p,q)-property,*

$$\nu^*(\mathcal{C}_0) \leq K.$$

PROOF. Let $K = 1/\beta$ for β from Lemma 2.2.19 with the (pq,q)-property.

Fix a finite $\mathcal{C}_0 \subseteq \mathcal{C}$ with the (p,q)-property. Let $\sigma : \mathcal{C}_0 \to [0,1]$ be an optimal fractional packing, so

$$\nu^*(\mathcal{C}_0) = \sum\{\sigma(f) : f \in \mathcal{C}_0\}.$$

By Theorem 2.2.3, σ takes on rational values. Let D be a positive integer and $m : \mathcal{C}_0 \to \omega$ such that $\sigma(f) = m(f)/D$. Let $n = \sum\{m(f) : f \in \mathcal{C}_0\}$. Finally, let $\langle f_i : i < n\rangle$ be an enumeration of $\mathcal{C}_0$ so that, for each $f \in \mathcal{C}_0$, $|\{i < n : f_i = f\}| = m(f)$. That is, we can think of $\overline{f}$ as a "multiset" with $m(f)$ copies of f for each $f \in \mathcal{C}_0$.

Now $\overline{f}$ has the (pq,q)-property. To see this, consider $J \in \binom{n}{pq}$. Either $|\{f_i : i \in J\}| \geq p$ or there exists f so that $|\{i \in J : f_i = f\}| \geq q$ (i.e., if each element $f \in \mathcal{C}_0$ is repeated fewer than q times, then there are $pq/q = p$ distinct f_i's). Therefore, by Lemma 2.2.18, there exists $a \in X$ so that

$$|\{i < n : f_i(a) = 1\}| \geq \beta n.$$

However, by the definition of fractional packing,

$$\sum\{\sigma(f) : f \in \mathcal{C}_0, f(a) = 1\} \leq 1.$$

Therefore,

$$1 \geq \sum\{m(f)/D : f \in \mathcal{C}_0, f(a) = 1\} = \frac{1}{D}|\{i < n : f_i(a) = 1\}| \geq \frac{\beta n}{D}.$$

However, $D\nu^*(\mathcal{C}_0) = n$, so $\beta n/D = \beta\nu^*(\mathcal{C}_0)$. Therefore,

$$\nu^*(\mathcal{C}_0) \leq 1/\beta = K.$$

□

Proof of Theorem 2.2.18. By Theorem 2.2.20, there exists a number K depending only on p and k such that, for all finite $\mathcal{C}_0 \subseteq \mathcal{C}$ with the (p,k)-property, $\nu^*(\mathcal{C}_0) \leq K$. By Theorem 2.2.10, there is a bound, depending only on d and $\nu^*(\mathcal{C}_0)$, on the size of transversals for any finite $\mathcal{C}_0 \subseteq \mathcal{C}$. Therefore, there is a global bound depending only on p, k, and d. □

3. Learnability and Compression

3.1. PAC-Learning. Fix a concept class $\mathcal{C}$ on X and define

$$\mathcal{C}_{\text{fin}} = \{f|_Y : Y \subseteq X, Y \text{ finite}\}.$$

This codes all the finite parts of elements of $\mathcal{C}$.

Suppose that $(X, \mathcal{B}, \mu)$ is a probability space and $\mathcal{C}$ is a concept class on X of μ-measurable functions. Consider a function $H : \mathcal{C}_{\text{fin}} \to {}^X 2$. This function takes a finite sample of a concept and returns a guess at a concept (we will call this the "hypothesis function"). We will say that the hypothesis function H is *consistent* if, for all $f \in \mathcal{C}$ and all $Y \subseteq X$ finite,

$$(\forall a \in Y) \left([H(f|_Y)](a) = f(a)\right).$$

That is, if the hypothesis matches the actual concept on the set Y. We can define the *error* of the hypothesis as follows: Given a fixed $f \in \mathcal{C}$, $n < \omega$, and $\overline{a} \in X^n$,

$$\text{err}_\mu(H, f, \overline{a}) := \mu\left(\left\{c \in X : f(c) \neq [H(f|_{\{a_1,\ldots,a_n\}})](c)\right\}\right).$$

That is, $\text{err}_\mu(H, f, \overline{a})$ is the μ-probability that f differs from the hypothesis of f on $\overline{a}$. Of course, we have to assume that all such sets are μ-measurable.

Now fix X a set and $\mathcal{C}$ a concept class on X. We say that $\mathcal{C}$ is *probably approximately correctly learnable* (or *PAC-learnable*) if there exists a hypothesis function $H : \mathcal{C}_{\text{fin}} \to {}^X 2$ such that, for all $\epsilon > 0$ and all $\delta > 0$, there exists $N_{\epsilon,\delta} < \omega$ such that, for all $n \geq N_{\epsilon,\delta}$, all $f \in \mathcal{C}$, and all probability measures μ on X (with the correct sets being μ-measurable),

$$\mu^n\left(\{\overline{a} \in X^n : \text{err}_\mu(H, f, \overline{a}) > \epsilon\}\right) < \delta.$$

That is, with low probability, the error in the hypothesis is high. This is where the "probably" and "approximately" come from. The function $\langle \epsilon, \delta \rangle \mapsto N_{\epsilon,\delta}$ (for minimal such choice of $N_{\epsilon,\delta}$) above is called the *sample complexity* of H.

Example 2.3.1 (Example 2.1 of [**4**]). Consider the example of learning the concept of "room temperature," say between 20 °C and 25 °C. To learn this concept, we randomly sample temperatures on Earth with an oracle that tells us which are "room temperature" and which aren't. Then, we make a hypothesis of what the definition of "room temperature" should be from this. In this example, our hypothesis will be the smallest closed interval of temperatures containing all of those sampled temperatures labeled as "room temperature." With high probability, our guess will classify most temperatures correctly. Notice that the probability measure on temperatures on Earth is not in any way uniform. This is why we need our hypothesis to work for *any* probability measure μ.

Formally, let $X = \mathbb{R}$ and let $\mathcal{C}$ be the concept class of closed intervals in $\mathbb{R}$. Given any closed interval $I \subseteq \mathbb{R}$ and any sample $Y \subseteq \mathbb{R}$ (finite), we set the hypothesis $h = [a, b]$ where a is the smallest element of $I \cap Y$ and b is the largest element of $I \cap Y$. If $I \cap Y = \emptyset$, set $h = \emptyset$. Formally, $I = \operatorname{supp}(f)$ and

$$[H(f|_Y)](x) = 1 \text{ iff. } (\exists a, b \in Y)(f(a) = f(b) = 1 \wedge a \leq x \leq b).$$

The claim is that H witnesses PAC-learnability with sample complexity

$$\langle \epsilon, \delta \rangle \mapsto N_{\epsilon,\delta} \leq \frac{2}{\epsilon} \log\left(\frac{2}{\delta}\right).$$

Suppose $f = \operatorname{supp}([a, b])$ and fix $\epsilon > 0$, $\delta > 0$, and μ a probability measure on X. If $\mu(f) < \epsilon$, then since $\operatorname{supp}(H(f|_Y)) \subseteq \operatorname{supp}(f)$, $\mu(H(f|_Y)) < \epsilon$. So we may assume that $\mu(f) \geq \epsilon$. As such, define the two side intervals as

$$\begin{aligned} L =& [a, x], \text{ where } x = \inf\{x \in [a, b] : \mu([a, x]) \geq \epsilon/2\}, \text{ and} \\ R =& [x, b], \text{ where } x = \sup\{x \in [a, b] : \mu([x, b]) \geq \epsilon/2\}. \end{aligned}$$

Then, $\mu(L) = \mu(R) = \epsilon/2$. Notice now that, for any n,

$$\mu^n\left(\{\overline{a} \in X^n : (\forall i < n)(a_i \notin L)\}\right) \leq \left(1 - \frac{\epsilon}{2}\right)^n.$$

Hence, the μ-probability that, given $\overline{a} \in X^n$, either each a_i misses L or each a_i misses R is

$$2\left(1 - \frac{\epsilon}{2}\right)^n \leq 2e^{-\epsilon n/2},$$

where the inequality holds by (1.2). Therefore, if $n \geq 2/\epsilon \log(2/\delta)$, then with μ-probabilitly at least $1 - \delta$, we have that each of L and R contains a sample. However, if each of L and R contains a sample, then the hypothesis will be an interval I such that $([a,b]\setminus I) \subseteq (L \cup R)$, and $\mu(L \cup R) = \epsilon$. Hence, the μ-measure of the error is $< \epsilon$. Therefore,

$$\mu^n\left(\{\overline{a} \in X^n : \operatorname{err}(H, f, \overline{a}) < \epsilon\}\right) > 1 - \delta.$$

The conclusion follows.

EXERCISE 2.3.2. Fix $m > 0$ and suppose $X = \mathbb{R}^m$ and $\mathcal{C}$ is the concept class of axes-parallel boxes in $\mathbb{R}^m$. Show that the above example generalizes, with sample complexity

$$\langle \epsilon, \delta \rangle \mapsto N_{\epsilon,\delta} \leq \frac{2m}{\epsilon} \log\left(\frac{2m}{\delta}\right).$$

EXERCISE 2.3.3. Suppose $X = \mathbb{R}^2$ and $\mathcal{C}$ is the concept class of all closed disks in $\mathbb{R}^2$. Show that $\mathcal{C}$ is PAC-learnable and find an appropriate sample complexity (use geometry and not Proposition 2.3.6 below).

What is the expected value of the error function for a PAC-learnable class?

LEMMA 2.3.4. *Let $\mathcal{C}$ be a PAC-learnable concept class on X and let H be a witnessing hypothesis function with sample complexity $\langle \epsilon, \delta \rangle \mapsto N_{\epsilon,\delta}$. Then, for all $\epsilon > 0$, $\delta > 0$, μ a probability measure on X, and $n \geq N_{\epsilon,\delta}$,*

$$E\left(\overline{a} \mapsto \operatorname{err}_\mu(H, f, \overline{a})\right) \leq \delta + \epsilon(1 - \delta).$$

PROOF. Let

$$Y_0 = \{\overline{a} \in X^n : \operatorname{err}_\mu(H, f, \overline{a}) > \epsilon\}$$

and let $Y_1 = (X^n \setminus Y_0)$. By definition, $\mu^n(Y_0) < \delta$, so

$$E\left(\overline{a} \mapsto \operatorname{err}_\mu(H, f, \overline{a})\right) = \int_{X^n} \operatorname{err}_\mu(H, f, -) d\mu^n \leq \mu^n(Y_0) + \epsilon\mu^n(Y_1) \leq \delta + \epsilon(1 - \delta).$$

□

The main result of this subsection is the following:

THEOREM 2.3.5 (Theorem 2.1 of [**4**]). *Let X be a set and $\mathcal{C}$ be a concept class on X. Then, the following are equivalent:*

(1) *$\mathcal{C}$ is a VC-class.*
(2) *$\mathcal{C}$ is PAC-learnable.*

More precisely, we have the following bounds:

PROPOSITION 2.3.6. *Fix $d < \omega$, let X be a set, and let $\mathcal{C}$ be a concept class on X with VC-dimension $\leq d$. Then, any consistent hypothesis function $H : \mathcal{C}_{\text{fin}} \to \mathcal{C}$ is a witness to PAC-learnability for $\mathcal{C}$ with sample complexity*

$$\langle \epsilon, \delta \rangle \mapsto N_{\epsilon,\delta} \leq \max\left\{ \frac{4}{\epsilon} \log_2\left(\frac{2}{\delta}\right), \frac{8d}{\epsilon} \log_2\left(\frac{13}{\epsilon}\right) \right\}.$$

PROPOSITION 2.3.7. *Fix $d < \omega$, let X be a set, and let $\mathcal{C}$ be a concept class on X with VC-dimension $\geq d$. Then, any hypothesis function $H : \mathcal{C}_{\text{fin}} \to {}^{X}2$ that is a witness to PAC-learnability for $\mathcal{C}$ has sample complexity*

$$\langle \epsilon, \delta \rangle \mapsto N_{\epsilon,\delta} \geq d(1 - 2(\epsilon(1-\delta) + \delta)).$$

PROPOSITION 2.3.8. *Let X be a set and let $\mathcal{C}$ be a concept class on X with at least three elements. Then, any hypothesis function $H : \mathcal{C}_{\text{fin}} \to {}^{X}2$ that is a witness to PAC-learnability for $\mathcal{C}$ has sample complexity*

$$\langle \epsilon, \delta \rangle \mapsto N_{\epsilon,\delta} \geq \frac{1-\epsilon}{\epsilon} \log\left(\frac{1}{\delta}\right).$$

We will leave the proof of Proposition 2.3.8 as an exercise to the reader. As a hint, concentrate the measure onto two points. This proposition is interesting because it gives an absolute lower bound for the sample complexity regardless of any conditions on $\mathcal{C}$ (except non-triviality). The other two propositions give bounds in both directions for the sample complexity of a concept class with a finite VC-dimension.

PROOF OF PROPOSITION 2.3.7. Let $H : \mathcal{C}_{\text{fin}} \to {}^{X}2$ be any hypothesis function witnessing the PAC-learnability of $\mathcal{C}$. Since $\mathcal{C}$ has VC-dimension $\geq d$, there is a subset of X of size d that is shattered. By concentrating the measure on the shattered set, we may assume that $|X| = d$ and $\mathcal{C} = {}^{X}2$. Let μ denote the uniform probability measure on X and let μ also denote the uniform probability measure on $\mathcal{C}$ (i.e., for all $a \in X$, $\mu(\{a\}) = 1/d$ and for all $f \in \mathcal{C}$, $\mu(\{f\}) = 2^{-d}$). Fix $n \leq d$.

Let $\overline{a} \in X^n$, let $Y = \{a_1, ..., a_n\}$ enumerate $\overline{a}$, and let $\ell = |Y|$ (which is $\leq n$). For each $g \in {}^{Y}2$, the number of functions $f \in \mathcal{C}$ extending g is $2^{d-\ell}$. Regardless of the value of $H(g)$, for each $b \in (X \setminus Y)$, exactly half of the functions f extending g differ from $H(g)$ on b. That is,

$$|\{f \in \mathcal{C} : f \text{ extends } g \text{ and } [H(g)](b) \neq f(b)\}| = \frac{1}{2}\left(2^{d-\ell}\right).$$

Hence, for each $g \in {}^{Y}2$,

$$\sum \left\{ |H(g)(b) - f(b)| : b \in X, f \in {}^{X}2, f \text{ extends } g \right\} \geq \frac{1}{2} 2^{d-\ell}(d-\ell).$$

So, since $|{}^{Y}2| = 2^{\ell}$, we get

$$\sum \left\{ |H(f|_Y)(b) - f(b)| : b \in X, f \in {}^{X}2 \right\} \geq \frac{1}{2} 2^{d}(d-\ell) \geq \frac{1}{2} 2^{d}(d-n).$$

Thus,

$$\sum \left\{ \mathrm{err}_{\mu}(H, f, \overline{a}) : b \in X, f \in {}^{X}2 \right\} \geq \frac{2^d}{2d}(d-n).$$

Thereby, since $\left|{}^{X}2\right| = 2^d$, the expected error is

$$E\left(f \mapsto \mathrm{err}_{\mu}(H, f, \overline{a})\right) \geq \frac{d-n}{2d}.$$

Since this holds for any $\overline{a} \in X^n$, we get (on the probability space $\mathcal{C} \times X^n$),

$$E\left(\langle f, \overline{a}\rangle \mapsto \mathrm{err}_{\mu}(H, f, \overline{a})\right) \geq \frac{d-n}{2d}.$$

By Lemma 1.3.1, there exists $f \in \mathcal{C}$ so that,

$$E\left(\overline{a} \mapsto \mathrm{err}_{\mu}(H, f, \overline{a})\right) \geq \frac{d-n}{2d}.$$

By Lemma 2.3.4,

$$E\left(\overline{a} \mapsto \mathrm{err}_{\mu}(H, f, \overline{a})\right) \leq \epsilon(1-\delta) + \delta.$$

Therefore,

$$n \geq d(1 - 2(\epsilon(1-\delta) + \delta)).$$

□

To prove the other direction, i.e. Proposition 2.3.6, we use ϵ-nets. First, do the following exercise:

Exercise 2.3.9 (Lemma 2A.3 of [**4**]). Suppose $\mathcal{C} \subseteq {}^{X}2$ and $f \in {}^{X}2$. Show that, for all m, $\pi_{\mathcal{C}}(m) = \pi_{(\mathcal{C} \triangle f)}(m)$, where

$$(\mathcal{C} \triangle f) = \{|g - f| : g \in \mathcal{C}\}.$$

Therefore, the VC-dimension of $\mathcal{C}$ is equal to the VC-dimension of $(\mathcal{C} \triangle f)$. Hint: For all $Y \in \binom{X}{m}$, show that, for all $g_1, g_2 \in \mathcal{C}$, $g_1|_Y = g_2|_Y$ if and only if $(|g_1 - f|)|_Y = (|g_2 - f|)|_Y$.

Proof of Proposition 2.3.6. Fix some $f \in \mathcal{C}$. By Exercise 2.3.9, since $\mathcal{C}$ has VC-dimension $\leq d$, $(\mathcal{C} \triangle f)$ has VC-dimension $\leq d$. By Theorem 2.2.7, we get that

$$\mu\big(\{\overline{a} \in X^n : \{a_1, ..., a_n\} \text{ is not an } \epsilon\text{-net for } (\mathcal{C} \triangle f)\}\big) \leq 2(2n)^d 2^{-\epsilon n/2}.$$

This is independent of the choice of probability measure, μ. One can show that, if

$$n \geq \max\left\{\frac{4}{\epsilon}\log_2\left(\frac{2}{\delta}\right), \frac{8d}{\epsilon}\log_2\left(\frac{13}{\epsilon}\right)\right\},$$

then

$$2(2n)^d 2^{-\epsilon n/2} \leq \delta$$

(it is easy to show that

$$\lim_{n\to\infty} 2(2n)^d 2^{-\epsilon n/2} = 0,$$

so it is not difficult to exhibit that there is some bound, dependent only on ϵ, δ, and d, that works).

For each $\overline{a} \in X^n$,

$$\mathrm{err}_\mu(H, f, \overline{a}) = \mu\left(|f - H(f|_{\overline{a}})|\right).$$

That is, the error of a hypothesis $h = H(g) \in \mathcal{C}$ from f is exactly the μ-measure of $|f - h|$. Suppose $\overline{a} \in X$ is such that

(1) $\{a_1, ..., a_n\}$ is an ϵ-net for $(\mathcal{C}\triangle f)$, and
(2) $\mathrm{err}_\mu(H, f, \overline{a}) \geq \epsilon$.

Then, by definition of ϵ-net, for some $i = 1, ..., n$, $|f(a_i) - [H(f|_{\overline{a}})](a_i)| = 1$ (note that it is vital here that we are assuming that the image of H is in $\mathcal{C}$ and not all of ${}^X2$). That is,

$$f(a_i) \neq [H(f|_{\overline{a}})](a_i).$$

However, this is contrary to the fact that H is consistent. Therefore,

$$\mu\left(\{\overline{a} \in X^n : \mathrm{err}_\mu(H, f, \overline{a}) > \epsilon\}\right) < 2(2n)^d 2^{-\epsilon n/2} \leq \delta.$$

□

We combine these to prove the main theorem.

Proof of Theorem 2.3.5. If $\mathcal{C}$ is a VC-class, then there exists $d < \omega$ so that $\mathcal{C}$ has VC-dimension $\leq d$. Hence, by Proposition 2.3.6, any consistent hypothesis function is a witness to the fact that $\mathcal{C}$ is PAC-learnable.

Conversely, if $\mathcal{C}$ is not a VC-class, then, for all $d < \omega$, $\mathcal{C}$ has VC-dimension $\geq d$. Hence, by Proposition 2.3.7, any hypothesis function has is a supposed witness to PAC-learnability has sample complexity

$$\langle \epsilon, \delta \rangle \mapsto N_{\epsilon,\delta} \geq d(1 - 2(\epsilon(1-\delta) + \delta)).$$

Hence, no such function exists and $\mathcal{C}$ is not PAC-learnable. □

3.2. Compression Schemes. Another measure of complexity is that of compressibility. Fix a concept class $\mathcal{C}$ on X and recall that

$$\mathcal{C}_{\text{fin}} = \{f|_Y : Y \subseteq X, Y \text{ finite}\}.$$

This codes all the finite parts of elements of $\mathcal{C}$.

Definition 2.3.10 ([**7**]). We say that $\mathcal{C}$ has a *d-dimensional (sequence) compression scheme* if there exists a compression function $\kappa : \mathcal{C}_{\text{fin}} \to X^d$ with $\kappa(f) \in \text{dom}(f)^d$ for each non-empty $f \in \mathcal{C}_{\text{fin}}$ and a finite number of recovery functions $\rho_i : X^d \to {}^X 2$ (for $i = 1, ..., N$) such that, for each $f \in \mathcal{C}_{\text{fin}}$, there exists i such that

$$\rho_i(\kappa(f))|_Y = f.$$

The idea here is that we can code the complexity of $\mathcal{C}$ on finite parts. For each finite $Y \subseteq X$ and $f \in \mathcal{C}$, there exists an element in Y^d that "codes" $f|_Y$. That is, we can recover $f|_Y$ from the element via ρ_i for some i.

Example 2.3.11 (Example 2.3.1, revisited). We will consider compressing the notion of "room temperature" as above. Consider $X = \mathbb{R}$ and $\mathcal{C} = \{\chi_{(a,b)} : a, b \in \mathbb{R}, a < b\}$. For the compression function $\kappa : \mathcal{C}_{\text{fin}} \to X^2$, for any $Y \subseteq X$ finite non-empty and $f \in \mathcal{C}$, let $\kappa(f|_Y)$ be $\langle a, b\rangle \in Y^2$, where $a \in Y$ is minimal such that $f(a) = 1$ and $b \in Y$ is maximal such that $f(b) = 1$. If f is uniformly zero, then return $\langle a, a\rangle$ for any $a \in Y$. For recovery functions, let $\rho_0 : X^d \to {}^X 2$ be given by

$$[\rho_0(\langle a, b\rangle)](c) = 1 \text{ iff. } a \leq c \leq b$$

and let $\rho_1 : X^d \to {}^X 2$ be identically zero. Then, we check that this is a 2-dimensional compression scheme. Suppose $Y \subseteq X$ finite non-empty and $f \in \mathcal{C}$. In the first case, if $f(a) = 1$ for some $a \in Y$, then $\rho_0(\kappa(f|_Y))|_Y = f|_Y$. On the other hand, if $f(a) = 0$ for all $a \in Y$, then $\rho_1(\kappa(f|_Y))|_Y = f|_Y = \mathbf{0}_Y$.

Exercise 2.3.12. Show that $\mathcal{C}$ in Example 2.3.11 above does not have a 1-dimensional compression scheme.

Exercise 2.3.13. Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the concept class of open intervals in $\mathbb{R}$. Show that $\mathcal{C}$ has a 2-dimensional compression scheme (Hint: Endpoints). In general, let $X = \mathbb{R}^n$ and let $\mathcal{C}_n$ be the concept class of axes-parallel open n-boxes. Show that $\mathcal{C}_n$ has a $2n$-dimensional compression scheme.

Lemma 2.3.14. *If $\mathcal{C}$ is a concept class with a d-dimensional compression scheme, then $\mathcal{C}$ has VC-density $\leq d$. In particular, $\mathcal{C}$ is a VC-class.*

PROOF. Let κ and ρ_i for $i = 1, ..., N$ be a d-dimensional compression scheme for $\mathcal{C}$ (as in Definition 2.3.10 above). We aim to show that $\pi_{\mathcal{C}}(m) \leq Nm^d$. For each set $Y \subseteq X$ with $|Y| = m$, an element of $\mathcal{C}|_Y$ is determined by an element in Y^d together with ρ_i for a choice of $i \in \{1, ..., N\}$. In other words,

$$\mathcal{C}|_Y \subseteq \{\rho_i(\overline{x})|_Y : \overline{x} \in Y^d \text{ and } i = 1, ..., N\}.$$

This amounts to Nm^d possibilities. Hence,

$$|\mathcal{C}|_Y| \leq Nm^d,$$

as desired. □

Does this reverse? This is a major open question.

CONJECTURE 2.3.15 (Warmuth Conjecture [**7**]). *If $\mathcal{C}$ is a VC-class, then $\mathcal{C}$ has a compression scheme.*

We will provide many partial results to this question using model theory.

4. Link to Model Theory

Let T be a first-order theory in a fixed language L. Let $M \models T$ and fix $\varphi(x; y)$ an L-formula. This generates a concept class on M_y, namely

$$\mathcal{C}_\varphi^M = \left\{\chi_{\varphi(a;M)} : a \in M_x\right\},$$

where $\chi_{\varphi(a;M)}(b) = 1$ if and only if $M \models \varphi(a; b)$.

We will use this translation to measure the complexity of φ. The *VC-dimension* of $\varphi(x; y)$ in M is the VC-dimension of $\mathcal{C}_\varphi^M$ as in Definition 2.1.1 above. The VC-density of $\varphi(x; y)$ in M is the VC-density of $\mathcal{C}_\varphi^M$ as in Definition 2.1.9 above. We will say that $\varphi(x; y)$ has *NIP* (for "not the independence property", or sometimes called *dependent*) in M if $\mathcal{C}_\varphi^M$ is a VC-class. We say the L-structure M has *NIP* (or is *dependent*) if all formulas have NIP in M. How do we translate "compression schemes?"

DEFINITION 2.4.1. We say a formula $\varphi(x; y)$ has *UDTFS* (for "uniform definability of types over finite sets") in M if there exists a finite collection of formulas $\psi_i(y; z_1, ..., z_d)$ for $i = 1, ..., N$ such that, for each finite $B \subseteq M_y$ and each $a \in M_x$, there exists $c_1, ..., c_d \in B$ and i such that

$$(\forall b \in B)[M \models \varphi(a; b) \leftrightarrow \psi_i(b; c_1, ..., c_d)].$$

We will say that $\varphi(x; y)$ has *UDTFS rank* d in M for the minimal such d (regardless of the size of N).

PROPOSITION 2.4.2. *If $\varphi(x;y)$ has UDTFS rank d in M, then $\mathcal{C}_{\varphi}^{M}$ has a d-dimensional compression scheme (as in Definition 2.3.10 above).*

PROOF. Let $\psi_i(y; z_1, ..., z_d)$ witness that $\varphi(x;y)$ has UDTFS rank d. Define $\kappa : (\mathcal{C}_{\varphi}^{M})_{\text{fin}} \to M_y^d$ in the following manner. For $a \in M_x$ and finite $B \subseteq M_y$, consider $f = \chi_{\varphi(a;M)}|_B$ (the general form of elements in $(\mathcal{C}_{\varphi}^{M})_{\text{fin}}$). Let $\kappa(f)$ be the element $\langle c_1, ..., c_d \rangle \in B^d$ given by the definition of UDTFS rank d. Now define $\rho_i : M_y^d \to {}^{M_y}2$ as follows:

$$[\rho_i(c_1, ..., c_d)](b) = 1 \text{ if and only if } M \models \psi_i(b; c_1, ..., c_d).$$

Now it is easy to check that κ and ρ_i form a compression scheme for $\mathcal{C}_{\varphi}^{M}$. □

COROLLARY 2.4.3. *If $\varphi(x;y)$ has UDTFS rank d in M, then φ has VC-density $\leq d$ in M.*

PROOF. Follows immediately from Proposition 2.4.2 and Lemma 2.3.14. □

We can go the other direction, too. Suppose X is any set and $\mathcal{C}$ is a concept class on X. Let $L = \{R\}$ be the language with a binary relation symbol $R(x, y)$ on two different sorts. Let $M = M_{\mathcal{C}}$ be the two-sorted L-structure whose universe is $X \sqcup \mathcal{C}$ and where we interpret R as

$$R^M(x, f) \text{ if and only if } f(x) = 1. \tag{2.5}$$

It is easy now to check the VC-dimension of $\mathcal{C}$ is equal to the VC-dimension of R in M. The same holds for VC-density. In a manner similar to the proof of Proposition 2.4.2, one can show the following:

PROPOSITION 2.4.4. *Let $\mathcal{C}$ be a concept class on a set X. If R has UDTFS rank d in $M_{\mathcal{C}}$, then $\mathcal{C}$ has a d-dimensional compression scheme.*

The main goal the remainder of this book will be to prove Chernikov and Simon's partial result to the Warmuth Conjecture.

THEOREM 2.4.5. *If $\mathcal{C}$ is a concept class such that $M_{\mathcal{C}}$ has NIP, then $\mathcal{C}$ has a compression scheme.*

Along the way we will also use Model Theory to prove other results. The following theorem lists all such results:

THEOREM 2.4.6. *Let $\mathcal{C}$ be a concept class on X.*

(1) *(Theorem 3.3.14 below) If $\mathcal{C}$ has VC-density < 2, then $\mathcal{C}$ has a compression scheme.*

(2) *(Theorem 4.1.3 below) If R is stable in $M_{\mathcal{C}}$, then $\mathcal{C}$ has a compression scheme (whose dimension is bounded by the Shelah 2-rank of R).*
(3) *(Theorem 3.3.10 below) If $\mathcal{C}$ is maximum of dimension d, then $\mathcal{C}$ has a compression scheme of dimension d.*
(4) *(Corollary 3.3.5 below) If $\varphi(x; y)$ is a formula in a weakly o-minimal theory T and $M \models T$, then $\mathcal{C}_{\varphi}^{M}$ has a compression scheme of dimension $|x|$.*
(5) *(Theorem 4.3.2 below) If $\varphi(x; y)$ is a formula in a strongly minimal theory T and $M \models T$, then $\mathcal{C}_{\varphi}^{M}$ has VC-density $\leq |x|$.*

So now we switch gears and move into the realm of model theory.

CHAPTER 3

Theories with NIP

1. Introduction to NIP Theories

Let T be a first-order theory in a fixed language L and let $\mathcal{U}$ be a large, sufficiently saturated and homogeneous model of T (a so called "monster model" of T).

We say that a formula $\varphi(x;y)$ does not have the *independence property* (has *NIP*) if $\mathcal{C}_{\varphi}^{\mathcal{U}}$ is a VC-class. In other words, there exists $n < \omega$ such that, for all $b_0, ..., b_{n-1} \in \mathcal{U}_y$, there do not exist $a_I \in \mathcal{U}_x$ for each $I \subseteq n$ so that

$$\text{for all } i < n, I \subseteq n, \models \varphi(a_I; b_i) \text{ iff. } i \in I.$$

Say that the theory T is *NIP* if all formulas have NIP.

This is related to indiscernible sequences.

Definition 3.1.1. Fix a set of formulas $\Delta(y_1, ..., y_n)$ (each y_i of the same sort), $\langle I; < \rangle$ a linear order, and $b_i \in \mathcal{U}_{y_1}$ for each $i \in I$. We say that the sequence $\langle b_i : i \in I \rangle$ is Δ*-indiscernible* if, for all $i_1 < ... < i_n$ from I and $j_1 < ... < j_n$ from I, for all $\delta \in \Delta$,

$$\models \delta(b_{i_1}, ..., b_{i_n}) \leftrightarrow \delta(b_{j_1}, ..., b_{j_n}).$$

For a set $C \subseteq \mathcal{U}$, we say that $\langle b_i : i \in I \rangle$ is *indiscernible over* C if, for all sets of formulas $\Delta(y_1, ..., y_n)$ over C (for all $n < \omega$), $\langle b_i : i \in I \rangle$ is Δ-indiscernible. If $C = \emptyset$, we drop "over C."

Definition 3.1.2. Fix a formula $\varphi(x;y)$ and $n < \omega$. Say that $\varphi(x;y)$ has *alternation rank* $\geq n$ if, for all indiscernible sequences $\langle b_i : i \in I \rangle$ of elements from $\mathcal{U}_y$ and all $a \in \mathcal{U}_x$, there exist $i_0 < ... < i_n$ from I such that, for all $i < n$,

$$\models \varphi(a; b_i) \leftrightarrow \neg\varphi(a; b_{i+1}).$$

That is, the truth value of $\varphi(a; -)$ alternates at least n times. We get the following:

Proposition 3.1.3. *For any formula $\varphi(x;y)$, φ has NIP if and only if φ has finite alternation rank.*

Proof. ($\Leftarrow$): Suppose φ has the independence property (does not have NIP). We proceed in two steps. First we show, by compactness, there exists $b_i \in \mathcal{U}_y$ for $i < \omega$ and $a_I \in \mathcal{U}_x$ for $I \subseteq \omega$ such that, for all $i < \omega$ and $I \subseteq \omega$,

$$\models \varphi(a_I; b_i) \text{ iff. } i \in I.$$

To do this, let $L^* = L \cup \{c_I : I \subseteq \omega\} \cup \{d_i : i < \omega\}$ be a new language where each c_I is a new constant of the x sort and each d_i is a new constant of the y sort. Let

$$\Sigma = T \cup \{\varphi(c_I; d_i) : i \in I\} \cup \{\neg\varphi(c_I; d_i) : i \notin I\}$$

be a set of L^*-sentences. To show Σ is consistent, by compactness, we only need to show that Σ_0 is consistent for each finite $\Sigma_0 \subseteq \Sigma$. Fix such a Σ_0. Since Σ_0 is finite, there exists $N < \omega$ such that the only c_I and d_i mentioned are such that $I \subseteq N$ and $i < N$. Since φ has NIP, a set of size N is shattered by instances of φ. Thus, there exists $b_i \in \mathcal{U}_y$ for $i < N$ and $a_I \in \mathcal{U}_x$ for $I \subseteq N$ such that $\models \varphi(a_I; b_i)$ if and only if $i \in I$. Hence, interpreting d_i as b_i and c_I as a_I we realize Σ_0.

Next we show, by compactness and Ramsey's Theorem, there exists an infinite indiscernible sequence $\langle b_i : i < \omega \rangle$ (of the y sort) and $a_I \in \mathcal{U}_x$ for $I \subseteq \omega$ such that, for all $i < \omega$ and $I \subseteq \omega$,

$$\models \varphi(a_I; b_i) \text{ iff. } i \in I.$$

Still using L^* as above, let

$$\Gamma = \Sigma \cup \{\delta(b_{i_0}, ..., b_{i_{n-1}}) \leftrightarrow \delta(b_{j_0}, ..., b_{j_{n-1}}) : \\ i_0 < ... < i_{n-1} < \omega, j_0 < ... < j_{n-1} < \omega, \delta \in L\}$$

be a set of L^*-sentences (here we let δ range over all L-formulas with n variables of the y sort for all varying $n < \omega$). Again, it suffices to show Γ_0 is consistent for some finite $\Gamma_0 \subseteq \Gamma$, so fix such a Γ_0. As above, fix $N < \omega$ sufficiently large. Let n be the maximum for the δ's appearing in Γ_0 and we may suppose, by adding dummy variables, that all the δ have n variables of the y sort. Let $\Delta(y_0, ..., y_{n-1})$ list off all such formulas used in Γ_0. Now Ramsey's Theorem states that, for any finite set X and any map

$$f : \binom{\omega}{n} \to X$$

there exists $J \subseteq \omega$ with $|J| = \aleph_0$ such that f is constant on $\binom{J}{n}$. In particular, let $X = 2^\Delta$ and, for $I \in \binom{\omega}{n}$, we can set $f(i_0, ..., i_{n-1})$ (where $I = \{i_0, ..., i_{n-1}\}$ in ascending order) equal to $\langle \epsilon_\delta : \delta \in \Delta \rangle \in 2^\Delta$, where $\epsilon_\delta = 1$ if and only if $\models \delta(b_{i_0}, ..., b_{i_{n-1}})$ for each $\delta \in \Delta$. Then, applying Ramsey's Theorem, we get $J \subseteq \omega$ with $|J| = \aleph_0$ so that

$\langle b_i : i \in J\rangle$ is Δ-indiscernible (since each n-tuple matches all truth values on Δ). Reindexing, we may suppose that $\langle b_i : i < \omega\rangle$ is Δ-indiscernible. Therefore, with our same interpretation as above, Γ_0 is realized.

Finally, if we take I to be the set of even numbers, the truth value of $\varphi(a_I; -)$ alternates infinitely many times.

($\Rightarrow$): Fix $n < \omega$ and suppose φ has alternation rank $\geq 2n$. Then, there exists an indiscernible sequence $\langle b_i : i < 2n\rangle$ from $\mathcal{U}_y$ and $a \in \mathcal{U}_x$ so that

$$\models \varphi(a; b_i) \text{ iff. } i \text{ is even}$$

(take a sequence witnessing alternation rank $2n$ and take an appropriate subsequence). Then, for any $s : n \to 2$

$$\models \bigwedge_{i<n} \varphi(a; b_{2i+s(i)})^{s(i)}.$$

Therefore, $\models \delta_s(b_{s(0)}, ..., b_{2n-2+s(n-1)})$ holds, where

$$\delta_s(y_0, ..., y_{n-1}) = \exists x \left(\bigwedge_{i<n} \varphi(a; y_i)^{s(i)} \right),$$

As witnessed by a. Hence, by indiscernibility, $\models \delta_s(b_0, ..., b_{n-1})$ (notice that $s(0) < 2 + s(1) < ... < 2n - 2 + s(n-1)$, so we can "slide" this down to $0 < ... < n-1$). Therefore,

$$\models \exists x \left(\bigwedge_{i<n} \varphi(a; b_i)^{s(i)} \right).$$

Since s was arbitrary, $\langle b_i : i < n\rangle$ is shattered by instances of φ. Therefore, φ has independence dimension $\geq n$. Since n was arbitrary, φ has NIP. □

Suppose that I and J are two infinite linear orders. We say that two indiscernible sequences $\overline{a} = \langle a_i : i \in I\rangle$ and $\overline{b} = \langle b_j : j \in J\rangle$ *have the same EM-type* if, for all $n < \omega$, all formulas $\delta(y_0, ..., y_{n-1})$, all $i_0 < ... < i_{n-1}$ from I, and all $j_0 < ... < j_{n-1}$ from J,

$$\models \delta(a_{i_0}, ..., a_{i_{n-1}}) \leftrightarrow \delta(b_{j_0}, ..., b_{j_{n-1}}).$$

Exercise 3.1.4. If $\varphi(x; y)$ has alternation rank $\geq n$ witnessed by an indiscernible sequence $\overline{a}$ and the sequence $\overline{b}$ has the same EM-type as $\overline{a}$, show that $\overline{b}$ is also a witness to the fact that $\varphi(x; y)$ has alternation rank $\geq n$.

Proposition 3.1.5. *If T is a theory such that, for all formulas $\varphi(x; y)$ with $|x| = 1$, we have that $\varphi(x; y)$ has NIP, then T has NIP.*

To prove this, consider the following lemma.

LEMMA 3.1.6. *Fix a variable* x, $\kappa > |T|$ *(i.e.,* $\kappa > |L| + \aleph_0$*),* $\langle b_i : i < \kappa \rangle$ *is an indiscernible sequence, and* $a \in \mathcal{U}_x$*. Suppose that, for all formulas* $\delta(x; z)$*,* δ *has NIP. Then, there exists* $\alpha < \kappa$ *such that* $\langle b_i : \alpha < i < \kappa \rangle$ *is indiscernible over* $\{a\}$*.*

PROOF. Suppose not. Then, for each $\alpha < \kappa$, there exists an L-formula $\delta_\alpha(x; y_0, ..., y_{n_\alpha})$, $\alpha < i_0 < ... < i_{n_\alpha} < \kappa$, and $\alpha < j_0 < ... < j_{n_\alpha} < \kappa$ such that

$$\models \delta_\alpha(a; b_{i_0}, ..., b_{i_{n_\alpha}}) \wedge \neg\delta_\alpha(a; b_{j_0}, ..., b_{j_{n_\alpha}}).$$

Hence, by pigeon hole principle, there exists $A \subseteq \kappa$ cofinal and an L-formula $\delta(x; y_0, ..., y_n)$ so that $\delta = \delta_\alpha$ for all $\alpha \in A$. Now construct a new indiscernible sequence $\langle c_i : i < \omega \rangle$ by induction as follows.

Choose $\alpha_0 \in A$ arbitrary, and $\alpha_0 < i_0^0 < ... < i_n^0 < \kappa$ such that $\models \neg\delta(a; b_{i_0^0}, ..., b_{i_n^0})$. Set $c_0 = \langle b_{i_0^0}, ..., b_{i_n^0} \rangle$ and choose $\alpha_1 \in A$ such that $\alpha_1 > i_n^0$. In general, if α_ℓ is chosen, pick

$$\alpha_\ell < i_0^\ell < ... < i_n^\ell < \kappa \text{ such that } \models \neg\delta(a; b_{i_0^\ell}, ..., b_{i_n^\ell})^{\ell \pmod 2}.$$

Let $c_\ell = \langle b_{i_0^\ell}, ..., b_{i_n^\ell} \rangle$ and fix $\alpha_{\ell+1} \in A$ such that $\alpha_{\ell+1} > i_n^\ell$. This concludes the construction of $\langle c_i : i < \omega \rangle$. Notice, however, that $\delta(x; y_0, ..., y_n)$ has NIP by assumption. Therefore, it has finite alternation rank by Proposition 3.1.3. However, $\models \delta(a; c_\ell)$ if and only if ℓ is odd, a contradiction. □

PROOF OF PROPOSITION 3.1.5. Fix a variable x. By induction on $n = |x|$ and repeated applications of Lemma 3.1.6, we get that if $a \in \mathcal{U}_x$ and $\langle b_i : i < \kappa \rangle$ is an indiscernible sequence for $\kappa > |T|$, then there exists $\alpha < \kappa$ such that $\langle a^\frown b_i : \alpha < i < \kappa \rangle$ is an indiscernible sequence. Therefore, if $\varphi(x; y)$ is any formula, $a \in \mathcal{U}_x$, and $\langle b_i : i < \kappa \rangle$ is an indiscernible sequence, then there exists $\alpha < \kappa$ so that the truth value of $\varphi(a; b_i)$ is constant on $\alpha < i < \kappa$.

On the other hand, suppose that φ had IP. Then, by compactness, there exists $a \in \mathcal{U}_x$ and $\langle b_i : i < \kappa \rangle$ such that $\models \varphi(a; b_i)$ for cofinally many $i \in \kappa$ and $\models \neg\varphi(a; b_i)$ for cofinally many $i \in \kappa$. Contradiction. Therefore, φ has NIP. □

EXERCISE 3.1.7. Suppose that $\varphi(x; y)$ and $\psi(x; z)$ have NIP. Show that $(\varphi \wedge \psi)(x; y, z)$ has NIP and $\neg\varphi(x; y)$ has NIP. Exhibit a specific example to show that NIP is *not* closed under existential quantification.

We can use this produce many examples of theories that have NIP. For example, all stable theories have NIP and all weakly o-minimal theories have NIP (see definitions below).

Exercise 3.1.8. Many algebraic examples have NIP. Show that the following theories have NIP (you are allowed to assume quantifier elimination in the given language):

(1) Divisible ordered abelian groups, $T = \mathrm{Th}(\mathbb{Q}; +, <)$.
(2) Algebraically closed fields, $T = \mathrm{Th}(\mathbb{C}; +, \cdot)$.
(3) Real closed fields, $T = \mathrm{Th}(\mathbb{R}; +, \cdot, <)$.

2. Invariant Types

Let T be a first-order theory in a fixed language L and let $\mathcal{U}$ be a monster model of T. Let $A \subseteq \mathcal{U}$ be a small subset (where "small" means of cardinality much less than the saturation of $\mathcal{U}$). For a fixed variable x, consider $S_x(\mathcal{U})$ the space of types over $\mathcal{U}$ in x. These are called the *global types*, as we think of $\mathcal{U}$ as being the universe.

Definition 3.2.1. We say that $p(x) \in S_x(\mathcal{U})$ is *invariant* over A if, for all automorphisms σ of $\mathcal{U}$ fixing A, $p(x)$ is fixed under σ. That is, for all formulas $\varphi(x; y)$ and all $b \in \mathcal{U}_y$, $\varphi(x; b) \in p(x)$ if and only if $\varphi(x; \sigma(b)) \in p(x)$.

Remark 3.2.2. Under our assumptions (namely that of homogeneity), this is equivalent to non-splitting. In general, we say that $p(x)$ *does not split over* A if, for all $b_1, b_2 \in \mathcal{U}_y$ with $\mathrm{tp}(b_1/A) = \mathrm{tp}(b_2/A)$, $\varphi(x; b_1) \in p(x)$ if and only if $\varphi(x; b_2) \in p(x)$.

There are two ways to be invariant.

Definition 3.2.3. Fix $p(x) \in S_x(\mathcal{U})$ and $A \subseteq \mathcal{U}$ small.

(1) We say p is *definable* over A if, for all formulas $\varphi(x; y)$, there exists a formula $\psi(y; a)$ over A such that, for all $b \in \mathcal{U}_y$, $\varphi(x; b) \in p(x)$ if and only if $\models \psi(b; a)$.
(2) We say p is *finitely satisfiable* over A if, for all formulas $\varphi(x; b)$ in $p(x)$, there exists $a \in A_x$ such that $\models \varphi(a; b)$.

Lemma 3.2.4. *If $p(x) \in S_x(\mathcal{U})$ is definable over A or finitely satisfiable over A, then $p(x)$ is invariant over A.*

Proof. (1): Suppose $p(x)$ is definable over A, $\varphi(x; y)$ is any formula, and $b_1, b_2 \in \mathcal{U}_y$ are such that $\mathrm{tp}(b_1/A) = \mathrm{tp}(b_2/A)$. Then, in particular, if $\psi(y; a)$ is the definition of φ in p, $\models \psi(b_1; a)$ if and only if $\models \psi(b_2; a)$. Thus, $\varphi(x; b_1) \in p(x)$ if and only if $\varphi(x; b_2) \in p(x)$.

(2): Suppose $p(x)$ is finitely satisfiable over A. On the other hand, suppose $\varphi(x; b_1), \neg\varphi(x; b_2) \in p(x)$. Then, there exists $a \in A_x$ so that

$$\models \varphi(a; b_1) \wedge \neg\varphi(a; b_2).$$

Therefore, $\mathrm{tp}(b_1/A) \neq \mathrm{tp}(b_2/A)$. $\square$

It may be hard to find definable types, but it is easy to find finitely satisfiable types. Fix A a small subset of $\mathcal{U}$ and let $\mathcal{F}$ be an ultrafilter on A. Define the average type

$$\mathrm{Av}_x(\mathcal{F},\mathcal{U}) = \{\varphi(x;b) : \varphi(x;b) \in L(\mathcal{U}), \varphi(A_x;b) \in \mathcal{F}\}.$$

EXERCISE 3.2.5. Show that $\mathrm{Av}_x(\mathcal{F},\mathcal{U})$ is a type over $\mathcal{U}$ that is finitely satisfiable over A.

Therefore, in particular, $\mathrm{Av}_x(\mathcal{F},\mathcal{U})$ is an invariant type over A.

DEFINITION 3.2.6. Fix a small model M and a type $p(x) \in S_x(M)$. A *coheir* of p is a global extension $q \in S_x(\mathcal{U})$ of p that is finitely satisfiable over M.

To see that coheirs always exist, consider the set $\mathcal{A} := \{\theta(M) : \theta(x) \in p(x)\} \subseteq \mathcal{P}(M)$. Since M is a model, this has the finite intersection property. So, by Proposition 1.4.2, there exists $\mathcal{F} \supseteq \mathcal{A}$ an ultrafilter on M. By Exercise 3.2.5, the type $\mathrm{Av}_x(\mathcal{F},\mathcal{U})$ is a coheir of p.

For a fixed small $A \subseteq \mathcal{U}$, Consider the set of types invariant over A:

$$S_x^{\mathrm{inv}}(\mathcal{U}, A) = \{p \in S_x(\mathcal{U}) : p \text{ is invariant over } A\}.$$

LEMMA 3.2.7. *The set $S_x^{\mathrm{inv}}(\mathcal{U}, A)$ is a closed subset of $S_x(\mathcal{U})$, hence is compact.*

PROOF. A type $p(x) \in S_x(\mathcal{U})$ is not invariant over A if and only if there is some formulas $\varphi(x;b_1), \neg\varphi(x;b_2) \in p(x)$ with $\mathrm{tp}(b_1/A) = \mathrm{tp}(b_2/A)$. Thereby, the open set

$$\bigcup\{[\varphi(x;b_1) \wedge \neg\varphi(x;b_2)] : \varphi(x;y) \in L, b_1, b_2 \in \mathcal{U}_y \text{ with } \mathrm{tp}(b_1/A) = \mathrm{tp}(b_2/A)\}$$

is equal to $S_x(\mathcal{U}) \setminus S_x^{\mathrm{inv}}(\mathcal{U}, A)$. Hence $S_x^{\mathrm{inv}}(\mathcal{U}, A)$ is a closed. □

EXERCISE 3.2.8. Let

$$S_x^{\mathrm{fin}}(\mathcal{U}, A) = \{p \in S_x(\mathcal{U}) : p \text{ is finitely satisfiable over } A\}.$$

Prove that $S_x^{\mathrm{fin}}(\mathcal{U}, A)$ is a closed subset of $S_x(\mathcal{U})$.

PROPOSITION 3.2.9. *If $p \in S_x(\mathcal{U})$ is invariant over some small $A \subseteq \mathcal{U}$ and $\mathcal{V} \succeq \mathcal{U}$, then there exists a unique type $q \in S_x(\mathcal{V})$ that q extends p and q is invariant over A.*

PROOF. Given any such $p(x)$, define

$$q(x) = \{\varphi(x;b) : \varphi(x;y) \in L, b \in \mathcal{V}_y, (\exists b' \in \mathcal{U}_y)(\mathrm{tp}(b/A) = \mathrm{tp}(b'/A) \wedge \varphi(x;b') \in p(x))\}.$$

Since $\mathcal{U}$ is $|A|^+$-saturated, for each $b \in \mathcal{V}$, there exists $b' \in \mathcal{U}$ with $\text{tp}(b/A) = \text{tp}(b'/A)$. Therefore, $q(x) \in S_x(\mathcal{V})$. Since p is invariant over A, clearly q must be also. It is clear by definition that this is the unique such type. □

PROPOSITION 3.2.10. *If $p \in S_x(\mathcal{U})$ is invariant over some small $A \subseteq \mathcal{U}$ and $\langle a_i : i < \omega \rangle$ is such that, for all $i < \omega$,*

$$a_i \models p|_{A \cup \{a_j : j < i\}},$$

then $\langle a_i : i < \omega \rangle$ is an indiscernible sequence over A.

PROOF. We check by induction that, for all $i_0 < ... < i_n$,

$$\text{tp}(a_{i_0}, ..., a_{i_n}/A) = \text{tp}(a_0, ..., a_n/A)$$

Suppose it holds for $i_0 < ... < i_n$ and consider $j_0 < ... < j_n$ where we change exactly one coordinate by increasing by one (say $j_t = i_t$ for all $t \neq s$ and $j_s = i_s + 1$). Then, by induction,

$$\text{tp}(a_{i_0}, ..., a_{i_{n-1}}/A) = \text{tp}(a_{j_0}, ..., a_{j_{n-1}}/A).$$

By invariance of $p(y)$, for any formula $\theta(z_0, ..., z_n)$ over A,

$$\theta(a_{i_0}, ..., a_{i_{n-1}}, y) \in p(y) \text{ iff. } \theta(a_{j_0}, ..., a_{j_{n-1}}, y) \in p(y).$$

Hence,

$$\text{tp}(a_{i_0}, ..., a_{i_n}/A) = \text{tp}(a_{j_0}, ..., a_{j_n}/A).$$

□

We call such a sequence a *Morley sequence* of p over A.

EXERCISE 3.2.11. Fix a small A and $p \in S_x(\mathcal{U})$ an A-invariant type. Show that p is definable over A if and only if, for all $\varphi(x; y)$, the set

$$D_\varphi(p) = \{q(y) \in S_y(A) : \varphi(x; b) \in p(x) \text{ for some (any) } b \models q(y)\}$$

is closed. Hint: $D_\varphi(p) \subseteq S_y(A)$ is closed if and only if it is compact.

Simon uses this fact to show that an invariant global type p is definable if and only if it commutes with every finitely satisfiable type (Lemma 5 of [**23**]). See [**23**] for the definition of "commutes." He goes on to show that every global invariant type of dp-rank 1 is either finitely satisfiable or definable (Theorem 10 of [**23**]). See Definition 4.2.9 below for the definition of dp-rank.

3. UDTFS

As in the previous section, let T be a first-order theory in a fixed language L and let $\mathcal{U}$ be a monster model of T. Recall the defintion of UDTFS, Definition 2.4.1. We restate it here in terms of types.

REMARK 3.3.1. A formula $\varphi(x; y)$ has UDTFS (of rank $\leq d$) if there exists $\psi_i(y; z_1, ..., z_d)$ for $i = 1, ..., N$ such that, for all finite $B \subseteq \mathcal{U}_y$ and all $p \in S_\varphi(B)$, there exists $c_1, ..., c_d \in B$ and i such that $\psi_i(y; c_1, ..., c_d)$ defines $p(x)$ (i.e., $\varphi(a; B) = \psi_i(B; c_1, ..., c_d)$).

EXERCISE 3.3.2. If $\varphi(x; y)$ has UDTFS-rank $\leq d$, show that the UDTFS-rank of $\neg\varphi(x; y)$ is $\leq d$. If, in addition, $\psi(x; z)$ has UDTFS-rank $\leq n$, find a bound for the UDTFS-rank of $(\varphi \wedge \psi)(x; y, z)$.

3.1. Subadditivity and Weak o-Minimality. In this subsection, we show that UDTFS-rank is subadditive. We use this to compute the UDTFS-rank for weakly o-minimal theories.

PROPOSITION 3.3.3 (Lemma 2.6 of [**11**]). *Fix $k < \omega$. If T is such that, for each formula $\varphi(x; y)$ with x in the homesort (i.e., $|x| = 1$), the UDTFS-rank of φ is $\leq k$, then a general formula $\varphi(x; y)$ has UDTFS-rank $\leq k \cdot |x|$.*

PROOF. By induction on $n = |x|$. Fix $\varphi(x; y)$ with $|x| = n$ and write $x = x_1{}^\frown x_2$ with $|x_1| = 1$ and $|x_2| = n - 1$. Repartition and let $\hat{\varphi}(x_1; x_2, y) = \varphi(x; y)$. By hypothesis, there exists finitely many $\hat{\varphi}$-definitions

$$\psi_i(x_2, y; w_1, z_1, ..., w_k, z_k) \text{ with } |w_i| = |x_2| \text{ and } |z_i| = |y|$$

for $i = 1, ..., N_1$. For each i, let

$$\hat{\psi}_i(x_2; y, z_1, ..., z_k) = \psi_i(x_2, y; x_2, z_1, ..., x_2, z_k)$$

(that is, replace all w_ℓ with x_2 and repartition). Finally, by induction, there exists finitely many $\hat{\psi}_i$-definitions

$$\gamma_{i,j}(y, z_1, ..., z_k; u_1, v_{1,1}, ..., v_{1,k}, ..., u_{k(n-1)}, v_{k(n-1),1}, ..., v_{k(n-1),k})$$

for $j = 1, ..., N_2$. For each j, let

$$\hat{\gamma}_{i,j}(y; z_1, ..., z_k, u_1, ..., u_{k(n-1)}) = \gamma_{i,j}(y, z_1, ..., z_k; u_1, z_1, ..., z_k, ...)$$

(that is, replace all $v_{\ell,t}$ with z_t). Finally, we claim that $\hat{\gamma}_{i,j}$ are a set of φ-definitions witnessing that φ has UDTFS rank $\leq nk$.

Take $a_1 \in \mathcal{U}, a_2 \in \mathcal{U}^{n-1}$, $B \subseteq \mathcal{U}_y$ finite, and consider $p(x) = \mathrm{tp}_\varphi(a_1, a_2/B)$. Consider instead $\hat{p}(x_1) = \mathrm{tp}_{\hat{\varphi}}(a_1/a_2{}^\frown B)$ (where $a_2{}^\frown B =$

$\{a_2{}^\frown b : b \in B\}$). By definition, there exists $c_1, ..., c_k \in B$ and i so that $\psi_i(x_2, y; a_2, c_1, ..., a_2, c_k)$ defines $\hat{p}$. Now, consider

$$q(x_2) = \mathrm{tp}_{\hat{\psi}_i}(a_2/B^\frown\langle c_1, ..., c_k\rangle).$$

By definition, there exists j and $d_1, ..., d_{k(n-1)}$ so that

$$\gamma_{i,j}(y, z_1, ..., z_k; d_1, c_1, ..., c_k, ..., d_{k(n-1)}, c_1, ..., c_k) \text{ defines } q(x_2).$$

It is now easy to check that

$$\hat{\gamma}_{i,j}(y; c_1, ..., c_k, d_1, ..., d_{k(n-1)}) \text{ defines } p(x).$$

□

In this section, we will use Proposition 3.3.3 to compute some UDTFS ranks for formulas in common theories. This will give us a bound on the dimension of compression schemes using Proposition 2.4.2 above.

First we focus on weakly o-minimal theories. A theory T with a total linear order $<$ is *weakly o-minimal* if, for all formulas $\varphi(x; b) \in L(\mathcal{U})$ with x in the homesort, $\varphi(\mathcal{U}; b)$ is a finite union of $<$-convex sets. By compactness, there is a uniform bound on the number of convex components of instances of a fixed formula $\varphi(x; y)$. A theory T with a total linear order $<$ is *o-minimal* if, for all formulas $\varphi(x; b) \in L(\mathcal{U})$ with x in the homesort, $\varphi(\mathcal{U}; b)$ is a finite union of points and $<$-intervals. Clearly o-minimality implies weak o-minimality, but these are not the same. For example, consider $T = \mathrm{Th}(\mathbb{Q}; <, (-\pi, \pi))$ (with a predicate for the convex set between $-\pi$ and π, which has no endpoints in $\mathbb{Q}$). This is weakly o-minimal, but clearly not o-minimal. We show the following proposition.

Proposition 3.3.4 (Theorem 1.2 of [**14**]). *If $\varphi(x; y)$ is a formula in a weakly o-minimal theory with x of the homesort, then φ has UDTFS-rank ≤ 1.*

Corollary 3.3.5 (Corollary 5.8 of [**1**]). *If T is a weakly o-minimal theory and $\varphi(x; y)$ is any formula, then φ has UDTFS-rank $\leq |x|$.*

Proof. This follows from Proposition 3.3.4 and Proposition 3.3.3 □

Remark 3.3.6. By Proposition 2.4.2, if $\varphi(x; y)$ is a formula in a weakly o-minimal theory T, the concept class $\mathcal{C}_\varphi$ has a $|x|$-dimensional compression scheme. Consider, for example $T = \mathrm{Th}(\mathbb{R}; +, \cdot, <)$, the theory of real closed fields. This is (weakly) o-minimal. So, for example, if we take

$$\varphi(x_1, x_2, x_3; y_1, y_2) = [(y_1 - x_1)^2 + (y_2 - x_2)^2 \leq x_3^2].$$

we see that the concept class of closed disks in $\mathbb{R}^2$ has a 3-dimensional compression scheme. In general for real closed fields, concept class obtained by a uniform first-order definable family has a d-dimensional compression scheme, where d is the number of parameters. Now Exercise 2.3.13 above becomes trivial.

Exercise 3.3.7. Write an explicit compression scheme of dimension 3 for the concept class of closed disks in $\mathbb{R}^2$ using the proofs of Proposition 3.3.3 and Proposition 3.3.4.

Proof of Proposition 3.3.4. Let $\varphi(x;y)$ be a formula where $|x| = 1$. By compactness, there exists $K < \omega$ such that, for all $b \in \mathcal{U}_y$, $\varphi(\mathcal{U};b)$ is a union of at most K $<$-convex sets. For $i = 1, ..., 2K+1$, let $\psi_i(x;y)$ be the ith convex component of $\pm\varphi(x;y)$. That is, $\psi_i(x;y) = \theta_i(x;y) \wedge \neg\theta_{i+1}(x;y)$ where $\theta_i(x;y)$ is given by

$$\exists x_0 ... \exists x_i \left(\bigwedge_{j<i} \big(x_j < x_{j+1} \wedge \varphi(x_j;y) \leftrightarrow \neg\varphi(x_{j+1};y)\big) \wedge x = x_i \right).$$

Finally, let $\gamma_i(y;z)$ say that $\varphi(x;y)$ contains a final segment of $\psi_i(x;z)$. That is,

$$\gamma_i(y;z) = \exists x \forall w \big[(w > x \wedge \psi_i(w;z)) \to \varphi(w;y)\big].$$

We claim that these γ_i are a set of φ-definitions witnessing that φ has UDTFS-rank ≤ 1.

Choose $a \in \mathcal{U}$, $B \subseteq \mathcal{U}_y$ finite, and consider $p(x) = \mathrm{tp}_\varphi(a/B)$. Of all the sets $\psi_i(\mathcal{U};c)$ for various $c \in B$, choose i and c so that the set

$$\{d \in \mathcal{U} :\models \psi_i(d;c) \text{ and } d \geq a\}$$

is $\subseteq$-minimal non-empty. Since each of these is $<$-convex, there exists a minimal such. Finally, we show that $\gamma_i(y;c)$ defines $p(x)$.

Choose $b \in B$. Then, $\gamma_i(b;c)$ if and only if $\psi_i(\mathcal{U};c)$ is cofinal in $\varphi(\mathcal{U};b)$. However, by minimality of choice, this happens if and only if $a \in \varphi(\mathcal{U};b)$. Therefore, if and only if $\varphi(x;b) \in p(x)$. □

Corollary 3.3.8 (Theorem 6.1 of [**1**]). *If T is weakly o-minimal, then $\varphi(x;y)$ has VC-density $\leq |x|$.*

This trick does not work for strongly minimal theories.

Exercise 3.3.9. Show that in the theory of algebrically closed fields of characteristic zero, $T = \mathrm{Th}(\mathbb{C};+,\cdot)$, the formula

$$\varphi(x;y_0,y_1) = [x^2 + y_1x + y_0 = 0]$$

has UDTFS-rank 2 (Hint: Consider $(x-\alpha_i)(x-\alpha_j)$ for α_i's algebraically independent). On the other hand, show that the VC-density of $\varphi(x; y_0, y_1)$ is actually 1.

3.2. Maximum of Dimension n. Recall that a concept class $\mathcal{C}$ on X is maximum of dimension n if, for all $Y \subseteq X$ with $n \leq |Y| < \omega$, $|C|_Y| = \Phi_n(|Y|)$. In terms of model theory, we will say that a formula $\varphi(x; y)$ is *maximum of dimension n* if, for all finite $B \subseteq \mathcal{U}_y$ with $|B| \geq n$,

$$|S_\varphi(B)| = \Phi_n(|B|).$$

In particular, such a formula has independence dimension n. The main result of this section is the following:

THEOREM 3.3.10. *Fix a formula $\varphi(x; y)$ and $n < \omega$. If φ is maximum of dimension n, then φ has UDTFS-rank $\leq n$.*

First, we exhibit some lemmas about maximum concept classes. Let $\mathcal{C}$ be a concept class on a finite set X. For any $a \in X$, define the three concept classes on $(X \setminus \{a\})$:

$$\begin{aligned}
\mathcal{C}_a &= \mathcal{C}|_{X\setminus\{a\}}, \\
\mathcal{C}_a^2 &= \{f \in \mathcal{C}_a : \text{ both extensions of } f \text{ are in } \mathcal{C}\}, \text{ and} \\
\mathcal{C}_a^1 &= \mathcal{C}_a \setminus \mathcal{C}_a^2.
\end{aligned}$$

LEMMA 3.3.11 (Theorem 9 of [**7**]). *If $\mathcal{C}$ has VC-dimension n, then $\mathcal{C}$ is maximum of dimension n if and only if $|\mathcal{C}| = \Phi_n(|X|)$.*

PROOF. If $\mathcal{C}$ is maximum, then, by definition, $|\mathcal{C}| = \Phi_n(|X|)$, so we show the other direction.

Let $m = |X|$, suppose $|\mathcal{C}| = \Phi_n(m)$, fix $Y \subseteq X$, and we show, by reverse induction on $|Y|$, that $|\mathcal{C}|_Y| = \Phi_n(|Y|)$. Fix any $a \in X$ and suppose $Y = (X \setminus \{a\})$. So $\mathcal{C}|_Y = \mathcal{C}_a$ as defined above, which has VC-dimension $\leq n$, so $|\mathcal{C}_a| \leq \Phi_n(m-1)$ by Theorem 2.1.6. On the other hand, if $Z \in \binom{Y}{n}$ is shattered by $\mathcal{C}_a^2$, then clearly $Z \cup \{a\}$ is shattered by $\mathcal{C}$, hence $\mathcal{C}$ has VC-dimension $n+1$, contrary to assumption. Therefore, $\mathcal{C}_a^2$ has VC-dimension $n-1$. Again, by Theorem 2.1.6, $|\mathcal{C}_a^2| \leq \Phi_{n-1}(m-1)$.

On the other hand, for each $f \in \mathcal{C}_a$, either $f \in \mathcal{C}_a^1$ or $f \in \mathcal{C}_a^2$. In the first case, there is exactly one extension to a function in $\mathcal{C}$ and, in the second case, there are exactly two extensions to a function in $\mathcal{C}$. Therefore,

$$|\mathcal{C}| = |\mathcal{C}_a^1| + 2 \cdot |\mathcal{C}_a^2| = |\mathcal{C}_a| + |\mathcal{C}_a^2|.$$

Hence,

$$\Phi_n(m) = |\mathcal{C}_a| + |\mathcal{C}_a^2| \leq \Phi_n(m-1) + \Phi_{n-1}(m-1) = \Phi_n(m).$$

Therefore, $|\mathcal{C}|_Y| = \Phi_n(m-1)$, as desired. □

LEMMA 3.3.12 (Corollary 1 of [**7**]). *If $\mathcal{C}$ is maximum of dimension n, then, for any $a \in X$, $\mathcal{C}_a^2$ is a concept class that is maximum of dimension $n-1$.*

PROOF. By Lemma 3.3.11, it suffices to show $|\mathcal{C}_a^2| = \Phi_{n-1}(|X|-1)$ and $\mathcal{C}_a^2$ has VC-dimension $n-1$. Both of these facts follow from the proof of Lemma 3.3.11. □

Finally, by induction on Lemma 3.3.12, if $\mathcal{C}$ is maximal of dimension n and $Y \in \binom{X}{n}$, then the concept class

$$\mathcal{C}_Y^2 = \{f \in \mathcal{C}|_{X\setminus Y} : (\forall g \in {}^Y 2)(f \cup g \in \mathcal{C})\} \tag{3.1}$$

is maximum of VC-dimension 0. That is, $|\mathcal{C}_Y^2| = 1$.

THEOREM 3.3.13 (Theorem 10 of [**7**]). *Suppose that $\mathcal{C}$ is a concept class on a finite set X that is maximum of dimension n. For any $f \in \mathcal{C}$, there exists $Y \in \binom{X}{n}$ so that $f|_{X\setminus Y}$ is the unique element of $\mathcal{C}_Y^2$ (as in* (3.1)*).*

PROOF. We show this by induction on $m = |X|$ and n. If $m \leq n$, this is trivial. If $n = 0$, then $\mathcal{C}$ has only one element.

Fix X with $|X| > n$ and fix $a \in X$. Fix $f \in \mathcal{C}$.

Case 1. $f|_{X\setminus\{a\}} \in \mathcal{C}_a^2$.

By Lemma 3.3.12, $\mathcal{C}_a^2$ is maximum of dimension $n-1$. By induction, there exists $Y \in \binom{X\setminus\{a\}}{n-1}$ such that

$$(\mathcal{C}_a^2)_Y^2 = \{f|_{X\setminus(\{a\}\cup Y)}\}.$$

However, it is easy to check that $(\mathcal{C}_a^2)_Y^2 = \mathcal{C}_{Y\cup\{a\}}^2$. This completes Case 1.

Case 2. $f|_{X\setminus\{a\}} \in \mathcal{C}_a^1$.

By induction, there exists $Y \in \binom{X\setminus\{a\}}{n}$ such that

$$(\mathcal{C}_a)_Y^2 = \{f|_{X\setminus(\{a\}\cup Y)}\}.$$

Suppose $\{f'\} = \mathcal{C}_Y^2$. In particular, $f'' = f' \cup f|_Y \in \mathcal{C}$. First, suppose that $f'(b) \neq f(b)$ for some $b \in X \setminus (\{a\}\cup Y)$. Then, $\{b\}\cup Y$ is shattered by $\mathcal{C}$, showing that $\mathcal{C}$ has VC-dimension $\geq n+1$. Contradiction. So $f|_{X\setminus\{a\}} = f''|_{X\setminus\{a\}}$. Secondly, suppose $f'(a) \neq f(a)$. Then, $f|_{X\setminus\{a\}}$ has two extensions to $\mathcal{C}$, namely f and f''. Contrary to the assumption of Case 2. Hence, $f' = f|_Y$, as desired. □

PROOF OF THEOREM 3.3.10. First, define

$$\theta(y; z_0, ..., z_{n-1}) = \bigwedge_{t \in {}^n 2} \exists x \left(\bigwedge_{i<n} \varphi(x; z_i)^{t(i)} \wedge \varphi(x; y) \right).$$

and then, for each $s \in {}^n 2$, let

$$\psi_s(y; z_0, ..., z_{n-1}) = \left[\bigwedge_{i<n} y \neq z_i \to \theta(y; z) \right] \wedge \left[\bigvee_{i<n, s(i)=1} y = z_i \right].$$

Finally, we show that $\{\psi_s : s \in {}^n 2\}$ witness the fact that φ has UDTFS-rank $\leq n$.

Suppose $B \subseteq \mathcal{U}_y$ finite and $p \in S_\varphi(B)$. By Theorem 3.3.13 applied to the concept class

$$\mathcal{C} = \{\varphi(a; B) : a \in \mathcal{U}_x\},$$

there exists $B_0 \in \binom{B}{n}$ such that $p^* = p|_{B \setminus B_0}$ is the unique type in $S_\varphi(B \setminus B_0)$ such that

$$p^*(x) \cup \{\varphi(x; c_i)^{t(i)} : i < n\}$$

is consistent for each $t \in {}^n 2$, where $B_0 = \{c_i : i < n\}$ is an enumeration. Therefore, for all $b \in (B \setminus B_0)$, $\varphi(x; b) \in p(x)$ if and only if $\models \theta(b; c)$. Hence, for $s \in {}^n 2$ such that $\varphi(x; c_i)^{s(i)} \in p(x)$, we have that $\psi_s(y; c)$ defines $p(x)$. □

3.3. VC-density < 2**.** For this subsection, fix a formula $\varphi(x; y)$. We aim to prove the following theorem.

THEOREM 3.3.14 (Theorem 3.14 of [**11**]). *If $\varphi(x; y)$ has VC-density < 2, then φ has UDTFS.*

Fix $B \subseteq \mathcal{U}_y$ finite and $p \in S_\varphi(B)$. For any partial type $q(x)$ and any $b \in B$, we will say that q *decides* $\varphi(x; b)$ if

$$q(x) \vdash \varphi(x; b) \text{ or } q(x) \vdash \neg\varphi(x; b).$$

We will say that q *decides* $\varphi(x; b)$ *correctly* if $q(x) \vdash p|_{\{b\}}(x)$ (i.e., q implies the correct choice of $\pm\varphi(x; b)$ according to $p(x)$).

LEMMA 3.3.15. *If $q(x) \subseteq p(x)$ and $q(x)$ decides $\varphi(x; b)$ for some $b \in B$, then it does so correctly.*

PROOF. Suppose $q \vdash \varphi(x; b)^t$ for some t. On the other hand, if $\varphi(x; b)^{1-t} \in p(x)$, then $q \cup \{\varphi(x; b)^{1-t}\}$ would be consistent, a contradiction. □

For the purposes of this subsection, for $B_0 \subseteq B_1 \subseteq B$, define

$$\begin{aligned} p_{B_1, B_0}(x) = &\{\varphi(x; b)^t : b \in (B_1 \setminus B_0), t < 2 \text{ so that } \varphi(x; b)^t \in p(x)\} \cup \\ &\{\neg\varphi(x; b)^t : b \in B_0, t < 2 \text{ so that } \varphi(x; b)^t \in p(x)\}. \end{aligned}$$

That is, $p_{B_1,B_0}(x)$ is simply $p|_{B_1}$ except we switch the truth value of φ on each $b \in B_0$. Notice that this may not be consistent, hence may not actually be a φ-type.

For $B_0 \subseteq B$ and $b \in B$, we will say that B_0 *1-decides* $\varphi(x;b)$ if $p|_{B_0}$ decides $\varphi(x;b)$ or there exists $b_0 \in B_0$ so that $p_{B_0,\{b_0\}}$ is consistent and decides $\varphi(x;b)$. That is, by peturbing at most one truth value in $p|_{B_0}$, we get something that decides $\varphi(x;B)$. We say that B_0 *1-decides $\varphi(x;b)$ correctly* if B_0 1-decides $\varphi(x;b)$ and, either

1. $p|_{B_0}$ decides $\varphi(x;b)$, or
2. for all $b_0 \in B$ such that $p_{B_0,\{b_0\}}$ is consistent and decides $\varphi(x;b)$, $p_{B_0,\{b_0\}}$ decides $\varphi(x;b)$ correctly.

By Lemma 3.3.15, if $p|_{B_0}$ decides $\varphi(x;b)$, then it does so correctly.

Remark 3.3.16. It is possible to generalize this to n-decides, but we will not do it here. Unfortunately, some of the argument breaks down for higher decisions. In particular, this construction does not seem to work for φ with VC-density ≥ 2.

For notational simplicity, we define a quasi-order on $\mathcal{P}(B)$. For $B_0, B_1 \subseteq B$, define

$$B_0 \trianglelefteq_p B_1 \text{ iff. } p|_{B_0}(x) \vdash p|_{B_1}(x).$$

We will write $B_0 \equiv_p B_1$ if $B_0 \trianglelefteq_p B_1$ and $B_1 \trianglelefteq_p B_0$. Write $B_0 \triangleleft_p B_1$ if $B_0 \trianglelefteq_p B_1$ but $B_1 \not\trianglelefteq_p B_0$.

Lemma 3.3.17. *For the quasi-ordering $\trianglelefteq_p$, the following hold:*

1. *For all $B_0, B_1 \in \mathcal{P}(B)$, $B_0 \subseteq B_1$ implies that $B_1 \trianglelefteq_p B_0$.*
2. *For all $B_0, B_1, B_2 \in \mathcal{P}(B)$ with $B_0 \subseteq B_1$, $B_1 \trianglelefteq_p B_2$ if and only if $B_1 \trianglelefteq_p B_0 \cup B_2$.*
3. *If $\mathcal{D} \subseteq \mathcal{P}(B)$ and $B_1 \in \mathcal{D}$, then there exists $B_0 \in \mathcal{D}$ with $B_0 \trianglelefteq_p B_1$ and, for all other $B_2 \in \mathcal{D}$, $B_2 \trianglelefteq_p B_0$ implies that $B_2 \equiv_p B_0$ (we call such elements B_0 $\trianglelefteq_p$-minimal elements of $\mathcal{D}$).*

The proof of this lemma is simple and we leave it as an exercise to the reader.

Exercise 3.3.18. Prove Lemma 3.3.17. Notice that (3) holds because B, hence $\mathcal{D}$, is finite.

Lemma 3.3.19 (Correct Decisions Lemma). *Fix $b \in B$, $\mathcal{D} \subseteq \mathcal{P}(B)$ non-empty, $B_0 \in \mathcal{D}$ $\trianglelefteq_p$-minimal, and $b_0 \in B_0$. If*

1. *$p_{B_0,\{b_0\}}$ is consistent and decides $\varphi(x;b)$,*
2. *$p|_{B_0}$ does not decide $\varphi(x;b)$, and*

(3) *there exists* $B_1 \trianglelefteq_p (B_0 \setminus \{b_0\})$ *such that* $(B_1 \cup \{b\}) \in \mathcal{D}$,

then $p_{B_0,\{b_0\}}$ *decides* $\varphi(x;b)$ *correctly.*

PROOF. Since $p|_{B_0}$ does not decide $\varphi(x;b)$, we have that $B_0 \ntrianglelefteq_p \{b\}$ (by definition of $\trianglelefteq_p$), hence, by Lemma 3.3.17 (2),

$$B_0 \not\equiv_p (B_0 \setminus \{b_0\}) \cup \{b\} \tag{3.2}$$

(put $B_0 = B_0 \setminus \{b_0\}$, $B_1 = B_0$, and $B_2 = \{b\}$). By means of contradiction, suppose $p_{B_0,\{b_0\}}$ decides $\varphi(x;b)$ incorrectly. That is,

$$p|_{B_0\setminus\{b_0\}}(x) \cup \{\neg\varphi(x;b_0)^{t_0}\} \vdash \neg\varphi(x;b)^t,$$

where $t_0, t < 2$ are such that $\varphi(x;b_0)^{t_0}, \varphi(x;b)^t \in p(x)$. By contrapositive,

$$p|_{B_0\setminus\{b_0\}}(x) \cup \{\varphi(x;b)^t\} \vdash \varphi(x;b_0)^{t_0}.$$

Hence, $(B_0 \setminus \{b_0\}) \cup \{b\} \trianglelefteq_p \{b_0\}$. By Lemma 3.3.17 (2),

$$(B_0 \setminus \{b_0\}) \cup \{b\} \trianglelefteq_p B_0$$

(put $B_0 = (B_0 \setminus \{b_0\})$, $B_1 = (B_0 \setminus \{b_0\}) \cup \{b\}$, and $B_2 = \{b_0\}$). By (3.2), we get

$$(B_0 \setminus \{b_0\}) \cup \{b\} \triangleleft_p B_0.$$

However, by condition (3) above, $(B_1 \cup \{b\}) \in \mathcal{D}$ and $B_1 \trianglelefteq_p (B_0 \setminus \{b_0\})$. Hence,

$$B_1 \cup \{b\} \trianglelefteq_p (B_0 \setminus \{b_0\}) \cup \{b\} \triangleleft_p B_0.$$

This contradicts the minimality of B_0. □

So now we only need to define a good $\mathcal{D}$. we construct $\mathcal{D}_n$ a set of sequences of elements from B inductively as follows. For $n = 1$,

$$\mathcal{D}_1 = \{\langle b\rangle : \varphi(x;b)^t \text{ is consistent for both } t < 2\}.$$

For $n > 1$, define

$$\mathcal{D}_n = \{\gamma^\frown\langle b\rangle : b \in B, \gamma \in \mathcal{D}_{n-1}, \gamma \text{ does not 1-decide } \varphi(x;b)\}.$$

LEMMA 3.3.20. *Suppose* $\gamma = \langle b_0, ..., b_{n-1}\rangle \in \mathcal{D}_n$. *Then the following hold:*

(1) *For all* $\ell < n$, $p_{\gamma,\langle b_\ell\rangle}$ *is consistent.*
(2) *For all* $\ell < k < n$, $p_{\langle b_i : i \leq k\rangle,\langle b_\ell, b_l\rangle}$ *is consistent.*
(3) *For all* $k \leq n$ *and all subsequences* $\gamma_0 \subseteq \gamma$ *of length* k, $\gamma_0 \in \mathcal{D}_k$.
(4) $|S_\varphi(\{b_i : i < n\})| \geq \Phi_2(n)$.
(5) *For all* $b \in B$, *if* γ *does not* 1*-decide* $\varphi(x;b)$, *then* $\gamma^\frown\langle b\rangle \in \mathcal{D}_{n+1}$.

PROOF. (1) and (2): By induction on n. If $n = 1$, this is by definition, so suppose $n > 1$. Let $\gamma' = \langle b_i : i < n-1 \rangle$ (the initial segment of γ of length $n-1$). By induction, $p_{\gamma',\langle b_\ell \rangle}$ is consistent for each $\ell < n-1$ and $p_{\langle b_i : i \leq k \rangle, \langle b_\ell, b_l \rangle}$ is consistent for each $\ell < k < n-1$. However, by definition of $\mathcal{D}_n$, β' does not 1-decide $\varphi(x; b_{n-1})$. Therefore, $p|_\beta$ does not decide $\varphi(x; b)$, so $p_{\beta,\langle b_{n-1} \rangle}$ is consistent. Moreover, for all $\ell < n-1$, $p_{\beta',\langle b_\ell \rangle}$ is consistent, therefore, by the failure of 1-decides, it does not decide $\varphi(x; b)$. Hence, $p_{\beta,\langle b_\ell \rangle}$ and $p_{\beta,\langle b_\ell, b_{n-1} \rangle}$ are consistent. This gives the desired conclusion.

(3): Fix $k < n$ and $\gamma_0 \subseteq \gamma$ of length k and let $\gamma_0 = \langle b_{i(0)}, ..., b_{i(k-1)} \rangle$. We show by induction on k that $\gamma_0 \in \mathcal{D}_k$. For $k = 1$, for both $t < 2$, $\varphi(x; b_{i(0)})^t$ is consistent by (1), hence $\langle b_{i(0)} \rangle \in \mathcal{D}_1$. So it suffices to show that $\langle b_{i(0)}, ..., b_{i(k-2)} \rangle$ does not 1-decide $\varphi(x; b_{i(k-1)})$. However, condition (1) tells us that $p_{\gamma_0,\langle b_{i(k-1)} \rangle}$ is consistent, hence $p|_{\langle b_{i(0)},...,b_{i(k-2)} \rangle}$ does not decide $\varphi(x; b_{i(k-1)})$. Moreover, condition (2) tells us that, for all $\ell < k-1$,

$$p|_{\langle b_{i(0)},...,b_{i(k-2)} \rangle, \langle b_{i(\ell)} \rangle} \text{ does not decide } \varphi(x; b_{i(k-1)}).$$

Therefore, $\langle b_{i(0)}, ..., b_{i(k-2)} \rangle$ does not 1-decide $\varphi(x; b_{i(k-1)})$ and hence $\gamma_0 \in \mathcal{D}_k$.

(4): By (1) and (2), the following types are all in $S_\varphi(\{b_i : i < n\})$

(i) $p|_{\{b_i : i < n\}}$,
(ii) $p_{\{b_i : i < n\}, \{b_\ell\}}$ for each $\ell < n$, and
(iii) some extension of $p_{\{b_i : i \leq k\}, \{b_\ell, b_l\}}$ for each $\ell < k < n$.

This gives us a total of at least $1 + n + \binom{n}{2} = \Phi_2(n)$ types.

(5): By definition. □

Now, for the remainder of this subsection, suppose φ has VC-density < 2. Therefore, there exists $K < \omega$ such that, for all finite $C \subseteq \mathcal{U}_y$ with $|C| \geq K$,

$$S_\varphi(C) < \Phi_2(|C|).$$

In particular, for our construction of $\mathcal{D}_n$ as above, by Lemma 3.3.20 (4),

$$\mathcal{D}_K = \emptyset. \tag{3.3}$$

That is, the construction above terminates when we assume φ has VC-density < 2. Fix $K_0 < K$ maximal such that

$$\mathcal{D}_{K_0} \neq \emptyset$$

and fix $\gamma_0 \in \mathcal{D}_{K_0}$ that is $\trianglelefteq_p$-minimal (which exists by Lemma 3.3.17 (3)). We define a set $H = H(\gamma_0)$ of elements in $\bigcup\{\mathcal{D}_n : n \leq K_0\}$

inductively as follows. If $\gamma \in \mathcal{D}_1$, then set $H(\gamma) = \{\gamma\}$. If $\gamma = \langle b_i : i < n\rangle \in \mathcal{D}_n$, then for each $i < n$, let

$$\gamma_i = \langle b_0, ...b_{i-1}, b_{i+1}, ..., b_{n-1}\rangle$$

(the same as γ with the ith element removed). By Lemma 3.3.20 (3), this is an element of $\mathcal{D}_{n-1}$. By Lemma 3.3.17 (3), there exists some $\gamma_i' \in \mathcal{D}_{n-1}$ $\trianglelefteq_p$-minimal such that $\gamma_i' \trianglelefteq_p \gamma_i$. Finally, let

$$H(\gamma) = \{\gamma\} \cup \bigcup \{H(\gamma_i') : i < n\}.$$

Notice that, by construction, each element of $H = H(\gamma_0)$ is $\trianglelefteq_p$-minimal in the $\mathcal{D}_n$ to which it belongs.

Lemma 3.3.21. *Fix $b \in B$. Then,*

(1) *There exists $\gamma \in H$ such that 1-decides $\varphi(x; b)$.*
(2) *Either $\emptyset$ decides $\varphi(x; b)$ or, if $\gamma \in H$ 1-decides $\varphi(x; b)$ and $n = |\gamma|$ is minimal such, then γ 1-decides $\varphi(x; b)$ correctly.*

Proof. (1): If γ_0 does not 1-decide b, then $\gamma^\frown\langle b\rangle \in \mathcal{D}_{K_0+1} = \emptyset$ by Lemma 3.3.20 (5), contradiction.

(2): By induction on n. Suppose $n \geq 1$ and $\gamma = \langle b_0, ..., b_{n-1}\rangle$ 1-decides $\varphi(x; b)$ with n minimal such. By Lemma 3.3.15, if $p|_\gamma$ decides $\varphi(x; b)$, then it does so correctly, so assume not. Suppose that, for some $\ell < n$, $p_{\gamma,\langle b_\ell\rangle}$ is consistent and decides $\varphi(x; b)$. If $n = 1$ and, for both $t < 2$, $\varphi(x; b)^t$ is consistent, then $\langle b\rangle \in \mathcal{D}_1$. If $n > 1$, consider $\gamma_\ell' \trianglelefteq_p \gamma_\ell$ as used in the definition of $H(\gamma)$ above. We have that $\gamma_\ell' \in H$ and $|\gamma_\ell'| = n - 1 < n$. So, by minimality of n, γ_ℓ' does not 1-decide $\varphi(x; b)$. Hence,

$$\gamma_\ell'^\frown\langle b\rangle \in \mathcal{D}_n$$

by Lemma 3.3.20 (5).

These are the conditions required in Lemma 3.3.19. Hence, $p_{\gamma,\langle b_\ell\rangle}$ decides $\varphi(x; b)$ correctly. Therefore, γ 1-decides $\varphi(x; y)$ correctly. ☐

We will now conclude the proof of the main theorem.

Proof of Theorem 3.3.14. We sketch a definition for φ-types. First, notice that K given above is a uniform bound for all finite sets $B \subseteq \mathcal{U}_y$. Moreover, $|H(\gamma_0)|$ is a function of $K_0 \leq K$, hence is also unformly bounded. Given a finite $B \subseteq \mathcal{U}_x$ and $p \in S_\varphi(B)$, construct $K_0 < K$ and H as above. Then, for any $b \in B$, choose $n \leq K_0$ minimal so that there is $\gamma \in (H \cap \mathcal{D}_n)$ that 1-decides $\varphi(x; b)$. Then, it does so correctly. This is all expressible uniformly as a set of first-order formulas over H. This shows UDTFS.

Formally, let $H = \{\langle b_{i,j} : j < m_i\rangle : i < K_0\}$, arranged in order of smallest to largest (so $m_i \leq m_{i+1}$ for all i), and choose $s \in {}^{K_0\times K_0}2$ so that

$$\varphi(x; b_{i,j})^{s(i,j)} \in p(x).$$

Then, we can express "$\langle b_{i,j} : j < m_i\rangle$ 1-decides $\varphi(x; y)$" as:

$$\bigvee_{t<2} \forall x \left(\bigwedge_{j<m_i} \varphi(x; b_{i,j})^{s(i,j)} \to \varphi(x; y)^t \right) \vee$$

$$\bigvee_{\ell<n_i} \Bigg[\exists x \left(\bigwedge_{j<m_i, j\neq\ell} \varphi(x; b_{i,j})^{s(i,j)} \wedge \neg\varphi(x; b_{i,\ell})^{s(i,\ell)} \right) \wedge$$

$$\forall x \left(\bigwedge_{j<m_i, j\neq\ell} \varphi(x; b_{i,j})^{s(i,j)} \wedge \neg\varphi(x; b_{i,\ell})^{s(i,\ell)} \to \varphi(x; y) \right) \Bigg].$$

We leave it to the reader to check that all other statements are first-order uniformly definable over H. $\square$

Exercise 3.3.22. Assume x is of the homesort. Show that if, for arbitrarily large $B \subseteq \mathcal{U}_y$ and arbitarily large n, if $\mathcal{D}_n \neq \emptyset$, then T is not dp-minimal (that is, $x = x$ has dp-rank ≥ 2, as in Definition 4.2.9). To do this, use Proposition 4.2.10, Ramsey's Theorem, and compactness to get two mutually indiscernible sequences that show the dp-rank is ≥ 2. Conclude that the proof of Theorem 3.3.14 shows that all dp-minimal theories have UDTFS. This is the original proof given in [**11**] (see Theorem 3.1 of [**11**]).

We will show the result of Chernikov and Simon in Theorem 3.5.2 below that a formula in any NIP theory has UDTFS.

4. Honest Definitions

In this section, fix L a language, M an L-structure, and $B \subseteq M_y$ for some variable y.

4.1. The Honest Definition. Define a new language $L_P = L \cup \{P(y)\}$ where $P(y)$ is a new predicate symbol. The pair $\langle M, B\rangle$ is the L_P-structure whose L-reduct is M and where $P(M) = B$.

Fix an L-formula $\varphi(x; y)$, $a \in M_x$, and $B \subseteq M_y$. Recall that a formula $\psi(y)$ defines the type $\mathrm{tp}_\varphi(a/B)$ if $\psi(B) = \varphi(a; B)$. This is expressible in L_P as

$$\langle M, B\rangle \models (\forall y)[P(y) \to (\psi(y) \leftrightarrow \varphi(a; y))].$$

Let $\varphi(x; y)$ be an L-formula, M an L-structure, $B \subseteq M_y$, and $a \in M_x$. We say that an L-formula $\psi(y; z)$ is an *honest definition* of the φ-type

$\mathrm{tp}_\varphi(a/B)$ if there exists an elementary extension $\langle M', B'\rangle \succeq \langle M, B\rangle$ (in the language L_P) and d a tuple from B' such that

$$\varphi(a;B) \subseteq \psi(B';d) \subseteq \varphi(a;B').$$

Notice that, in particular, $\varphi(a;B) = \psi(B;d)$. So $\psi(y;d)$ defines the φ-type $\mathrm{tp}_\varphi(a/B)$.

PROPOSITION 3.4.1. *For any φ-type $\mathrm{tp}_\varphi(a/B)$, $\psi(y;z)$ is an honest definition if and only if, for all finite $B_0 \subseteq \varphi(a;B)$, there is a tuple d from B such that*

$$B_0 \subseteq \psi(B;d) \subseteq \varphi(B;a).$$

PROOF. Suppose $\psi(y;z)$ is an honest definition of $\mathrm{tp}_\varphi(a/B)$ (where $z = \langle z_0, ..., z_{n-1}\rangle$ and each z_i is of the same sort as y) and fix $B_0 \subseteq \varphi(a;B)$ finite. Consider the L_P-formula $\theta_{B_0}(z)$ given by

$$\left[\bigwedge_{i<n} P(z_i) \wedge \bigwedge_{b\in B_0} \psi(b;z) \wedge (\forall y)(P(y) \to [\psi(y;z) \to \varphi(a;y)])\right].$$

Clearly $\langle M', B'\rangle \models \exists z\theta_{B_0}(z)$, hence, by elementarity, we get $\langle M, B\rangle \models \exists z\theta_{B_0}(z)$. Therefore, there exists some $d \in B^n$ so that

$$B_0 \subseteq \psi(B;d) \subseteq \varphi(a;B).$$

Conversely, suppose that, for all finite $B_0 \subseteq \varphi(a;B)$, there exists $d \in B^n$ such that

$$B_0 \subseteq \psi(B;d) \subseteq \varphi(a;B)$$

Consider the set of formulas

$$\Sigma(z) = \{\theta_{B_0}(z) : B_0 \subseteq \varphi(a;B) \text{ finite }\}.$$

By compactness, this is satisfied in some $\langle M', B'\rangle \succeq \langle M, B\rangle$, say by $d \in (B')^n$. Hence,

$$\varphi(a;B) \subseteq \psi(B';d) \subseteq \varphi(a;B').$$

$\square$

EXERCISE 3.4.2. Let $\varphi(x;y) = x < y$ in the theory of dense linear orders. Show that the φ-type of π over $\mathbb{Q}$ has an honest definition, namely $y \leq z$.

The main result of this section is the following theorem:

THEOREM 3.4.3 (Proposition 1.1 of [**5**]). *Let L be a language, M an L-structure, $\varphi(x;y)$ an L-formula with NIP, $a \in M_x$, and $B \subseteq M_y$. Then, $\mathrm{tp}_\varphi(a/B)$ has an honest definition.*

Proof. Let $\langle M', B'\rangle \succeq \langle M, B\rangle$ be a $|M|^+$-saturated elementary extension and let $\mathcal{U}$ be a monster model of $T = \mathrm{Th}(M)$, so $M \prec M' \prec \mathcal{U}$. Consider $S_y^{\mathrm{fin}}(\mathcal{U}, B)$ from Exercise 3.2.8, the set of all global types in y finitely satisfiable in B. By the exercise, this is a closed subset of $S_y(\mathcal{U})$, which is itself a compact Hausdorff space. Hence, $S_y^{\mathrm{fin}}(\mathcal{U}, B)$ is compact. Consider $p(y) \in S_y^{\mathrm{fin}}(\mathcal{U}, B)$.

We construct a sequence $\langle b_i : i < \omega\rangle$ such that, for all $i < \omega$,

(1) $b_i \in B'$,
(2) $b_i \models p|_{B\cup\{b_j:j<i\}}$, and
(3) $\models \neg(\varphi(a; b_i) \leftrightarrow \varphi(a; b_{i+1}))$.

By Proposition 3.2.10, $\langle b_i : i < \omega\rangle$ is indiscernible over B. However, condition (3) implies then that φ has infinite alternation. By Proposition 3.1.3, φ has the independence property, contrary to assumption. Therefore, this construction must terminate after finitely many steps, say n. That is $\langle b_i : i < n\rangle$ exists satisfying (1), (2), and (3), but no such b_n exists.

Set $t_p < 2$ such that $\varphi(a; y)^{t_p} \in p(y)$. We claim that $\models \varphi(a; b_{n-1})^{t_p}$. Consider the L_P-type

$$p^*(y) = p|_{\{a\}\cup B\cup\{b_i:i<n\}}(y) \cup \{P(y)\}$$

Since $p(y)$ is finitely satisfiable over B, $p^*(y)$ is finitely satisfiable over B. Therefore, by saturation, there exists $b \in B'$ satisfying p^*. Therefore, if $\models \neg\varphi(a; b_{n-1})^{t_p}$, setting $b_n = b$ would extend our sequence, contrary to assumption.

Similarly, we must have that, in $\langle M', B'\rangle$,

$$p|_{B\cup\{b_i:i<n\}}(y) \cup \{P(y)\} \cup \{\neg\varphi(a; y)^{t_p}\}$$

is inconsistent (otherwise, we could again extend our sequence). In other words,

$$p|_{B\cup\{b_i:i<n\}}(y) \cup \{P(y)\} \vdash \varphi(a; y)^{t_p}.$$

By compactness, there exists $\theta_p(y) \in p|_{B\cup\{b_i:i<n\}}(y)$ so that

$$\langle M', B'\rangle \models (\forall y)(P(y) \wedge \theta_p(y) \to \varphi(a; y)^{t_p}). \tag{3.4}$$

Since $\theta_p(y) \in p|_{B\cup\{b_i:i<n\}}(y) \subseteq p(y)$, we also have that

$$p(y) \vdash \theta_p(y).$$

Therefore, the set $\{[\theta_p(y)] : p \in S_y^{\mathrm{fin}}(\mathcal{U}, B)\}$ is an open cover of the space $S_y^{\mathrm{fin}}(\mathcal{U}, B)$. Hence, since $S_y^{\mathrm{fin}}(\mathcal{U}, B)$ is compact, there exists a finite subcover. That is, there exists a finite $S_0 \subseteq S_y^{\mathrm{fin}}(\mathcal{U}, B)$ so that

$$\{[\theta_p(y)] : p \in S_0\} \text{ covers } S_y^{\mathrm{fin}}(\mathcal{U}, B).$$

Set

$$\psi'(x) = \bigvee \{\theta_p(x) : p \in S_0, t_p = 1\}$$

and let $\psi'(y) = \psi(y; d)$ for $\psi(y, z) \in L(B)$ and $d \in B'$ (d is the concatination of the b_i's needed in the various θ_p's for $p \in S_0$). We claim that this works.

First, if $b \in \varphi(a; B)$, then $p(y) = \mathrm{tp}(b/\mathcal{U}) \in S_y^{\mathrm{fin}}(\mathcal{U}, B)$. By definition, there exists $p' \in S_0$ so that $p(y) \vdash \theta_{p'}(y)$, hence $\models \theta_{p'}(b)$. However, if $t_{p'} = 0$, then (3.4) would give $\models \neg\varphi(a; b)$, contrary to assumption. Therefore, $t_{p'} = 1$ and $\models \psi(b; d)$. This shows

$$\varphi(a; B) \subseteq \psi(B'; d).$$

Moreover, (3.4) implies that

$$\langle M', B' \rangle \models (\forall y)(P(y) \wedge \psi(y; d) \to \varphi(a; y)).$$

Therefore, $\psi(B', d) \subseteq \varphi(a; B')$. This concludes the proof. □

Exercise 3.4.4. Prove the original statement of Proposition 1.1 of [**5**]: Let $L \subseteq L'$ be languages, $\mathcal{U}$ a monster model for an L'-theory T', $p(x)$ a partial L'-type, $\varphi(x; y)$ an NIP L-formula, and $a \in \mathcal{U}_x$. For each $B \subseteq p(\mathcal{U})$, there exists $\psi(y; z)$ and L-formula and d a tuple from $p(\mathcal{U})$ such that

$$\varphi(a; B) \subseteq \psi(p(\mathcal{U}); d) \subseteq \varphi(a; p(\mathcal{U})).$$

Hint: Modify the proof of Theorem 3.4.3 above, replacing L_P with L' and $P(y)$ with $p(y)$.

Corollary 3.4.5 (Corollary 2.7 of [**10**]). *Let T be an L-theory, $M \models T$, $B \subseteq M$, $\varphi(x; y)$ an NIP L-formula, and $a \in M_x$. Then, there exists an elementary extension $\langle M', B' \rangle \succ \langle M, B \rangle$, an L-formula $\psi(y; z)$, and a tuple d from B' such that*

$$\varphi(a; B) = \psi(B; d).$$

Proof. Follows immedately from Theorem 3.4.3, noting that if $b \in \psi(B; d)$, then $b \in \psi(B'; d) \subseteq \varphi(a; B')$, hence $b \in \varphi(a; B)$. □

4.2. The Shelah Expansion. Let T be a theory with NIP, $M \models T$, and $\mathcal{U}$ a monster model for T (in particular, $M \prec \mathcal{U}$). A subset $D \subseteq M$ is *externally definable* if there exists an L-formula $\varphi(x; y)$ and $a \in \mathcal{U}_x$ such that $D = \varphi(a; M)$. The *Shelah Expansion* of M, denoted M^{sh}, is an expansion of M to the language L^{sh} where we add a predicate $P_D(y)$ for each externally definable subset of M (for various sorts) and interpret P_D as D. In particular, this expands the "Morleyization" of M.

LEMMA 3.4.6. *Let $D \subseteq M$ be an externally definable subset. Then, there is a L-formula $\varphi(x;y)$ and $a \in \mathcal{U}_x$ such that*

(1) *$D = \varphi(a;M)$ (so $\varphi(a;y)$ externally defines D), and*
(2) *For every L-formula $\theta(y;z)$ and $c \in M_z$ such that $D \subseteq \theta(M;c)$, we have*
$$\mathcal{U} \models (\forall y)[\varphi(a;y) \to \theta(y;c)].$$

PROOF. Let $M \prec N$ with N $|M|^+$-saturated. Therefore, there exists an L-formula $\varphi_0(x_0;y)$ and $a_0 \in N_{x_0}$ such that $D = \varphi_0(a_0;M)$. By Theorem 3.4.3, the type $\mathrm{tp}_{\varphi_0}(a_0/M)$ has an honest definition, say $\varphi(x;y)$. That is, there exists $\langle N', M'\rangle \succ \langle N, M\rangle$ and $a \in M'_x$ such that
$$D = \varphi_0(a_0;M) \subseteq \varphi(a;M') \subseteq \varphi_0(a_0;M').$$
In particular, $D = \varphi(a;M)$, so this is also a external definition. We show it satisfies the desired property.

Fix $\theta(y;z)$ and $c \in M_z$ such that $D \subseteq \theta(M;c)$. Thereby,
$$\langle N, M\rangle \models (\forall y)(P(y) \wedge \varphi_0(a_0;y) \to \theta(y;c)).$$
Therefore, $\langle N', M'\rangle$ is also a model of this, so
$$\varphi(a;M') \subseteq \varphi_0(a_0;M') \subseteq \theta(M';c).$$
Hence,
$$M' \models (\forall y)[\varphi(a;y) \to \theta(y;c)].$$
Since $M' \prec \mathcal{U}$, this gives us the desired conclusion. □

PROPOSITION 3.4.7. *For any M, the structure M^{sh} admits quantifier elimination.*

PROOF. We need to show that, given $P_D(x,y)$ in L^{sh}, that the formula $(\exists x)P_D(x,y)$ is equivalent to $P_{D'}(y)$ for some other externally definable D'. So, it suffices to show that, for any L-formula $\varphi(x,y;z)$ and $c \in \mathcal{U}_z$, we have
$$\{b \in M_y : (\exists a \in M_x)(\varphi(a,b;c))\} = \{b \in M_y : (\exists a \in \mathcal{U}_x)(\varphi(a,b;c))\}.$$
We may assume that $\varphi(x,y;z)$ is of the form given in Lemma 3.4.6. One direction is clear, namely if there exists $a \in M_x$ so that $\varphi(a,b;c)$ holds, then clearly there exists $a \in \mathcal{U}_x$. Suppose the other direction fails. That is, there is $b \in M_y$ and $a \in \mathcal{U}_x$ with $\models \varphi(a,b;c)$ but, for all $a' \in M_x$, $\models \neg\varphi(a',b;c)$. Let $\theta(x,y;b) = [y \neq b]$. Then, clearly
$$\varphi(M;c) \subseteq \theta(M;b) = M \times (M \setminus \{b\}).$$
Hence, by Lemma 3.4.6,
$$\mathcal{U} \models (\forall x,y)[\varphi(x,y;c) \to \theta(x,y;b)].$$
But this contradicts the fact that $\mathcal{U} \models \varphi(a,b;c)$. □

THEOREM 3.4.8 (Theorem 1.13 of [**19**]). *The structure M^{sh} has NIP.*

PROOF. Since $\mathcal{U}$ has NIP, any formula $\varphi(x; y, c)$ has NIP, so it has finite alternation rank. For $D = \varphi(M; c)$, the formula $P_D(x; y)$ must have finite alternation rank, so this follows from Proposition 3.4.7. □

Although this was originally proved by Shelah in [**19**], the proof presented here is much simplier and is due to Simon in [**24**].

5. The UDTFS Conjecture

Let T be an NIP theory in a language L with monster model $\mathcal{U}$. Recall that an L-formula $\varphi(x; y)$ has UDTFS if there exists finitely many L-formulas $\psi_0(y; z), ..., \psi_{m-1}(y; z)$ (with $z = \langle z_1, ..., z_n \rangle$ and each z_i of the same sort as y) such that, for all finite $B \subseteq \mathcal{U}_y$ and all $a \in \mathcal{U}_x$, there exists $c \in B^n$ and $j < m$ such that $\varphi(a; B) = \psi_j(B; c)$.

In this section, we prove the UDTFS Conjecture for theories with NIP. This was conjectured by Laskowski and shown for o-minimal theories by Johnson and Laskowski in [**14**]. The UDTFS Conjecture was shown for dp-minimal theories and formulas with VC-density < 2 in [**11**] (Theorem 3.3.14 above). Finally, it was proved for theories with NIP by Chernikov and Simon in [**6**]. The UDTFS Conjecture for formulas is still open.

CONJECTURE 3.5.1 (The UDTFS Conjecture). *Suppose $\varphi(x; y)$ is a formula. Then φ has NIP if and only if φ has UDTFS.*

Notice that, by Proposition 2.4.4, this implies the Warmuth Conjecture (Conjecture 2.3.15). We now exhibit Chernikov and Simon's partial solution.

THEOREM 3.5.2 (Theorem 15 of [**6**]). *If T is a NIP theory, then all formulas $\varphi(x; y)$ have UDTFS.*

Notice that Proposition 3.4.1 and Theorem 3.4.3 bring us close to the UDTFS Conjecture. In particular, for any NIP formula $\varphi(x; y)$, any $B \subseteq \mathcal{U}_y$, and any $a \in \mathcal{U}_x$, there exists $\psi(y; z)$ so that, for any finite $B_0 \subseteq \varphi(a; B)$, there exists $c \in B^n$ so that

$$B_0 \subseteq \psi(B; c) \subseteq \varphi(a; B).$$

All we need is uniformity. First, we get a weak uniformity condition.

LEMMA 3.5.3. *Let $\varphi(x; y)$ be an L-formula. For any function n from L-formulas to ω, there are finitely many formulas $\psi_0(y; z_0)$,..., $\psi_{m-1}(y; z_{m-1})$ such that, for all $M \models T$, $B \subseteq M_y$, and $a \in M_x$, there* *exists $j < m$ such that, for any $B_0 \subseteq \varphi(a;B)$ with $|B_0| \leq n(\psi_j)$, there exists c a tuple from B such that*

$$B_0 \subseteq \psi_j(B;c) \subseteq \varphi(a;B).$$

PROOF. Let $L' = L \cup \{P(y), c_a\}$ where P is a predicate and c_a is a constant in the same sort as x. For each formula $\psi(y;z)$ with $z = \langle z_1, ..., z_k\rangle$ with each z_i the same sort as y, consider the L'-sentence

$$\sigma_\psi = (\exists y_0 ... y_{n(\psi)-1})\Big[\bigwedge_{i<n(\psi)} (P(y_i) \wedge \varphi(c_a;y_i))\wedge$$
$$(\neg\exists z_1...z_k)\Big(\bigwedge_{j<k} P(z_j) \wedge \bigwedge_{i<n(\psi)} \psi(y_i;z)\wedge$$
$$(\forall y)(P(y) \wedge \psi(y;z) \to \varphi(c_a;y))\Big)\Big].$$

That is, σ_ψ says that ψ fails the conclusion of the lemma for all finite sets B_0 with $|B_0| \leq n(\psi)$. Let T' be the L'-theory axiomatized by T together with σ_ψ for all ψ. By Proposition 3.4.1 and Theorem 3.4.3 (as noted above), T' is inconsistent. Hence, by compactness, there exists $\psi_0, ..., \psi_{m-1}$ so that

$$T \cup \{\sigma_{\psi_j} : j < m\}$$

is inconsistent. Therefore, for any $M \models T$, $B \subseteq M_y$, and $a \in M_x$, if we interpret P as B and c_a as a, we get $j < m$ so that σ_{ψ_j} fails. That is, for any $B_0 \subseteq \varphi(a;B)$ with $|B_0| \leq n(\psi_j)$, there exists c a tuple from B such that $B_0 \subseteq \psi_j(B;c) \subseteq \varphi(a;B)$. □

We get full uniformity using the (p,q)-Theorem (Theorem 2.2.18).

PROPOSITION 3.5.4. *Let $\varphi(x;y)$ be an L-formula. There exists finitely many L-formulas $\psi_0(y;z), ...\psi_{m-1}(y;z)$ such that, for any $M \models T$, $B \subseteq M_y$, $a \in M_x$, and $B_0 \subseteq \varphi(a;B)$ finite, there exists c a tuple from B and $j < m$ such that*

$$B_0 \subseteq \psi_j(B;c) \subseteq \varphi(a;B).$$

PROOF. For each L-formula $\psi(y;z)$, let $n(\psi)$ be the VC-density of ψ. By Lemma 3.5.3, there exists finitely many formulas $\theta_0(y;z)$, ..., $\theta_{m-1}(y;z)$ satisfying the conclusion of the lemma for this function, n. For each $j < m$, let K_j be given by Theorem 2.2.18 (the (p,q)-Theorem) with $p = k = n(\theta_j)$. Finally, let $K = \max\{K_j : j < m\}$ and, for each $j < m$, define

$$\psi_j(y;z_0,...,z_{K-1}) = \bigvee_{\ell<K} \theta_j(y;z_\ell).$$

We claim that these ψ_j for $j < m$ work.

Let $M \models T$, $B \subseteq M_y$, $a \in M_x$, and $B_0 \subseteq \varphi(a;B)$ be finite. Let $j < m$ be given as in Lemma 3.5.3. Define

$$C = \{c \in B^n : \theta_j(B;c) \subseteq \varphi(a;B)\}.$$

So C is the collection of tuples from B^n that satisfy the second condition of being a honest definition, $\theta_j(B;c) \subseteq \varphi(a;B)$ (what Simon calls the "honesty condition"). Consider the concept class on C

$$\mathcal{D} = \{\theta_j(b;C) : b \in B_0\}.$$

This is a finite concept class that satisfies the $(n(\theta_j), n(\theta_j))$-property. That is, every $n(\theta_j)$-element subset of $\mathcal{D}$ has non-empty intersection. This is because, for every $B_1 \subseteq B_0$ with $|B_1| \leq n(\theta_j)$, by Lemma 3.5.3, there exists $c \in B^n$ such that $B_1 \subseteq \theta_j(B;c) \subseteq \varphi(a;B)$. That is, $c \in \bigcap_{b \in B_1} \theta_j(b;C)$, showing that this has a non-empty intersection.

By Theorem 2.2.18, there exists a transversal $C_0 \subseteq C$ of $\mathcal{D}$ with $|C_0| \leq K$. Let $d = \langle c_0, ..., c_{K-1}\rangle$, where $C_0 = \{c_\ell : \ell < K\}$ enumerates C_0. Then, for all $b \in B_0$, $\models \theta_j(b;c_\ell)$ for some $\ell < K$, hence $\models \psi_j(b;d)$. Moreover, by the definition of C, $\theta_j(B;c_\ell) \subseteq \varphi(a;B)$ for all $\ell < K$, hence $\psi_j(B;d) \subseteq \varphi(a;B)$. That is,

$$B_0 \subseteq \psi_j(B;d) \subseteq \varphi(a;B).$$

□

We are now ready to prove the partial result to the UDTFS Conjecture.

Proof of Theorem 3.5.2. For any L-formula $\varphi(x;y)$, consider $\psi_0(y;z), ..., \psi_{m-1}(y;z)$ as given in Proposition 3.5.4. For any finite $B \subseteq \mathcal{U}_y$ and any $a \in \mathcal{U}_x$, let $B_0 = \varphi(a;B)$. Then, by Proposition 3.5.4, there exists $j < m$ and c a tuple from B so that

$$B_0 \subseteq \psi_j(B;c) \subseteq \varphi(a;B).$$

However, since $B_0 = \varphi(a;B)$, we get that $\varphi(a;B) = \psi_j(B;c)$. Therefore, φ has UDTFS. □

Note that this highly depends on the use of the theory T having NIP. In the proof of Proposition 3.5.4, we need that each L-formula ψ have NIP to construct the function n. This leaves open the UDTFS Conjecture for formulas. However, this does give a partial result to the Warmuth Conjecture.

Proof of Theorem 2.4.5. Suppose $\mathcal{C}$ is a concept class on a set X such that $M_{\mathcal{C}}$ has NIP. Then, by Theorem 3.5.2, $R(x,y)$ has UDTFS. By Proposition 2.4.4, $\mathcal{C}$ has a compression scheme. □

Notice that the proof of Theorem 3.5.2 does not give us a bound on the UDTFS-rank (i.e., the dimension of the compression scheme).

CHAPTER 4

Ranks in NIP Theories

In this chapter, let T be a complete theory in a fixed language L and let $\mathcal{U}$ be a monster model for T. It is useful to define different measures of the complexity of formulas and theories. We have already met interesting ranks for formulas with NIP, including VC-density, VC-dimension, independence dimension, and UDTFS-rank. Here we develop some more ranks. In the first section, we will discuss Shelah 2-Rank. In the next section, we will discuss dp-rank, both local and global versions. After that, we discuss VC-density in more detail, introduce op-dimension, and have a deeper discussion of UDTFS-rank.

1. Shelah 2-Rank and Stability

In this section, we study Shelah 2-Rank.

Let $\varphi(x; y)$ be a formula, $B \subseteq \mathcal{U}_y$, and $p \in S_\varphi(B)$. The *Shelah 2-Rank* of p, denoted $R_{2,\varphi}(p)$, is an ordinal-valued function of φ-types defined inductively as follows:

(1) $R_{2,\varphi}(p) \geq 0$.
(2) $R_{2,\varphi}(p) \geq \delta$ for a limit ordinal δ if $R_{2,\varphi}(p) \geq \alpha$ for all $\alpha < \delta$.
(3) $R_{2,\varphi}(p) \geq \alpha + 1$ if, for all finite φ-types $q \subseteq p$, there exists $b \in \mathcal{U}_y$ such that

$$R_{2,\varphi}(q \cup \{\varphi(x; b)^t\}) \geq \alpha$$

for both choices of $t = 0, 1$.

This was originally denoted by $R^{|x|}(p, \varphi, 2)$ in Shelah's book. Notice that if $p(x)$ is a finite φ-type with finite Shelah 2-rank and $b \in \mathcal{U}_y$ is not in the domain of p, then either $p(x) \cup \{\varphi(x; b)\}$ or $p(x) \cup \{\neg\varphi(x; b)\}$ has Shelah 2-Rank strictly smaller than p. This is because, if both have the same rank, then condition (3) says that p has a larger rank.

Definition 4.1.1. We say that $\varphi(x; y)$ is *stable* if $R_{2,\varphi}(\emptyset) < \omega$. A theory T is *stable* if all partitioned formulas $\varphi(x; y)$ are stable.

Exercise 4.1.2. Consider $T = \mathrm{Th}(\mathbb{Q}; <)$ and show that the formula $x < y$ has infinite Shelah 2-Rank (i.e., $R_{2,x<y}(\emptyset) \geq \alpha$ for all ordinals α).

The main result of this section is to bound the UDTFS-rank (hence dimension of a compression scheme) for stable formulas $\varphi(x;y)$ by the Shelah 2-Rank of φ.

THEOREM 4.1.3 (Laskowski (unpublished)). *If $\varphi(x;y)$ is a stable formula, then the UDTFS-rank of φ is bounded above by $R_{2,\varphi}(\emptyset)$.*

PROOF. Since φ is stable, $R_{2,\varphi}(\emptyset)$ is finite, say $n = R_{2,\varphi}(\emptyset)$. For each $s \in {}^n2$, define

$$\theta_s(x; z_0, ..., z_{n-1}) = \bigwedge_{i<n} \varphi(x; z_i)^{s(i)}.$$

We will use θ_s to code a φ-type of size $\leq n$. For each $K < \omega$, define

$$\psi_{K,s}(y, z) = \exists \left(w_\nu : \nu \in {}^{\leq K}2\right) \bigwedge_{\eta \in {}^K 2} \exists x \left(\theta_s(x; z) \wedge \varphi(x; y) \wedge \bigwedge_{i<K} \varphi(x; w_{\eta|_i})^{\eta(i)} \right).$$

So $\psi_{K,s}$ codes the φ-type of θ_s together with a positive witness to the fact that φ has Shelah 2-Rank $\geq K$. We claim that the set

$$\{\psi_{K,s}(y; z) : K \leq n, s \in {}^n2\}$$

is a witness to the fact that φ has UDTFS-rank $\leq n$.

Fix $B \subseteq \mathcal{U}_y$ finite and $p \in S_\varphi(B)$. Say that $R_{2,\varphi}(p) = \ell \leq n$ and let $q_0 = \emptyset$. Recursively define $q_i \subseteq p$ as follows: Fix $i > 0$. If there exists $b \in B$ so that $q_{i-1} \cup p|_{\{b\}}$ has Shelah 2-Rank $< R_{2,\varphi}(q_{i-1})$, let $q_i = q_{i-1} \cup p|_{\{b\}}$ for any such $b \in B$. Otherwise, the construction halts and we set $q = q_i$ and $K = R_{2,\varphi}(q)$. Since $R_{2,\varphi}(q_0) = n$ and $R_{2,\varphi}(p) = \ell$, this construction halts in at most $n - \ell \leq n$ steps and $K \leq n$.

By construction, for each $b \in B$,

$$R_{2,\varphi}(q \cup p|_{\{b\}}) = R_{2,\varphi}(q) = K.$$

Therefore, $\varphi(x; b) \in p(x)$ if and only if $q(x) \cup \{\varphi(x; b)\}$ has Shelah 2-Rank $\geq K$ (equivalently, $= K$). Therefore, if we let

$$q(x) = \{\varphi(x; c_i)^{s(i)} : i < n\}$$

(allowing some repeats if $|q| < n$), then we see that $\psi_{K,s}(y, c)$ defines p, as desired. $\square$

REMARK 4.1.4. Fix $n < \omega$, X a set, and $\mathcal{C}$ a concept class on a set X. Suppose that R is stable with Shelah 2-Rank $\leq n$ in the structure $M_{\mathcal{C}}$ (as defined in (2.5) above). Then, by Theorem 4.1.3, $\mathcal{C}$ has a compression scheme of dimension n. In other words, this proves Theorem 2.4.6 (2).

EXERCISE 4.1.5. Let $T = \mathrm{Th}(\mathbb{C}; +, \cdot)$, which is a stable theory. For $a, b \in \mathcal{C}$, consider the algebraic curve

$$C_{a,b} = \{\langle x, y\rangle \in \mathbb{C}^2 : y^2 = x^3 + ax + b\}$$

(this is called an elliptic curve). Using Theorem 4.1.3, show that the concept class of elliptic curves

$$\mathcal{C} = \{C_{a,b} : a, b \in \mathbb{C} \text{ with } 4a^3 + 27b^2 \neq 0\}$$

has a compression scheme. What is its dimension?

The condition "$4a^3 + 27b^2 \neq 0$" is placed there to guarantee that the curve is non-singular.

The following is the fundamental theorem of stability theory. We prove it here using the techniques of 2-rank.

THEOREM 4.1.6 (Stable Formula Theorem, [**20**]). *Fix a formula $\varphi(x; y)$. The following are equivalent:*

(1) *φ is stable (i.e., $R_{2,\varphi}(\emptyset) < \omega$);*
(2) *There does not exist $\langle a_i : i < \omega\rangle$ from $\mathcal{U}_x$ and $\langle b_j : j < \omega\rangle$ from $\mathcal{U}_y$ such that, for all $i, j < \omega$, $\models \varphi(a_i; b_j)$ if and only if $i < j$.*
(3) *For all infinite $B \subseteq \mathcal{U}_y$, $|S_\varphi(B)| \leq |B|$.*
(4) *There exists finitely many $\psi_0(y; z), ..., \psi_{N-1}(y; z)$ such that, for all $B \subseteq \mathcal{U}_y$ and $p \in S_\varphi(B)$, there exists $c \in B^n$ and $i < N$ such that $\psi_i(y; c)$ defines $p(x)$.*

PROOF. (1) $\Rightarrow$ (4) : A study of the proof of Theorem 4.1.3 shows that there is no need to assume that $B \subseteq \mathcal{U}_y$ is finite. Therefore, this gives us (4).

(4) $\Rightarrow$ (3) : Suppose that $\psi_0(y; z), ..., \psi_{N-1}(y; z)$ satisfy (4). Then, for any infinite $B \subseteq \mathcal{U}_y$, each type $p \in S_\varphi(B)$ is determined by a choice of $c \in B^n$ and $i < N$. Hence,

$$|S_\varphi(B)| \leq N|B|^n = |B|.$$

(3) $\Rightarrow$ (2) : Suppose (2) fails. By compactness, we may assume that there exists $\langle a_r : r \in \mathbb{R}\rangle$ and $\langle b_q : q \in \mathbb{Q}\rangle$ such that, for all $q \in \mathbb{Q}$ and $r \in \mathbb{R}$, $\models \varphi(a_r; b_q)$ if and only if $r < q$. Let $B = \{b_q : q \in \mathbb{Q}\}$, which has cardinality $\aleph_0$. Now, for any choice $r_0 < r_1$ from $\mathbb{R}$, there exists $q \in \mathbb{Q}$ so that $r_0 < q < r_1$. Therefore,

$$\models \varphi(a_{r_0}; b_q) \wedge \neg\varphi(a_{r_1}; b_q).$$

Hence, $\mathrm{tp}_\varphi(a_{r_0}/B) \neq \mathrm{tp}_\varphi(a_{r_1}/B)$. This shows that

$$|S_\varphi(B)| \geq |\mathbb{R}| = 2^{\aleph_0} > \aleph_0 = |B|.$$

(2) $\Rightarrow$ (4) : We will hold off of providing a proof for this. The proof of Theorem 4.5.5 gives a specialized version of this statement.

(3) $\Rightarrow$ (1) : Suppose that (1) fails, so $R_{2,\varphi}(\emptyset) \geq \omega$. Then, we use this to construct $b_\tau \in \mathcal{U}_y$ for each $\tau \in {}^{<\omega}2$ such that, for all $n < \omega$ and $\sigma \in {}^n2$,

$$R_{2,\varphi}(\{\varphi(x; b_{\sigma|_i})^{\sigma(i)} : i < n\}) \geq \omega.$$

Choose $b_{\langle\rangle} \in \mathcal{U}_y$ such that, for both $t < 2$, $R_{2,\varphi}(\{\varphi(x; b_{\langle\rangle})^t\}) \geq \omega$. In general, if b_τ is constructed for $\tau \in {}^{n-1}2$, then for any $\tau' \in {}^n2$ extending τ (of which there are exactly two), choose $b_{\tau'} \in \mathcal{U}_y$ such that, for all $t < 2$,

$$R_{2,\varphi}(\{\varphi(x; b_{\tau|_i})^{\tau'(i)} : i < n\} \cup \{\varphi(x; b_{\tau'})^t\}) \geq \omega.$$

This concludes the construction.

Finally, consider the set $B = \{b_\tau : \tau \in {}^{<\omega}2\}$. We have that $|B| = \aleph_0$. On the other hand, for each $\sigma \in {}^\omega 2$ (which has cardinality $2^{\aleph_0}$), the type

$$\{\varphi(x; b_{\sigma|_i})^{\sigma(i)} : i < \omega\}$$

is consistent by compactness. This extends to a type $p_\sigma(x) \in S_\varphi(B)$. For $\sigma \neq \sigma'$, $p_\sigma \neq p_{\sigma'}$ since, if $i < \omega$ is minimal such that $\sigma|_i = \sigma'|_i$, then $\varphi(x; b_{\sigma|_i})$ is in one type but not the other. Therefore,

$$|S_\varphi(B)| \geq |{}^\omega 2| = 2^{\aleph_0} > \aleph_0 = |B|.$$

□

The properties of stability and NIP are closely related to the strict order property (SOP).

Definition 4.1.7. We say that a formula $\varphi(x; y)$ has the *strict order property* (*SOP*) if there is a sequence $\langle b_j : j < \omega\rangle$ such that, for all $i, j < \omega$,

$$\models (\forall x)[\varphi(x; b_i) \to \varphi(x; b_j)] \text{ if and only if } i \leq j.$$

We say a theory T has the *strict order property* (*SOP*) if there is some formula that does.

Theorem 4.1.8 (Theorem II.4.7 (1) of [**20**]). *A theory T is unstable if and only if T has IP or SOP.*

Proof. If T has IP, then there exists a formula $\varphi(x; y)$, $\langle a_I : I \subseteq \omega\rangle$, and $\langle b_j : j < \omega\rangle$ such that, for all $j < \omega$ and $I \subseteq \omega$,

$$\models \varphi(a_I; b_j) \text{ if and only if } j \in I.$$

For each $i < \omega$, let $a'_i = a_{\{i' < \omega : i' > i\}}$. Then, $\langle a'_i : i < \omega\rangle$ and $\langle b_j : j < \omega\rangle$ witness the fact that $\varphi(x; y)$ has the order property. By Theorem 4.1.6 (2), φ is unstable. Hence T is unstable.

If T has SOP, then let $\varphi(x;y)$ and $\langle b_j : j < \omega\rangle$ witness this fact. For each $i < \omega$, choose a_i such that

$$\models \neg\varphi(a_i;b_i) \wedge \varphi(a_i;b_{i+1})$$

(which is consistent by assumption). We see that $\langle a_i : i < \omega\rangle$ and $\langle b_j : j < \omega\rangle$ witness the fact that $\varphi(x;y)$ has the order property. As before, T is unstable.

Conversely, suppose T is unstable but has NIP. By Theorem 4.1.6 (2), there exists $\varphi(x;y)$ with the order property, so there exists $\langle a_i : i < \omega\rangle$ and $\langle b_j : j < \omega\rangle$ such that $\models \varphi(a_i;b_j)$ if and only if $i < j$. By the compactness theorem and Ramsey's theorem, we may assume that $\langle b_j : j < \omega\rangle$ is indiscernible. Let n be greater than the independence dimension of $\varphi(x;y)$. Therefore, there exists $s \in {}^n 2$ such that

$$\models \neg(\exists x)\left[\bigwedge_{j<n} \varphi(x;b_j)^{s(j)}\right].$$

On the other hand, take $s^* \in {}^n 2$ so that

(1) $|\mathrm{supp}(s^*)| = |\mathrm{supp}(s)|$ (i.e., $|\{j < n : s(j) = 1\}| = |\{j < n : s^*(j) = 1\}|$), and
(2) if $i < j < n$ and $s^*(i) = 1$, then $s^*(j) = 1$.

Then, as witnessed by $a_{n-|\mathrm{supp}(s)|-1}$,

$$\models (\exists x)\left[\bigwedge_{j<n} \varphi(x;b_j)^{s^*(j)}\right].$$

Since traspositions generate the symmetric group on n elements, there exists $s_0, s_1 \in {}^n 2$ and $i_0 < n-1$ such that

(1) $|\mathrm{supp}(s_0)| = |\mathrm{supp}(s_1)| = |\mathrm{supp}(s)|$,
(2) $s_0(i) = s_1(i)$ for all $i \neq i_0, i_0+1$,
(3) $s_0(i_0) \neq s_1(i_0)$, $s_0(i_0+1) \neq s_1(i_0+1)$, and

$$\models \neg(\exists x)\left[\bigwedge_{j<n} \varphi(x;b_j)^{s_0(j)}\right] \wedge (\exists x)\left[\bigwedge_{j<n} \varphi(x;b_j)^{s_1(j)}\right].$$

Let

$$\psi(x;y,y_0,...,y_{i_0-1},y_{i_0+2},...,y_{n-1}) = \bigwedge_{\substack{i<n,\\ i\neq i_0,i_0+1}} \varphi(x;y_i)^{s_0(i)} \wedge \varphi(x;y).$$

We show that ψ has SOP, giving us the desired conclusion. For any $N < \omega$, let

$$c = \langle b_0,...,b_{i_0-1},b_{i_0+2+N},...,b_{n+N-1}\rangle.$$

For any $i_0 < \ell < k < i_0 + 2 + N$,

$$\models (\exists x)[\psi(x; b_k, c) \wedge \neg\varphi(x; b_\ell)]$$

and

$$\models (\forall x)[\psi(x; b_\ell, c) \to \varphi(x; b_k)].$$

Hence, for all $i_0 < \ell, k < i_0 + 2 + N$

$$\models (\forall x)[\psi(x; b_\ell, c) \to \psi(x; b_k, c)] \text{ if and only if } \ell \leq k.$$

By compactness, ψ has SOP. □

For the remainder of this section, we will guide an exercise for a direct proof of Theorem 4.1.6, (2) ⇒ (1).

First, as in the proof of (3) ⇒ (1), we have the following: For a formula $\varphi(x; y)$, if $R_{2,\varphi}(\emptyset) \geq \omega$, then there exists $b_\tau \in \mathcal{U}_y$ for all $\tau \in {}^{<\omega}2$ such that, for each $\sigma \in {}^{\omega}2$, the type

$$\{\varphi(x; b_{\sigma|_i})^{\sigma(i)} : i < \omega\}$$

is consistent. Using this, for each $N < \omega$, we will construct a witness to the order property. That is, we will construct $a_i \in \mathcal{U}_x$ for $i < N$ and $b_j \in \mathcal{U}_y$ for $j < N$ such that, for all $i, j < N$,

$$\models \varphi(a_i; b_j) \text{ if and only if } i \leq j.$$

Fix $m < \omega$ and consider the tree $T = {}^{<m}2$. For $k \leq m$, we will say that a subtree $T' \subseteq T$ is *k-good* if

1. (1) $\langle T'; \subseteq\rangle$ is isomorphic to $\langle {}^{<k}2; \subseteq\rangle$, and
2. (2) for each non-terminal node $\tau \in T'$ and each $t < 2$, there exists a node $\tau' \in T'$ extending $\tau^\frown\langle t\rangle$.

Exercise 4.1.9. Show that, for all $k \leq m < \omega$, if $T = {}^{<m}2$ is partitioned into two sets $X, Y \subseteq T$, then either there exists a k-good subtree of X or there exists a $(m-k)$-good subtree of Y. Hint: Induction on m.

Fix $\varphi(x; y)$, assume that $R_{2,\varphi}(\emptyset) \geq \omega$, and fix $N < \omega$. Suppose that we have constructed the following sequence for $m \leq n < N$:

$$I_n = \langle a_0, b_0, ..., a_{m-1}, b_{m-1}, A', B', a_m, b_m, ..., a_{n-1}, b_{n-1}\rangle$$

such that

1. (1) $a_i \in \mathcal{U}_x$ for $i < n$, $b_j \in \mathcal{U}_y$ for $j < n$, $A' = \{a'_\sigma : \sigma \in {}^{3^{N-n+1}}2\}$, $B' = \{b'_\tau : \tau \in {}^{<3^{N-n+1}}2\}$ such that, for each $\sigma \in {}^{2^{N-n}}2$,
$$\models \bigwedge_{i < 3^{N-n+1}} \varphi(a_\sigma; b'_{\sigma|_i})^{\sigma(i)},$$
2. (2) for all $i, j < n$, $\models \varphi(a_i; b_j)$ if and only if $i \leq j$,

(3) for all $i < n$ and $b' \in B'$, $\models \varphi(a_i; b')$ if and only if $i < m$, and
(4) for all $a' \in A$ and $j < n$, $\models \varphi(a'; b_j)$ if and only if $m \leq j$.

Exercise 4.1.10. Show that $I_0 = \langle A', B' \rangle$ can be constructed to satisfy (1) through (4) (vacuously, just (1)).

For each $a' \in A'$ and $t < 2$, let $B'_t(a') = \{b' \in B' :\models \varphi(a'; b')^t\}$. Notice that $B'_0(a')$ and $B'_1(a')$ is a partition of B' (and we can view B' as a tree isomorphic to ${}^{<3^{N-n+1}}2$ via $\tau \mapsto b'_\tau$). Therefore, by Exercise 4.1.9, either

(i) there exists $B'' \subseteq B'_0(a')$ a subtree that is $(3^{N-n} + 1)$-good, or
(ii) there exists $B'' \subseteq B'_1(a')$ a subtree that is $(3^{N-n} + 1)$-good.

Exercise 4.1.11. Assume (ii) holds for some $a' \in A'$. Then, let $b'_\tau \in B''$ be the root of B'' and let B^* be the 3^{N-n}-good subtree of B'' extending $b'_{\tau^\frown\langle 0 \rangle}$. Let $A^* \subseteq A'$ witness the branches of B^*. Show that I_{n+1} obtained by replacing $\langle A', B' \rangle$ in I_n with $\langle a', b', A^*, B^* \rangle$ satisfies conditions (1) through (4) above.

Exercise 4.1.12. Assume (i) holds for all $a' \in A'$. Let $b' \in B'$ be the root of B' (i.e., $b' = b'_{\langle\rangle}$). In particular, (i) holds for $a' = a'_\sigma$ where σ extends $\langle 1 \rangle$. Choose $B^* \subseteq B''$ a 3^{N-n}-good subtree of B'' above $b'_{\langle 1 \rangle}$. Let $A^* \subseteq A'$ witness the branches of B^*. Show that I_{n+1} obtained by replacing $\langle A', B' \rangle$ in I_n with $\langle A^*, B^*, a', b' \rangle$ satisfies conditions (1) through (4) above.

This completes the proof of Theorem 4.1.6, (2) $\Rightarrow$ (1).

2. dp-Rank

The main one of interest in this section is dp-rank, but it is closely related to $\mathrm{VC_{ind}}$-density, so we introduce both. The idea behind dp-rank is the notion of crafting a general notion of "dimension" for NIP theories. It turns out that this generalizes the notion of dimension for o-minimal theories (which is the standard dimension in $\mathbb{R}$). Like dimension, dp-rank enjoys the property of subadditivity (see Theorem 4.2.16 below).

2.1. Local dp-Rank. First, we begin locally (i.e., with ranks on formulas).

Definition 4.2.1. Let $\varphi(x; y)$ be a formula, $p(x)$ a partial type, and $\ell \in \mathbb{R}$. We say that φ has *VC_{ind}-density* $\leq \ell$ *over* p if there exists $K < \omega$ such that, for all finite indiscernible sequences $\langle b_i : i < N \rangle$,

$$|S_\varphi(\{b_i : i < N\}) \cap [p(x)]| \leq KN^\ell$$

(as defined in (1.3)).

That is, the number of φ-types over $B = \{b_i : i < N\}$ consistent with $p(x)$ is bounded above by KN^ℓ.

This differs from VC-density since we only look over indiscernible sequences. We also define local dp-rank by studying the alternation of indiscernible sequences (similar to that of Definition 3.1.2).

To define dp-rank, we look at concept classes on $\mathbb{R}$. Fix $n < \omega$, $C \in \binom{\mathbb{R}}{n}$, and $f : (2n+1) \to 2$, define the function $g_{f,C} \in {}^{\mathbb{R}}2$ as follows: If $c_0 < ... < c_{n-1}$ is an enumeration of C, then set

$$\begin{aligned} &g_{f,C}(i) = f(0) \text{ for all } i < c_0, \\ &g_{f,C}(c_\ell) = f(2\ell+1) \text{ for all } \ell < n, \\ &g_{f,C}(i) = f(2\ell) \text{ if } c_{\ell-1} < i < c_\ell \text{ for } 0 < \ell < n \text{ and,} \\ &g_{f,C}(i) = f(2n) \text{ for all } i > c_{n-1}. \end{aligned} \tag{4.1}$$

We say that f is *n-minimizing* if, for all $\ell < n$,

$$\{f(2\ell), f(2\ell+1), f(2\ell+2)\} = \{0, 1\}.$$

That is, among the points around $2\ell + 1$ (corresponding to c_ℓ), the f takes on different values. We define the concept class on $\mathbb{R}$

$$\mathcal{C}_f = \left\{ g_{f,C} : C \in \binom{\mathbb{R}}{n} \right\}.$$

The idea behind this is that $\mathcal{C}_f$ is the concept class of unions of points and intervals, with a total of n "endpoints," coded by the function f. Instead of using the function f to code this, we could have used a quantifier-free formula $\psi(x; y_0, ..., y_{n-1})$ in the language of just order, $L = \{<\}$. For example, the function $f : 5 \to 2$ given by

$$f = \{\langle 0, 0\rangle, \langle 1, 1\rangle, \langle 2, 1\rangle, \langle 3, 0\rangle, \langle 4, 0\rangle\}$$

would code the concept class $\mathcal{C}_f$ of all intervals of the form $[a, b)$ for $a, b \in \mathbb{R}$, which could also be coded by the formula $\psi(x; y_0, y_1) = y_0 \leq x < y_1$. The function f is n-minimizing exactly when we cannot rewrite the union of points and intervals with fewer "endpoints."

Lemma 4.2.2. *For any f that is n-minimizing, $\mathcal{C}_f$ is maximum of dimension n.*

Proof. Fix a finite $D \subseteq \mathbb{R}$ and we proceed by induction on $m = |D|$ and n. If $n = 0$, this is trivial, so assume $n \geq 1$. If $m \leq n$, then we show that D is shattered by $\mathcal{C}_f$. Let $d_0 < ... < d_{m-1}$ enumerate D. Then, since f is n-minimizing, $\{f(2\ell), f(2\ell+1), f(2\ell+2)\} = \{0, 1\}$ for each $\ell < m \leq n$. Hence, if we want to represent $h \in {}^D2$, for each $\ell < m$, choose

(1) $c_\ell > d_\ell$ if $h(\ell) = f(2\ell)$,
(2) $c_\ell = d_\ell$ if $h(\ell) = f(2\ell+1)$, and
(3) $c_\ell < d_\ell$ if $h(\ell) = f(2\ell+2)$.

Therefore, D is shattered by $\mathcal{C}_f$.

So suppose $n \geq 1$ and $m > n$. Fix $a = \max D$, let $D^* = D \setminus \{a\}$, and define

$$\mathcal{C}_f^{a-} = \left\{ g_{f,C} : C \in \binom{\mathbb{R}}{n}, \max C \geq a \right\}.$$

and

$$\mathcal{C}_f^{a+} = \left\{ g_{f,C} : C \in \binom{\mathbb{R}}{n}, \max C < a \right\}.$$

By induction, $\mathcal{C}_f^{a-}|_{(-\infty,a)}$ is a concept class on $(-\infty, a)$ that is maximum of dimension $n-1$ and $D^* \subseteq (-\infty, a)$. Therefore,

$$|\mathcal{C}_f^{a-}|_{D^*}| = \Phi_{n-1}(m-1).$$

Also by induction,

$$|\mathcal{C}_f|_{D^*}| = \Phi_n(m-1).$$

It is easy to check that, for all $g \in \mathcal{C}_f$, either $g \in \mathcal{C}_f^{a-}$ or $g \in \mathcal{C}_f^{a+}$. Hence,

$$|\mathcal{C}_f^{a-}|_{D^*}| + |\mathcal{C}_f^{a+}|_{D^*}| = |\mathcal{C}_f|_{D^*}|.$$

If $g \in \mathcal{C}_f^{a+}$, then $g(a) = f(2n)$ is fixed. If $g \in \mathcal{C}_f^{a-}$, then both choices of $g(a) = 0$ and $g(a) = 1$ are always possible (as you can move the last element of C to either side of a). Therefore,

$$|\mathcal{C}_f|_D| = 2 \cdot |\mathcal{C}_f^{a-}|_{D^*}| + |\mathcal{C}_f^{a+}|_{D^*}| = \Phi_{n-1}(m-1) + \Phi_n(m-1) = \Phi_n(m).$$

□

Definition 4.2.3. Let $\varphi(x; y)$ be a formula, $p(x)$ a partial type, and $n < \omega$. We say that φ has *dp-rank* $\leq n$ *over* p if, for all $a \in \mathcal{U}_x$ with $a \models p$ and indiscernible sequence $\langle b_i : i \in \mathbb{R} \rangle$, there exists $C \subseteq \mathbb{R}$ with $|C| = n$ and $f \in {}^{2n+1}2$ such that

$$(\forall i \in \mathbb{R}) \left[\models \varphi(a; b_i)^{g_{f,C}(i)} \right], \tag{4.2}$$

where $g_{f,C}$ is defined as in (4.1).

That is, φ has dp-rank $\leq n$ over p if, for all $a \models p$ and indiscernible sequence $\langle b_i : i \in \mathbb{R} \rangle$, the set $I = \{i \in \mathbb{R} : \models \varphi(a; b_i)\}$ is a union of points and intervals with at most n "endpoints."

We introduce a third rank called VC-maximum-rank.

Definition 4.2.4 (Definition 3.4 of [**13**]). Fix a concept class $\mathcal{C}$ on a set X. We say that $\mathcal{C}$ has *VC-maximum-rank* $\geq n$ if there exists an infinite $Y \subseteq X$ and $\mathcal{D} \subseteq \mathcal{C}|_Y$ such that $\mathcal{D}$ is maximum of dimension n (i.e., for all $Y \subseteq X$ with $n \leq |Y| < \omega$, $|\mathcal{C}|_Y| = \Phi_n(|Y|)$).

Fix a formula $\varphi(x; y)$, $p(x)$ a partial type, and $n < \omega$. We say that $\varphi(x; y)$ has *VC-maximum-rank* $\geq n$ *over* p if the concept class

$$\mathcal{C} = \{\varphi(a; M_y) : a \models p, a \in \mathcal{U}_x\}.$$

on $\mathcal{U}_y$ has VC-maximum-rank $\geq n$.

That is, there exists $A \subseteq \mathcal{U}_x$ with $a \models p$ for all $a \in A$ and $B \subseteq \mathcal{U}_y$ infinite such that, for all $B_0 \subseteq B$ with $|B_0| \geq n$,

$$|S_\varphi(B_0) \cap [x \in A]| = \Phi_n(|B_0|).$$

This definition is especially interesting because it involves no model theory. This is really a measure on concept classes. Therefore, it is easy to translate back to the realm of pure theoretical computer science. Consequently, we can use this rank to "explain" the idea of dp-rank to non-model theorists.

The main objective of this subsection is to show the following theorem:

Theorem 4.2.5 (Theorem 3.3 of [**8**] and Theorem 3.4 of [**13**]). *If $\varphi(x; y)$ is a formula with NIP, $p(x)$ is a partial type, and $n < \omega$, then the following are equivalent:*

(1) *φ has VC$_{\text{ind}}$-density $\leq n$ over p.*
(2) *φ has VC$_{\text{ind}}$-density $< n + 1$ over p.*
(3) *φ has dp-rank $\leq n$ over p.*
(4) *φ has VC-maximum-rank $\leq n$ over p.*

We get the following corollary immediately from (1) $\Leftrightarrow$ (2):

Corollary 4.2.6 (Theorem 1.6 of [**8**]). *VC$_{\text{ind}}$-density is always integer valued.*

This is in sharp contrast to VC-density, which can be non-integer valued.

Proof of Theorem 4.2.5. (1) $\Rightarrow$ (2): Trivial.

(2) $\Rightarrow$ (3): Suppose $\varphi(x; y)$ has dp-rank $\geq n + 1$ over p, witnessed by $a \in \mathcal{U}_x$ with $a \models p$ and an indiscernible sequence $\langle b_i : i \in \mathbb{Q}\rangle$. Fix any $K < \omega$ and we show that

$$|S_\varphi(\{b_i : i < K\}) \cap [p(x)]| \geq \binom{K}{n+1},$$

showing that the VC$_{\text{ind}}$-density of φ is $\geq n + 1$.

Since φ has NIP, it has finite alternation rank by Proposition 3.1.3. Let $m < \omega$ be minimal such that there exists $C \in \binom{\mathbb{R}}{m}$ and $f \in {}^{2m+1}2$ such that (4.2) holds. Thus, f is m-minimizing. By assumption, $m \geq n+1$. Let $c_0 < ... < c_{m-1}$ enumerate the elements of C. For each $i_0 < ... < i_n < K$, choose an order-preserving map $h_{\bar{i}} : K \to \mathbb{R}$ such that

$$(\forall \ell \leq n) \left[h_{\bar{i}}(i_\ell) = c_\ell \right].$$

By indiscernibility, there exists $a_{\bar{i}} \models p$ such that, for all $j < K$,

$$\models \varphi(a; b_j) \leftrightarrow \varphi(a_{\bar{i}}; b_{h_{\bar{i}}(j)}).$$

We simply need to show that $\mathrm{tp}_\varphi(a_{\bar{i}}/\{b_i : i < K\})$ is distinct for each choice of $\bar{i}$ (as there are $\binom{K}{n+1}$ such choices).

Fix $i_0 < ... < i_n < K$ and $i'_0 < ... < i'_n < K$ distinct. Suppose ℓ is minimal such that $i_\ell \neq i'_\ell$. Since f is m-minimizing, it is easy to check that

$$\models \varphi(a_{\bar{i}}; b_{i_\ell}) \leftrightarrow \neg\varphi(a_{\bar{i}'}; b_{i_\ell})$$

or

$$\models \varphi(a_{\bar{i}}; b_{i'_\ell}) \leftrightarrow \neg\varphi(a_{\bar{i}'}; b_{i'_\ell}).$$

For example, if $i_\ell < i'_\ell$ and $f(2\ell+1) \neq f(2\ell+2)$, then the second condition holds. This gives the desired conclusion.

(3) $\Rightarrow$ (1): Suppose that the $\mathrm{VC}_{\mathrm{ind}}$-density of φ over p is $> n$. Therefore, by compactness, there exists an indiscernible sequence $\langle b_i : i \in \mathbb{R} \rangle$ and $K < \omega$ such that, for all finite $D \subseteq \mathbb{R}$ with $|D| \geq K$,

$$|S_\varphi(\{b_i : i \in D\}) \cap [p(x)]| > 2^{2n+1} \binom{|D|}{n}.$$

Fix such a D and let $m = |D| \geq K$. By means of contradiction, suppose that φ has dp-rank $\leq n$ over p. Then, for each $a \models p$, there exists $C \in \binom{\mathbb{R}}{n}$ and $f \in {}^{2n+1}2$ such that (4.2) holds. But this only depends on a choice of $C \in \binom{D}{n}$. Therefore, the number of φ-types over $\{b_i : i \in D\}$ is bounded by the number of choices of C and f. Hence,

$$|S_\varphi(\{b_i : i \in D\}) \cap [p(x)]| \leq 2^{2n+1} \binom{m}{n}.$$

Hence, $m^\ell \leq 2^{2n+1}\binom{m}{n} \leq 2^{2n+1} m^n$. Since this holds for arbitrarily large m and $n < \ell$, this is a contradiction.

(1) $\Rightarrow$ (4): Suppose φ has VC-maximum-rank $\geq n+1$ over p. Thus, there exists $B \subseteq \mathcal{U}_y$ infinite and $A \subseteq \mathcal{U}_x$ with $a \models p$ for all $a \in A$ such that, for all $B_0 \subseteq B$ finite,

$$|S_\varphi(B_0) \cap [x \in A]| = \Phi_{n+1}(|B_0|).$$

In particular, since $A \subseteq p(\mathcal{U})$,

$$|S_\varphi(B_0) \cap [p(x)]| \geq \Phi_{n+1}(|B_0|).$$

By Ramsey's Theorem and compactness, we may assume that B is the image of an indiscernible sequence, $\langle b_i : i \in \mathbb{R}\rangle$. Therefore, for any $m < \omega$ and $D \in \binom{\mathbb{R}}{m}$,

$$|S_\varphi(\{b_i : i \in D\}) \cap [p(x)]| \geq \Phi_{n+1}(m) \geq \binom{m}{n+1}$$

However, for any $K < \omega$, for sufficiently large m, $\binom{m}{n+1} > Km^n$. Therefore, the $\mathrm{VC_{ind}}$-density of φ is $> n$.

(4) $\Rightarrow$ (3): Suppose that the dp-rank of φ is $\geq n+1$ over p. Let $\langle b_i : i \in \mathbb{R}\rangle$ and $a \models p$ witness this. Let $C \in \binom{\mathbb{R}}{m}$ and $f \in {}^{2m+1}2$ be so that (4.2) holds and choose $m < \omega$ minimal such (some such choice exists, otherwise φ would have infinite alternation rank, hence would not be NIP). Hence, f is m-minimizing. By assumption, $m > n$ and, by truncating the sequence if necessary, we may assume $m = n+1$.

By indiscernibility and compactness, for any choice of $C' \in \binom{\mathbb{R}}{m}$, there exists $a_{C'} \models p$ so that

$$(\forall i \in \mathbb{R}) \left[\models \varphi(a_{C'}; b_i)^{g_{f,C'}(i)}\right],$$

where $g_{f,C'}$ is defined as in (4.1). Let

$$A = \left\{a_{C'} : C' \in \binom{\mathbb{R}}{m}\right\}$$

and $B = \{b_i : i \in \mathbb{R}\}$. We claim that A and B witness that φ has VC-maximum-rank $\geq n+1$ over p. But this follows immediately from Lemma 4.2.2, since

$$\mathcal{C}_f = \{\chi_{\varphi(a';B)} : a' \in A\}.$$

□

Example 4.2.7 (Proposition 4.6 of [**1**]). Suppose K is an infinite field in the language $L = \{+, \cdot, 0, 1\}$. Consider the formula

$$\begin{aligned}\varphi(x_0, x_1; y_0, y_1, y_2, y_3) = &[y_1 = y_2 y_0 + y_3] \wedge \\ &[\langle x_0, x_1\rangle = \langle y_0, y_1\rangle \vee \langle x_0, x_1\rangle = \langle y_2, y_3\rangle].\end{aligned}$$

Then,

(1) If K has characteristic zero, the VC-density of φ is $\frac{4}{3}$.
(2) If K has characteristic $p > 0$, then the VC-density of φ is $\frac{3}{2}$.
(3) The $\mathrm{VC_{ind}}$-density of φ is 1.

Open Question 4.2.8. *Does there exist a formula $\varphi(x;y)$ in a theory T with VC-density irrational?*

We continue this discussion in Section 4.3, where we discuss VC-density in more depth.

2.2. Global dp-Rank. Let $p(x)$ be any partial type, let α and β be ordinals, let $\overline{\varphi} = \langle \varphi_i(x;y_i) : i < \alpha \rangle$ be a set of functions, and let $\overline{b} = \langle b_{i,j} : i < \alpha, j < \beta \rangle$ where $b_{i,j} \in \mathbb{U}_{y_i}$. We say that $\langle \overline{\varphi}, \overline{b} \rangle$ is a *ICT-pattern in p of depth α and length β* if, for all functions $f : \alpha \to \beta$, the following type is consistent:

$$p(x) \cup \{\varphi_i(x; b_{i,f(i)}) : i < \alpha\} \cup \{\neg\varphi_i(x; b_{i,j}) : i < \alpha, j \neq f(i)\}. \tag{4.3}$$

Definition 4.2.9. We say that $p(x)$ has *dp-rank* $\leq \alpha$ if there exists no ICT-pattern in p of depth $\alpha + 1$ and length ω, and we denote this by $\mathrm{dpR}(p) \leq \alpha$.

Finally, we say that a set of indiscernible sequences

$$\{\langle b_{i,j} : j < \omega \rangle : i < \alpha\}$$

is mutually indiscernible if, for each i, $\langle b_{i,j} : j < \omega \rangle$ is indiscernible over $\{b_{i',j} : i' \neq i, j < \omega\}$.

Proposition 4.2.10 ([**15**], [**22**]). *For a partial type $p(x)$ and ordinal α, the following are equivalent:*

(i) *p has dp-rank $< \alpha$.*
(ii) *For $\{\langle b_{i,j} : j < \omega \rangle : i < \alpha\}$ mutually indiscernible and $a \models p$, there exists $i < \alpha$ such that $\langle b_{i,j} : j < \omega \rangle$ is indiscernible over a.*

Proof. (i) $\Rightarrow$ (ii): Suppose p has dp-rank $\geq \alpha$, so there exists an ICT-pattern in p of depth α and length ω. Fix $\langle \overline{\varphi}, \overline{b} \rangle$ such a pattern. Add each $b_{i,j}$ and a_f for each $f : \alpha \to \omega$ as constant symbols and consider the set of formulas Σ stating T, $\langle \overline{\varphi}, \overline{b} \rangle$ is an ICT-pattern so that $a_f \models p$ in (4.3), and $\{\langle \overline{b}_{i,j} : j < \omega \rangle : i < \alpha\}$ is mutually indiscernible. By compactness, it suffices to show that each finite $\Sigma_0 \subseteq \Sigma$ is consistent, which amounts to showing that, for some finite $A \subseteq \alpha$ and $n < \omega$, this holds for $i \in A$ and $j < n$. Focus on $\{b_{i,j} : j < \omega, i \in A\}$ and, by Ramsey's Theorem on $|A|$ disjoint sets, we may assume that $\{\langle b_{i,j} : j < n \rangle : i \in A\}$ is mutually indiscernible. It now follows that we may assume $\overline{b}$ is mutually indiscernible.

Now, for any choice of $f : \alpha \to \omega$, take a realizing the type (4.3). Then, for each $i < \alpha$, $\langle b_{i,j} : j < \omega \rangle$ is clearly not indiscernible over a, since $\models \neg\varphi_i(a; b_j)$ for $j \neq f(i)$ yet $\models \varphi_i(a; b_{f(i)})$.

(ii) ⇒ (i): Suppose $\{\langle b_{i,j} : j < \omega\rangle : i < \alpha\}$ is mutually indiscernible, $a \models p$, yet for each $i < \alpha$, $\langle b_{i,j} : j < \omega\rangle$ is not indiscernible over a. Therefore, there exists $\psi_i(x; y_1, ..., y_{n_i})$, $j_1 < ... < j_n < \omega$, and $j'_1 < ... < j'_n < \omega$ (where $n = n_i$) such that

$$\models \psi_i(a; b_{i,j_1}, ..., b_{i,j_n}) \wedge \neg\psi_i(a; b_{i,j'_1}, ..., b_{i,j'_n}).$$

By choosing $k > \max\{j_n, j'_n\}$ and possibly replacing the j_ℓ's with j'_ℓ's and ψ_i with $\neg\psi_i$, we may assume that

$$\models \psi_i(a; b_{i,j_1}, ..., b_{i,j_n}) \wedge \neg\psi_i(a; b_{i,k+1}, ..., b_{i,k+n}).$$

Define

$$\varphi_i(x; z_1, z_2) = \neg(\psi_i(x; z_1) \leftrightarrow \psi_i(x; z_2))$$

and we claim that $\langle \varphi_i : i < \alpha\rangle$ forms an ICT-pattern of depth α, showing that $\mathrm{dpR}(p) \geq \alpha$.

By compactness, it suffices to show that, for each finite $A \subseteq \alpha$ and $m < \omega$, there exists $\{\langle c_{i,j} : j < m\rangle : i \in A\}$ that, together with $\overline{\varphi}$, form an ICT pattern of depth $|A|$ and length m. By indiscernibility, there exists $a' \models p$ such that, for each $i \in A$ and $\ell \leq 2m$,

$$\models \varphi_i(a'; b_{i,2\ell n_i}, ..., b_{i,2(\ell+1)n_i-1}) \text{ if and only if } \ell = m.$$

By indiscernibility again, for each $f : A \to m$, there exists $a' \models p$ such that, for each $i \in A$ and $\ell < m$,

$$\models \varphi_i(a'; b_{i,2\ell n_i}, ..., b_{i,2(\ell+1)n_i-1}) \text{ if and only if } \ell = f(i).$$

Therefore, $\{\langle\langle b_{i,2\ell n_i}, ..., b_{i,2(\ell+1)n_i-1}\rangle : \ell < m\rangle : i \in A\}$ witnesses the desired ICT pattern. □

Local dp-rank is related to global dp-rank in the following way:

THEOREM 4.2.11 (Proposition 2.3 of [**8**]). *If $p(x)$ is any partial type, the dp-rank of $p(x)$ is $\leq n$ if and only if, for all formulas $\varphi(x; y)$, the dp-rank of $\varphi(x; y)$ over p is $\leq n$.*

PROOF. Suppose the dp-rank of p is $> n$, witnessed by an ICT-pattern $\langle\overline{\varphi}, \overline{b}\rangle$ of depth $n + 1$ and length ω. By compactness, we may assume that $\overline{b} = \langle b_{i,j} : i \leq n, j \in \mathbb{R}\rangle$. By Ramsey's Theorem, we may assume that $\langle\langle b_{0,j}, ..., b_{n,j}\rangle : j \in \mathbb{R}\rangle$ is indiscernible. Let

$$\psi(x; y_0, ..., y_n) = \bigvee_{i \leq n} \varphi_i(x; y_i).$$

Now, choose $a \models p$ so that $\models \varphi_i(a; b_{i,j})$ if and only if $j = i$, hence $\models \psi_i(a; b_{0,j}, ..., b_{n,j})$ if and only if $j = 0, ..., n$. Hence, any finite $C \subseteq \mathbb{R}$ with function f witnessing (4.2) in the definition of dp-rank for ψ must at least satisfy $j \in C$ for each $j = 0, ..., n$. Hence, $|C| \geq n+1$, showing that the dp-rank of ψ over p is at least $n+1$.

Conversely, suppose there exists $\varphi(x; y)$ where the dp-rank of φ over p is $\geq n+1$. Fix $a \models p$ and $\langle b_i : i \in \mathbb{R} \rangle$ witnessing this. Define the convex equivalence relation E on $\mathbb{R}$ as follows: rEq if the truth value of $\varphi(a; b_i)$ is constant on the $\mathbb{R}$-interval between r and q. By definition of the dp-rank of φ, there are more than $n+1$ infinite E-classes.

Hence, we can partition $\mathbb{R}$ into $n+1$ convex sets $C_0 < ... < C_n$ so that the truth value of $\varphi(a; b_i)$ is non-constant on each C_ℓ. Then,

$$\{\langle b_j : j \in C_i \rangle : i \leq n\}$$

is mutually indiscernible, yet no sequence is indiscernible over a. Therefore, the dp-rank of p is $\geq n+1$. $\square$

2.3. Subadditivity of dp-Rank. In this subsection, we suppose that T has NIP. In order to prove subadditivity of global dp-rank, we need a notion of limit type. Fix an ordinal α and, for each $i < \alpha$, J_i some linear order. Let

$$\mathcal{J} = \{\langle b_{i,j} : j \in J_i \rangle : i < \alpha\}$$

be a mutually indiscernible sequence. For each $i < \alpha$, choose $\sigma_i : \omega \to J_i$ a function that is strictly monotone (i.e., $(\forall k < k' < \omega)[\sigma_i(k) < \sigma_i(k')]$ or $(\forall k < k' < \omega)[\sigma_i(k) > \sigma_i(k')]$). Let $\overline{y} = \langle y_{i,k} : k < K, i < \alpha \rangle$ be some tuple of variables.

Lemma 4.2.12. *For any formula $\delta(\overline{y})$ over any set of parameters, there exists $N < \omega$ and $t < 2$ such that, for all*

$$N < \ell_{i,0} < ... < \ell_{i,K-1} < \omega$$

for each $i < \alpha$, we have

$$\models \delta(b_{i,\sigma_i(\ell_{i,k})})^t_{k<K,i<\alpha}.$$

Proof. Let $\delta(\overline{y}) = \delta(a; \overline{y})$ for some a and $\delta(x; \overline{y})$ over $\emptyset$. Since T has NIP, $\delta(x; \overline{y})$ has NIP, so suppose that it has independence dimension $< N$.

Suppose the conclusion fails. Then, there exists $\ell^0_{i,0} < ... < \ell^0_{i,K-1} < \omega$ for each $i < \alpha$ such that

$$\models \neg\delta(a; b_{i,\sigma_i(\ell^0_{i,k})})_{k<K,i<\alpha}.$$

by assumption. In general, suppose that

$$\ell^0_{i,0} < ... < \ell^0_{i,K-1} < ... < \ell^n_{i,0} < ... < \ell^n_{i,K-1} < \omega$$

have been constructed for each $i < \alpha$ so that

$$\models \delta(a; b_{i,\sigma_i(\ell^n_{i,k})})^{n \pmod 2}_{k<K,i<\alpha}.$$

Setting $N = \ell^n_{i,K-1}$ and $t = n \pmod 2$, the failure of the conclusion of the lemma gives us $N < \ell^{n+1}_{i,0} < ... < \ell^{n+1}_{i,K-1} < \omega$ for each $i < \alpha$ such that

$$\models \delta(a; b_{i,\sigma_i(\ell^{n+1}_{i,k})})^{n+1 \pmod 2}_{k<K,i<\alpha}.$$

This concludes our construction.

Now, we use mutual indiscernibility to contradict the fact that δ has independence dimension $< N$. For each $\eta : N \to 2$, consider the formula

$$\exists x \left(\bigwedge_{n<N} \delta(x; b_{,\sigma_i(\ell^{2n+\eta(n)}_{i,k})})^{\eta(n)}_{k<K,i<\alpha} \right).$$

This holds, as witnessed by a. By mutual indiscernibility, we have that

$$\models \exists x \left(\bigwedge_{n<N} \delta(x; b_{i,\sigma_i(\ell^n_{i,k})})^{\eta(n)}_{k<K,i<\alpha} \right).$$

This contradicts the fact that δ has independence dimension $< N$. $\square$

Definition 4.2.13. With $\mathcal{J}$, $\overline{\sigma}$, and $\overline{y}$ as above, define the *limit type* of $\mathcal{J}$ over a parameter set A with respect to $\overline{\sigma}$ in the variables $\overline{y}$ as follows: For any formula $\delta(\overline{y})$ over B,

$$\delta(\overline{y}) \in \lim_{\overline{\sigma}}(\mathcal{J}/A)(\overline{y})$$

if and only if there exists $N < \omega$ such that, for all $N < \ell_{i,0} < ... < \ell_{i,K-1} < \omega$ for each $i < \alpha$, we have

$$\models \delta(b_{i,\sigma_i(\ell_{i,k})})_{k<K,i<\alpha}.$$

By Lemma 4.2.12 (and compactness), $\lim_{\overline{\sigma}}(\mathcal{J}/A)(\overline{y})$ is a complete type over A in the variables $\overline{y}$.

Exercise 4.2.14. Suppose $\alpha < \omega$ and $\mathcal{J} = \{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i < \alpha\}$ is mutually indiscernible but not mutually indiscernible over A. Show that there exists a formula $\delta(y_{i,k})_{k<K,i<\alpha}$ over A such that

(1) $\delta(\overline{y}) \in \lim_{\overline{\sigma}}(\mathcal{J}/A)(\overline{y})$ for $\sigma_i : \omega \to \mathbb{Z}$ the standard injection for each $i < \alpha$, and
(2) there exists $j_{i,0} < ... < j_{i,K-1}$ from $\mathbb{Z}$ for each $i < \alpha$ such that
$$\models \neg\delta(b_{i,j_{i,k}})_{k<K,i<\alpha}.$$

We show the following proposition, which is an improvement on Proposition 4.2.10:

Proposition 4.2.15 (Proposition 4.4 of [**15**], Proposition 4.15 of [**24**]). *For a partial type $p(x)$ over A, the following are equivalent:*

(i) *p has dp-rank $\leq n$.*

(ii) *For all $k < \omega$, all $\{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i < k\}$ mutually indiscernible over A, and $a \models p$, there exists $I \subseteq k$ with $|I| \geq k - n$ such that $\{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i \in I\}$ is mutually indiscernible over $A \cup \{a\}$.*

Proof. (ii) $\Rightarrow$ (i) : Follows from Proposition 4.2.10, (ii) $\Rightarrow$ (i) by compactness.

(i) $\Rightarrow$ (ii) : Suppose that (ii) fails, witnessed by $k < \omega$, $\mathcal{J} = \{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i < k\}$, and $a \models p$. Clearly $k > n$, otherwise we could take $I = \emptyset$. Fix $N < \omega$ and $\theta(x) \in p(x)$ and we construct an ICT-pattern in θ of length N and depth $n+1$. By compactness, we conclude that (i) fails.

We proceed by induction on $m \leq n$. Consider $X = \{m, m+1, ..., k\}$ and suppose that L_{m-1} has been constructed ($L_{-1} = 0$). By assumption, $\mathcal{J}_X = \{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i \in X\}$ is not mutually indiscernible over $A \cup \{a\}$. For each choice of $\epsilon \in \{-1, 1\}$, let $\sigma_\epsilon : \omega \to \mathbb{Z}$ be given by $\sigma_\epsilon(\ell) = \epsilon\ell$. Then, for each $s \in {}^X\{-1,1\}$, let $\overline{\sigma}_s = \langle \sigma_{s(i)} : i \in X\rangle$. Now, for any choice of variables $\overline{y}$ and s as above, consider $\lim_{\overline{\sigma}_s}(\mathcal{J}/A \cup \{a\})(\overline{y})$.

Case 1. There exists $\overline{y}$, s_0, s_1 such that $\lim_{\overline{\sigma}_{s_0}}(\mathcal{J}_X/A \cup \{a\})(\overline{y}) \neq \lim_{\overline{\sigma}_{s_1}}(\mathcal{J}_X/A \cup \{a\})(\overline{y})$.

In this case, by switching one at a time, we may assume that there exists $i_0 \in X$ such that, for all $i \in X$, $s_0(i) = s_1(i)$ if and only if $i \neq i_0$. By rearranging sequences, we may assume $i_0 = m$. Then, there exists

$$\delta(a; \overline{y}) \in \left[\lim_{\overline{\sigma}_{s_1}}(\mathcal{J}_X/A \cup \{a\})(\overline{y}) \setminus \lim_{\overline{\sigma}_{s_0}}(\mathcal{J}_X/A \cup \{a\})(\overline{y})\right].$$

Set

$$\psi_m(x; \overline{y}^0, \overline{y}^1) = \delta(x; \overline{y}^0) \leftrightarrow \neg\delta(x; \overline{y}^1).$$

Now choose $L^* < \omega$ with $L^* > L_{m-1}$ sufficiently large such that, for all $L^* \leq \ell_{i,0} < ... < \ell_{i,K-1} < \omega$ for each $i \in X$, for all $t < 2$,

$$\models \delta(a; b_{i,\sigma_{s_t(i)}(\ell_{i,k'})})^t_{k'<K, i \in X}.$$

Now, for $j < N$, define

$$\begin{aligned} c_{m,N} = &\langle b_{i,\sigma_{s_0(i)}(L^*+2Kj+k')} : k' < 2K, m < i < k\rangle^\frown \\ &\langle b_{m,\sigma_{s_0(m)}(L^*+2KN-(2Kj+k'))} : k' < 2K\rangle \end{aligned}$$

For $j = N$, define

$$\begin{aligned} c_{m,N} = &\langle b_{i,\sigma_{s_0(i)}(L^*+2Kj+k')} : k' < 2K, m < i < k\rangle^\frown \\ &\langle b_{m,\sigma_{s_0(m)}(L^*+2Kj+k')} : k' < K\rangle^\frown \langle b_{m,\sigma_{s_1(m)}(L^*+2Kj+k')} : k' < K\rangle. \end{aligned}$$

For $N < j < 2N$, define

$$c_{m,j} = \langle b_{i,\sigma_{s_1(i)}(L^*+2Kj+k')} : k' < 2K, i \in X \rangle.$$

Finally, set $L_m = L^* + 2KN$. Notice that, for all $j < 2N$,

$$\models \psi_m(a; c_{m,j}) \text{ if and only if } j = N.$$

This concludes our construction for Case 1.

Case 2. For all $\overline{y}$, s_0, s_1, $\lim_{\overline{\sigma}_{s_0}}(\mathcal{J}_X/A \cup \{a\})(\overline{y}) = \lim_{\overline{\sigma}_{s_1}}(\mathcal{J}_X/A \cup \{a\})(\overline{y})$.

Since $\mathcal{J}_X$ is not mutually indiscernible over $A \cup \{a\}$, there exists some $\overline{y}$, $j_{i,0} < ... < j_{i,K-1}$ for each $i \in X$, and $\psi_m(x; \overline{y})$ over A such that

(1) $\psi_m(a; \overline{y}) \notin \lim_{\overline{\sigma}_1}(\mathcal{J}_X/A \cup \{a\})(\overline{y})$, and
(2) $\models \psi_m(a; b_{i,j_{i,k'}})_{k'<K, i \in X}$.

Let $s_0 \in {}^X\{-1, 1\}$ be the function that assigns -1 to m and 1 to everything else. Similarly, let s_1 be the function that assigns 1 to all elements of X. Choose $L^* > L_{m-1}$ sufficiently large so that, for all $L^* \le \ell_{i,0} < ... < \ell_{i,K-1} < \omega$ for each $i \in X$, for all $t < 2$,

$$\models \psi_m(a; b_{i,\sigma_{s_t(i)}(\ell_{i,k'})})^t_{k'<K, i \in X}.$$

Construct $c_{m,j}$ for $j < 2N$ as before so that, for all $j < 2N$,

$$\models \psi_m(a; c_{m,j}) \text{ if and only if } j = N.$$

This concludes our construction for Case 2 (and hence all cases).

Finally, by mutual indiscernibility, for any choice of $\eta : (n+1) \to N$,

$$\models \exists x \left(\theta(x) \wedge \bigwedge_{m \le n} \psi_m(x; c_{m,j})^{\text{iff } j=\eta(m)} \right).$$

This forms an ICT-pattern of depth $n + 1$ and length N in $\theta(x)$. By compactness, p has dp-rank $> n$. □

If a is a tuple and A is a set, then let $\text{dpR}(a/A) = \text{dpR}(\text{tp}(a/A))$. We now immediately obtain the subadditivity of dp-rank from the previous proposition.

THEOREM 4.2.16 (Subadditivity of dp-rank, Theorem 4.8 of [**15**], Proposition 4.18 of [**24**]). *For tuples a and b and a set A,*

$$\text{dpR}(a, b/A) \le \text{dpR}(a/A) + \text{dpR}(b/A \cup \{a\}).$$

PROOF. Suppose $\text{dpR}(a/A) = n_1$ and $\text{dpR}(b/A \cup \{a\}) = n_2$. Fix any $k < \omega$ and mutually indiscernible sequence $\{\langle b_{i,j} : j \in \mathbb{Z} \rangle : i < k\}$ over A. By Proposition 4.2.15, there exists $I_1 \subseteq k$ with $|I_1| = k - n_1$ such that $\{\langle b_{i,j} : j \in \mathbb{Z} \rangle : i \in I_1\}$ is mutually indiscernible over $A \cup \{a\}$.

Again by Proposition 4.2.15, there exists $I_2 \subseteq I_1$ with $|I_2| = k - n_1 - n_2$ such that $\{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i \in I_2\}$ is mutually indiscernible over $A \cup \{a, b\}$. By Proposition 4.2.15 in the other direction (since $\mathcal{J}$ was arbitrary), $\mathrm{dpR}(a, b/A) \leq n_1 + n_2$. $\square$

The original proof was done through expanding mutually indiscernible sequences using limit types. We outline this in the exercises below.

EXERCISE 4.2.17. Let $\mathcal{J} = \{\langle b_{i,j} : j \in J_i\rangle : i \in I\}$ be mutually indiscernible over A and suppose each J_i has no largest element (and, for simplicity, has cofiniality ω). For each $i \in I$, let J_i' be any linear order. Show that there exists $b_{i,j}$ for $j \in J_i'$ and $i \in I$ such that

$$\{\langle b_{i,j} : j \in J_i + J_i'\rangle : i \in I\} \text{ is mutually indiscernible over } A.$$

Hint: By compactness, we may assume each J_i' is finite and are all equal to some J'. Consider the limit type

$$\lim_{\overline{\sigma}}(\mathcal{J}/A \cup \{b_{i,j} : j \in J_i + J', i \in I\})(y_i)_{i\in I}.$$

Here σ_i is any sequence cofinal in J_i for each $i \in I$. Extend the sequence by placing elements between the new elements and the original sequence.

Fix a partial type $p(x)$ over A with dp-rank $\leq n$ and $n < k < \omega$. Fix $\mathcal{J} = \{\langle b_{i,j} : j \in J_i\rangle : i < k\}$ mutually indiscernible over A, $J' = \omega$, and extend to $\mathcal{J}' = \{\langle b_{i,j} : j \in J_i + J'\rangle : i < k\}$ mutually indiscernible over A as in Exercise 4.2.17. Let

$$B = A \cup \{b_{i,j} : j \in J', i < k\}.$$

Clearly $\mathcal{J}$ is mutually indiscernible over B. By Proposition 4.2.10, (ii) $\Rightarrow$ (i), there exists $i_0 < k$ so that $\langle b_{i_0,j} : j \in J_{i_0}\rangle$ is indiscernible over $B \cup \{a\}$. Without loss of generality, suppose $i_0 = 0$. Let

$$A' = A \cup \{b_{0,j} : j \in J_0\}.$$

By induction on k, we may suppose that there exists $I \subseteq \{1, ..., k\}$ with $|I| = k - n - 1$ such that

$$\{\langle b_{i,j} : j \in J_i + J'\rangle : i \in I\}$$

is mutually indiscernible over $A' \cup \{a\}$. Without loss, suppose $I = \{1, ..., k - n - 1\}$.

EXERCISE 4.2.18. Show $\langle b_{0,j} : j \in J_0\rangle$ is indiscernible over $A \cup \{a\} \cup \{b_{i,j} : j \in J_i, 0 < i < k - n\}$. Hint: By means of contradiction, suppose not. Use indiscernibility to slide parameters from J_i to J'. However, recall we assumed $\langle b_{0,j} : j \in J_0\rangle$ is indiscernible over $B \cup \{a\}$.

Thus, $\{\langle b_{i,j} : j \in J_i \rangle : i < k - n\}$ is mutually indiscernible over $A \cup \{a\}$, as we aimed to prove.

3. VC-density

In this section, we will use results from model theory to give bounds for the VC-density of concept classes. One main result is an immediate corollary of Corollary 3.3.5.

THEOREM 4.3.1. *If T is a weakly o-minimal theory and $\varphi(x; y)$ is any formula, then the VC-density of φ is $\leq |x|$.*

We also prove another result from [**2**], this one about strongly minimal theories.

THEOREM 4.3.2 (Theorem 3.1 of [**2**]). *If T is a strongly minimal theory and $\varphi(x; y)$ is any formula, then the VC-density of φ is $\leq |x|$.*

EXAMPLE 4.3.3. Consider the example of elliptic curves from Exercise 4.1.5. Let $T = \mathrm{Th}(\mathbb{C}; +, \cdot)$, which is a strongly minimal theory. Consider the concept class of elliptic curves

$$\mathcal{C} = \{C_{a,b} : a, b \in \mathbb{C} \text{ with } 4a^3 + 27b^2 \neq 0\}$$

where $C_{a,b} = \{\langle x, y \rangle \in \mathbb{C}^2 : y^2 = x^3 + ax + b\}$. The VC-density of this concept class is ≤ 2.

EXERCISE 4.3.4. Show that $\mathcal{C}$ in Example 4.3.3 has VC-density 2.

To prove Theorem 4.3.2, we will need the definition of Morley rank. Let $\varphi(x)$ be a formula over $\mathcal{U}$ and we inductively define the ordinal-valued $\mathrm{MR}(\varphi)$.

(1) $\mathrm{MR}(\varphi(x)) \geq 0$ if $\varphi(x)$ is consistent.
(2) $\mathrm{MR}(\varphi(x)) \geq \gamma$ for limit ordinal γ if $\mathrm{MR}(\varphi(x)) \geq \alpha$ for all $\alpha < \gamma$.
(3) $\mathrm{MR}(\varphi(x)) \geq \alpha + 1$ if, there exists $\{\psi_i(x) : i < \kappa\}$ (for $\kappa \geq \aleph_0$) formulas over $\mathcal{U}$ mutually inconsistent, each implying $\varphi(x)$, such that $\mathrm{MR}(\psi_i(x)) \geq \alpha$ for each $i < \kappa$.

When $T = \mathrm{Th}(\mathbb{C}; +, \cdot)$, the Morley Rank of $\varphi(x)$ is exactly equal to the Krull Dimension of $\varphi(\mathbb{C})$. We define the Morley Degree of $\varphi(x)$, denoted $\mathrm{Md}(\varphi(x))$, to be the maximal size N such that there exists formulas $\{\psi_i(x) : i < N\}$ over $\mathcal{U}$ mutually inconsistent, each implying $\varphi(x)$, such that $\mathrm{MR}(\psi_i(x)) = \mathrm{MR}(\varphi(x))$. Assuming that $\mathrm{MR}(\varphi(x))$ is finite, if no such N exists, then by compactness we would have a witness to the fact that $\mathrm{MR}(\varphi(x)) \geq \mathrm{MR}(\varphi(x)) + 1$, a contradiction. Therefore, Morley Degree exists and is finite.

Theorem 4.3.5. *Suppose T is a strongly minimal theory and $\varphi(x)$ and $\psi(x)$ are formulas. Then,*

1. $-1 \leq \mathrm{MR}(\varphi) \leq |x|$ *and* $1 \leq \mathrm{Md}(\varphi)$.
2. *If* $\psi \vdash \varphi$, *then* $\mathrm{MR}(\psi) \leq \mathrm{MR}(\varphi)$ *and, if* $\mathrm{MR}(\psi) = \mathrm{MR}(\varphi)$, *then* $\mathrm{Md}(\psi) \leq \mathrm{Md}(\varphi)$.
3. $\mathrm{MR}(\varphi \wedge \psi) = \mathrm{MR}(\varphi)$ *or* $\mathrm{MR}(\varphi \wedge \neg\psi) = \mathrm{MR}(\varphi)$.
4. *If* $\mathrm{MR}(\varphi) = \mathrm{MR}(\varphi \wedge \psi)$ *and* $\mathrm{Md}(\varphi \wedge \psi) < M(\varphi)$, *then* $\mathrm{MR}(\varphi) = \mathrm{MR}(\varphi \wedge \neg\psi)$.
5. *If* $\psi'(x)$ *is another formula over* $\mathcal{U}$, $\psi \vdash \varphi$, $\psi' \vdash \varphi$, *and* $\mathrm{MR}(\varphi) = \mathrm{MR}(\psi) = \mathrm{MR}(\psi') > \mathrm{MR}(\psi \wedge \psi')$, *then* $\mathrm{Md}(\varphi) \geq \mathrm{Md}(\psi) + \mathrm{Md}(\psi')$.
6. *For all L-formulas $\theta(x; y)$, there exists $k < \omega$ such that, for all $b \in \mathcal{U}_y$,* $\mathrm{Md}(\theta(x; b)) \leq k$.

We will omit the proof of this theorem here. Notice that these statements hold for Krull Dimension. Using this, we prove the following lemma. First, for any L-formula $\varphi(x; y)$ and any finite φ-type $p(x)$, let $\mathrm{MR}(p) = \mathrm{MR}(\bigwedge p)$ and $\mathrm{Md}(p) = \mathrm{Md}(\bigwedge p)$.

Lemma 4.3.6. *If $\varphi(x; y)$ is an L-formula, then*

1. *If $B \subseteq \mathcal{U}_y$ is finite, $B_0 \subseteq B$, $p \in S_\varphi(B)$, and, for all $B_1 \subseteq B$ with $B_0 \subset B_1$ and $|B_1 \setminus B_0| = 1$, we have that* $\mathrm{Md}(p|_{B_1}) = \mathrm{Md}(p|_{B_0})$ *and* $\mathrm{MR}(p|_{B_1}) = \mathrm{MR}(p|_{B_0})$, *then* $\mathrm{Md}(p) = \mathrm{Md}(p|_{B_0})$ *and* $\mathrm{MR}(p) = \mathrm{MR}(p|_{B_0})$.
2. *There exists $g : (|x| + 1) \to \omega$ such that, for all finite φ-types p, there exists $p_0 \subseteq p$ with $|p_0| \leq g(\mathrm{MR}(p))$ such that* $\mathrm{MR}(p_0) = \mathrm{MR}(p)$.
3. *There exists $K < \omega$ such that, for all finite φ-types p,* $\mathrm{Md}(p) \leq K$.

Proof. (1): By induction on $m = |B \setminus B_0|$. If $m = 0$, this is trivial, so suppose $m \geq 1$. Write $B = B_1 \cup B_2$ for $B_1, B_2 \subseteq B$ with $|B_1 \setminus B_0| < m$ and $|B_2 \setminus B_0| < m$. By induction,

$$\mathrm{Md}(p|_{B_t}) = \mathrm{Md}(p|_{B_0}) \text{ and } \mathrm{MR}(p|_{B_t}) = \mathrm{MR}(p|_{B_0})$$

for both $t = 1, 2$. If $\mathrm{MR}(p) < \mathrm{MR}(p|_{B_0})$, then by Theorem 4.3.5 (5), $\mathrm{Md}(p|_{B_0}) \geq \mathrm{Md}(p|_{B_1}) + \mathrm{Md}(p|_{B_2})$, contrary to assumption. Hence $\mathrm{MR}(p) = \mathrm{MR}(p|_{B_0})$. Similarly, Theorem 4.3.5 (4) shows us $\mathrm{Md}(p) = \mathrm{Md}(p|_{B_0})$.

(2): By Theorem 4.3.5 (6), for each $\ell < \omega$, there exists $k_\ell < \omega$ such that, for all φ-types p with $|p| = \ell$, $\mathrm{Md}(p) \leq k_\ell$. Define $g(|x|) = 0$ and define $g(i) = g(i + 1) + k_{g(i+1)} + 1$ for each $0 \leq i < |x|$. We claim this works.

By induction on $r = \mathrm{MR}(p)$. If $r = \mathrm{MR}(\emptyset)$ (where $\emptyset$ is the empty φ-type), then $p_0 = \emptyset$ witnesses the type. So assume $r < \mathrm{MR}(\emptyset)$. Choose $p' \subseteq p$ with $\mathrm{MR}(p') > r$ and $\mathrm{MR}(p')$ is minimal such. By induction, there exists $p_0 \subseteq p'$ with $|p_0| \leq g(r+1)$ and $\mathrm{MR}(p_0) = \mathrm{MR}(p')$.

Case 1. There exists $b \in (\mathrm{dom}(p) \setminus \mathrm{dom}(p_0))$ with $\mathrm{MR}(p_0 \cup p|_{\{b\}}) < \mathrm{MR}(p_0)$.

In this case, $\mathrm{MR}(p) = \mathrm{MR}(p_0 \cup p|_{\{b\}})$ by choice of p'. Moreover,

$$|p_0 \cup p|_{\{b\}}| = |p_0| + 1 \leq g(r+1) + 1 \leq g(r).$$

Case 2. For all $b \in (\mathrm{dom}(p) \setminus \mathrm{dom}(p_0))$, $\mathrm{MR}(p_0 \cup p|_{\{b\}}) = \mathrm{MR}(p_0)$.

If there exists $b \in (\mathrm{dom}(p) \setminus \mathrm{dom}(p_0))$ such that $\mathrm{Md}(p_0 \cup p|_{\{b\}}) < \mathrm{Md}(p_0)$. Then, replace p_0 by $p_0 \cup p|_{\{b\}}$ and repeat this at most $\mathrm{Md}(p_0) < k_{g(r+1)}$ steps to produce $p_1 \subseteq p$ with

$$|p_1| \leq |p_0| + k_{g(r+1)}, \mathrm{MR}(p_1) = \mathrm{MR}(p_0),$$

and $\mathrm{Md}(p_1)$ is minimal. Suppose now that, for all $b \in (\mathrm{dom}(p) \setminus \mathrm{dom}(p_1))$, $\mathrm{MR}(p_1 \cup p|_{\{b\}}) = \mathrm{MR}(p_1)$. Then, by choice of p_1, $\mathrm{Md}(p_1 \cup p|_{\{b\}}) = \mathrm{Md}(p_1)$. By (1) of this lemma, $\mathrm{MR}(p) = \mathrm{MR}(p_1)$. Notice now that

$$|p_1 \cup p|_{\{b\}}| \leq |p_0| + k_{g(r+1)} + 1 \leq g(r),$$

as desired.

(3): Let $K = \max\{k_\ell : \ell \leq g(0)\}$ for g and k_ℓ as in (2) above. For any finite φ-type p, by (2) of this lemma, there exists $p_0 \subseteq p$ with $|p_0| \leq g(\mathrm{MR}(p)) \leq g(0)$ such that $\mathrm{MR}(p_0) = \mathrm{MR}(p)$. Hence, $\mathrm{Md}(p_0) \leq k_{|p_0|} \leq K$ and $\mathrm{Md}(p) \leq \mathrm{Md}(p_0)$, so $\mathrm{Md}(p) \leq K$. ☐

Proof of Theorem 4.3.2. Fix $\varphi(x; y)$ and let $n = |x|$. For any finite φ-type p and $B \supseteq \mathrm{dom}(p)$, recall the definition

$$S_\varphi(B) \cap [p] = \{q \in S_\varphi(B) : q \supseteq p\}.$$

Let K be as in Lemma 4.3.6 (3) above, so $\mathrm{Md}(p) \leq K$ for all finite φ-types p. We claim that, for all finite φ-types p and all $B \supseteq \mathrm{dom}(p)$,

$$|S_\varphi(B) \cap [p]| \leq K^{\mathrm{MR}(p)}\mathrm{Md}(p)(|B \setminus \mathrm{dom}(p)| + 1)^{\mathrm{MR}(p)}.$$

Applying this to $p = \emptyset$, we get that

$$|S_\varphi(B)| \leq K^{\mathrm{MR}(p)+1}(|B| + 1)^{\mathrm{MR}(p)} \leq K^{n+1}(|B| + 1)^n,$$

which shows that φ has VC-density $\leq n = |x|$ (since K^{n+1} is constant).

We prove this by induction. For the base case, consider p such that $\mathrm{dom}(p) = B$. That is, $p \in S_\varphi(B)$. Then, clearly $S_\varphi(B) \cap [p] = \{p\}$, therefore $|S_\varphi(B)| = 1$, as desired.

In general, fix p with $\mathrm{dom}(p) \subsetneq B$. Let $r = \mathrm{MR}(p)$, $\ell = \mathrm{Md}(p)$, and $m = |B \setminus \mathrm{dom}(p)| + 1$. So we aim to show that

$$|S_\varphi(B) \cap [p]| \leq \ell K^r m^r.$$

Case 1. There exists $b \in (B \setminus \mathrm{dom}(p))$ such that

$$\mathrm{MR}(p \cup \{\varphi(x;b)\}) = \mathrm{MR}(p \cup \{\neg\varphi(x;b)\}) = \mathrm{MR}(p) = r.$$

In this case, set $p_t = p \cup \{\varphi(x;b)^t\}$ for $t < 2$. By Theorem 4.3.5 (5), $\ell \geq \mathrm{Md}(p_0) + \mathrm{Md}(p_1)$. By induction hypothesis

$$|S_\varphi(B) \cap [p_t]| \leq K^r \mathrm{Md}(p_t)(m-1)^r$$

for both $t < 2$. On the other hand,

$$S_\varphi(B) \cap [p] = (S_\varphi(B) \cap [p_0]) \cup (S_\varphi(B) \cap [p_1]),$$

therefore,

$$|S_\varphi(B) \cap [p]| \leq K^r \ell (m-1)^r \leq \ell K^r m^r.$$

Case 2. For all $b \in (B \setminus \mathrm{dom}(p))$, $\mathrm{MR}(p \cup \{\varphi(x;b)^t\}) < r$ for some $t < 2$.

In this case, fix any such b, and let $p_0 = p \cup \{\varphi(x;b)^t\}$ and $p_1 = p \cup \{\varphi(x;b)^{1-t}\}$. By Theorem 4.3.5 (3), $\mathrm{MR}(p_1) = r$. By induction hypothesis,

$$|S_\varphi(B) \cap [p_0]| \leq K^{r-1}\mathrm{Md}(p_0)(m-1)^{r-1} \leq K^r(m-1)^{r-1}, \text{ and}$$
$$|S_\varphi(B) \cap [p_1]| \leq K^r \mathrm{Md}(p_1)(m-1)^r \leq K^r \ell (m-1)^r.$$

As in Case 1, we get

$$|S_\varphi(B) \cap [p]| \leq K^r(m-1)^{r-1} + K^r \ell (m-1)^r \leq \ell K^r m^r.$$

This concludes the proof. □

In the original paper, the statement of Theorem 3.1 of [**2**] is stronger. There they show that, for any formula $\varphi(x;y)$ in a theory without the finite cover property, the VC-density of φ is bounded by the Shelah $\aleph_0$-rank of φ (similar to the definition of Shelah 2-rank given above).

We say that a formula $\varphi(x;y)$ has the *finite cover property* if, for all $n < \omega$, there exists $b_0, ..., b_{n-1} \in \mathcal{U}_y$ such that

$$\models \neg\exists x \left(\bigwedge_{i<n} \varphi(x;b_i) \right)$$

but, for all $\ell < n$,

$$\models \exists x \left(\bigwedge_{i<n, i\neq \ell} \varphi(x;b_i) \right).$$

We say T has the *finite cover property* if there is a formula with the finite cover property.

THEOREM 4.3.7 (Theorem II.4.2 of [**20**]). *If T does not have the finite cover property, then T is stable.*

EXERCISE 4.3.8. Prove Theorem 4.3.7. Hint: Suppose $\varphi(x; y)$ has the order property, witnessed by $\langle b_j : j < \omega \rangle$. Let

$$\psi(x; y_1, y_2, y_3, y_4) = [\varphi(x; y_1) \leftrightarrow \neg\varphi(x; y_2)] \wedge [\varphi(x; y_3) \leftrightarrow \varphi(x; y_4)].$$

Fix n and, for each $k < n$, let

$$c_k = \langle b_0, b_n, b_k, b_{k+1} \rangle.$$

Show that $\{\psi(x; c_k) : k < n\}$ is inconsistent but, for each $\ell < n$, $\{\psi(x; c_k) : k < n, k \neq \ell\}$ is consistent. Hence ψ has the finite cover property.

EXERCISE 4.3.9. Show that if T is strongly minimal, then T does not have the finite cover property. For simplicity, show that all formulas $\varphi(x; y)$ with x of the home sort do not have the finite cover property (sufficiency of a single variable is shown in Theorem II.4.4 of [**20**]).

See [**2**] for more details.

4. op-Dimension

In this section, we talk about op-dimension. Since much of the work here involves simple modifications of 2-rank and dp-rank, we will leave some of the proofs as exercises.

For each $C \subseteq \mathbb{R}$, consider the equivalence relation $\sim_C$ on $\mathbb{Q}$ defined as follows: For $i, j \in \mathbb{Q}$, $i \sim_C j$ if and only if, for all $r \in C$, $r < i$ if and only if $r < j$ and $r = i$ if and only if $r = j$. That is, i and j have the same $\{x < y, x = y\}$-type over C. For example, if $C = \{-2, \pi\}$, then $\sim_C$ has 4 classes, namely $(-\infty, -2)$, $\{-2\}$, $(-2, \pi)$, and (π, ∞).

DEFINITION 4.4.1. Let $\varphi(x; y)$ be a formula, $p(x)$ a partial type, and $n < \omega$. We say that φ has *op-dimension* $\leq n$ over $p(x)$ if, for all indiscernible sequences $\overline{b} = \langle b_i : i \in \mathbb{Q} \rangle$ and all $a \models p$, there exists $C \in \binom{\mathbb{R}}{n}$ such that, for all $\sim_C$ classes $D \subseteq \mathbb{Q}$, either $\{i \in D : \models \varphi(a; b_i)\}$ is finite or $\{i \in D : \models \neg\varphi(a; b_i)\}$ is finite.

Compare this to the definition of (local) dp-rank, Definition 4.2.3 above. There is a subtle difference between the two notions, which we will explore in the following exercise.

Exercise 4.4.2. Let $L = \{E_1, E_2\}$ where E_1 and E_2 are two binary relation symbols. Let T be the L-theory which says that each E_i is an equivalence relation with infinitely many infinite classes for both $i = 1, 2$. Furthermore, T says that each E_1-class intersects each E_2-class with infinitely many elements. Check that T is a complete theory that admits quantifier elimination. Let $\varphi(x; y_1, y_2) = E_1(x, y_1) \wedge E_2(x, y_2)$. Show that φ has dp-rank 2 and op-dimension 0. Generalize this to find a theory with a formula that has arbitrarily large dp-rank and op-dimension 0.

Exercise 4.4.3. Show that $\varphi(x; y)$ is stable if and only if φ has op-dimension 0. Hint: Consider a modification of the proof Proposition 3.1.3 above, when we assume $\varphi(x; y)$ is stable instead of NIP. Use the order property characterization of stability.

Exercise 4.4.4. Let $\varphi(x; y)$ be any formula and $p(x)$ any partial type. Show that the op-rank of φ over p is at most the dp-rank of φ over p.

We turn from this toward a global definition of op-dimension. Instead of ICT-patterns, we will use IRD-patterns (first introduced by Shelah in [**20**], where he comments that "IRD" stands for "independent orders").

Let $p(x)$ be any partial type, let α and β be ordinals, let $\overline{\varphi} = \langle \varphi_i(x; y_i) : i < \alpha \rangle$ be a set of functions, and let $\overline{b} = \langle b_{i,j} : i < \alpha, j < \beta \rangle$ where $b_{i,j} \in \mathbb{U}_{y_i}$. We say that $\langle \overline{\varphi}, \overline{b} \rangle$ is a *IRD-pattern in p of depth α and length β* if, for all functions $f : \alpha \to \beta$, the following type is consistent:

$$p(x) \cup \{\varphi_i(x; b_{i,j}) : i < \alpha, j < f(i)\} \cup \{\neg\varphi_i(x; b_{i,j}) : i < \alpha, f(i) \leq j < \beta\}. \tag{4.4}$$

Definition 4.4.5. We say that $p(x)$ has *op-dimension* $\leq \alpha$ if there exists no IRD-pattern in p of depth $\alpha + 1$ and length ω, and we denote this by $\mathrm{opD}(p) \leq \alpha$.

We can transform IRD-patterns into ICT-patterns as follows. Suppose that $\overline{\varphi} = \langle \varphi_i(x; y_i) : i < \alpha \rangle$ and $\overline{b} = \langle b_{i,j} : i < \alpha, j < \omega \rangle$ form an IRD-pattern of depth α and length ω in a partial type $p(x)$. Then, for each $i < \alpha$, let

$$\psi_i(x; y_i^0, y_i^1) = \varphi(x; y_i^0) \leftrightarrow \neg\varphi(x; y_i^1)$$

and, for each $j < \omega$, let $c_{i,j} = \langle b_{i,2j}, b_{i,2j+1} \rangle$.

EXERCISE 4.4.6. Show that $\langle\overline{\psi}, \overline{c}\rangle$ is an ICT-pattern in $p(x)$ of depth α and length β. This shows, in particular, that the op-dimension of $p(x)$ is bounded above by the dp-rank of $p(x)$.

We show that local and global op-dimension coincide in the natural way.

THEOREM 4.4.7 (Theorem 1.21 of [**9**]). *If $p(x)$ is any partial type, the op-dimension of $p(x)$ is $\leq n$ if and only if, for all formulas $\varphi(x; y)$, the op-dimension of $\varphi(x; y)$ over p is $\leq n$.*

Actually, you show it.

EXERCISE 4.4.8. Prove Theorem 4.4.7. Take the proof of Theorem 4.2.11 and modify it to work for IRD-patterns. If $\varphi(x; y)$ has op-dimension $> n$ witnessed by $\overline{b} = \langle b_i : i \in \mathbb{Q}\rangle$ and $a \models p$, build an IRD-pattern with $\pm\varphi$ and parts of $\overline{b}$.

We can also build op-dimension up in terms of a 2-rank-esque construction. This was originally done in [**9**].

Let $\varphi(x; y)$ be a formula, $p(x)$ a partial type, and $n < \omega$. The *nth op-Rank* of p, denoted $\mathrm{opR}_{n,\varphi}(p)$, is an ordinal-valued function of partial types defined inductively as follows:

(1) $\mathrm{opR}_{n,\varphi}(p) \geq 0$.
(2) $\mathrm{opR}_{n,\varphi}(p) \geq \delta$ for a limit ordinal δ if $\mathrm{opR}_{n,\varphi}(p) \geq \alpha$ for all $\alpha < \delta$.
(3) $\mathrm{opR}_{n,\varphi}(p) \geq \alpha + 1$ if, for all finite subtypes $q \subseteq p$, there exists $b_0, ..., b_{n-1} \in \mathcal{U}_y$ such that, for each $s \in {}^n2$,

$$\mathrm{opR}_{n,\varphi}\left(q(x) \cup \left\{\varphi(x; b_i)^{s(i)} : i < n\right\}\right) \geq \alpha.$$

THEOREM 4.4.9. *Fix a partial type $p(x)$ and $n < \omega$. Then $p(x)$ has op-dimension $\leq n$ if and only if, for all formulas $\varphi(x; y)$, $\mathrm{opR}_{n+1,\varphi}(p) < \omega$.*

PROOF. First, suppose that $\mathrm{opR}_{n,\varphi}(p) \geq \omega$. Check that this implies there exists, for each $m < \omega$ and each $\tau : m \to {}^n2$, $\langle b_{\tau,i} : i < n\rangle \in \mathcal{U}_y^n$ such that, for all $\sigma : \omega \to {}^n2$, the type

$$p(x) \cup \left\{\varphi(x; b_{\sigma|_\ell,i})^{[\sigma(\ell)](i)} : \ell < \omega, i < n\right\}$$

is consistent. Compare this to the proof of Theorem 4.1.6, (2) $\Rightarrow$ (1) outlined in exercises in that section. Just as in this proof, we construct, for each $k < \omega$,

$$\overline{b} = \langle\langle b_{\tau,i} : i < n\rangle : \tau \in {}^nk\rangle$$

such that, for each $\sigma \in {}^n k$, there exists $a_\sigma \models p$ such that, for all $\tau \in {}^n k$ and $i < n$,

$$\models \varphi(a_\sigma, b_{\tau,i}) \text{ if and only if } \sigma(i) < \tau(i).$$

Then, check that $\varphi(x; y)$ together with $\bar{b}$ form an IRD-pattern of depth n and length k in $p(x)$. Hence, $\mathrm{opD}(p) \geq n$.

Conversely, suppose $\mathrm{opD}(p) \geq n$. By Theorem 4.4.7, there exists $\varphi(x; y)$ with op-dimension $\geq n$ in $p(x)$. Therefore, there exists $\langle b_i : i \in \mathbb{Q}\rangle$ and $a \models p$ such that, for minimal $C \subseteq \mathbb{R}$ so that, for all $\sim_C$-classes $D \subseteq \mathbb{Q}$, either $\{q \in D :\models \varphi(a; b_q)\}$ is finite or $\{q \in D :\models \neg\varphi(a; b_q)\}$ is finite, we have that $|C| \geq n$. By removing finitely many points, we may assume that each $\sim_C$ class D has constant truth value on $\varphi(a; b_q)$, hence this truth value alternates. Let $D_0 < D_1 < ... < D_n$ list off the first $n + 1$ classes. Without loss of generality, suppose $\models \neg\varphi(a; b_q)$ for all $q \in D_0$. Now we use this to construct the $q_{\tau,i} \in \mathbb{Q}$ for $\tau : m \to {}^n 2$ for $m < \omega$ and $i < n$ as in the first half of the proof, so that $b_{q_{\tau,i}}$ witness that $\mathrm{opR}_{n,\varphi}(p) \geq \omega$.

Fix $m < \omega$ and suppose that $q_{\tau,i} \in \mathbb{Q}$ has been constructed for all $m' < m$ and $\tau : m' \to {}^n 2$ such that,

(1) for each $i < n$ and τ, $q_{\tau,i} \in D_i \cup D_{i+1}$,
(2) for each $i < n$, $0 < j < m$, and $\tau : m \to {}^n 2$, $q_{\tau|_j,i} \in D_i$ if and only if $[\tau(j)](i) \equiv i \pmod 2$ (therefore $\models \varphi(a; b_{q_{\tau|_j,i}})^{[\tau(j)](i)}$).
(3) if $i < n$, $0 < j_0 < j_1 < m$ and $\tau : m \to {}^n 2$ is such that $[\tau(j_1)](i) = [\tau(j_0)](i)$, then $q_{\tau|_{j_1},i}$ is between $q_{\tau|_{j_0},i}$ and the cut between D_i and D_{i+1}.

By density, we can continue this construction. By indiscernibility, each one is consistent with any change at the top level. The details are left to the reader to check. This concludes the construction. □

Finally, we show that op-dimension is subadditive. For a tuple a and a set A, let $\mathrm{opD}(a/A) = \mathrm{opD}(\mathrm{tp}(a/A))$.

Theorem 4.4.10 (Subadditivity of op-dimension, Theorem 2.2 of [**9**]). *For tuples a and b and a set A,*

$$\mathrm{opD}(a, b/A) \leq \mathrm{opD}(a/A) + \mathrm{opD}(b/A \cup \{a\}).$$

To prove this, we use a notion called "almost mutually indiscernible sequences."

Fix an ordinal α, a collection of sequences

$$\mathcal{J} = \{\langle b_{j,i} : j \in J_i\rangle : i < \alpha\},$$

$K < \omega$, and a set of formulas

$$\Delta(y_{k,i})_{k<K,i<\alpha}.$$

We say that $\mathcal{J}$ is Δ-*mutually-indiscernible* if, for all sequences

$$j_{0,i} < ... < j_{K-1,i} \text{ and } \ell_{0,i} < ... < \ell_{K-1,i}$$

from J_i for each $i < \alpha$ and for all $\delta \in \Delta$, we have that

$$\models \delta(b_{j_{k,i},i})_{k<K,i<\alpha} \leftrightarrow \delta(b_{\ell_{k,i},i})_{k<K,i<\alpha}.$$

Notice that this notion coincides with our original definition of mutually indiscernible sequences. That is, $\mathcal{J}$ is mutually indiscernible if and only if, for all Δ, $\mathcal{J}$ is Δ-mutually-indiscernible.

Definition 4.4.11. Fix a set of parameters A. We say that $\mathcal{J}$ is *almost mutually indiscernible* over A if, for each $\delta(y_{k,i})_{k<K,i<\alpha}$ over A, there exists a finite $J'_i \subseteq J_i$ for each $i < \alpha$ such that

$$\{\langle b_{j,i} : j \in (J_i \setminus J'_i)\rangle : i < \alpha\}$$

is δ-mutually indiscernible.

Exercise 4.4.12. Show that limit types, as in Definition 4.2.13, makes sense for *almost* mutually indiscernible sequences $\mathcal{J}$. That is, show that Lemma 4.2.12 holds for almost indiscernible sequences. This is a simple modification of the proof of Lemma 4.2.12, noting that the "almost" condition means, after removing finitely much information, it is actually δ-mutually-indiscernible.

As in the case for dp-rank, we can characterize op-dimension in terms of almost mutual indiscernibility.

Proposition 4.4.13 (Proposition 2.1 of [**9**]). *For a partial type $p(x)$ over A, the following are equivalent:*

(i) *p has op-dimension $\leq n$.*
(ii) *For all $k < \omega$, all $\{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i < k\}$ almost mutually indiscernible over A, and $a \models p$, there exists $I \subseteq k$ with $|I| \geq k - n$ such that $\{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i \in I\}$ is almost mutually indiscernible over $A \cup \{a\}$.*

Exercise 4.4.14. Prove Proposition 4.4.13. Modify the proofs of Proposition 4.2.15 and Proposition 4.2.10 above, using Exercise 4.4.12.

Proof of Theorem 4.4.10. Suppose that $\mathrm{opD}(a/A) = n_1$ and $\mathrm{opD}(b/A \cup \{a\}) = n_2$. Fix any $k < \omega$ and almost mutually indiscernible sequence $\{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i < k\}$ over A. By Proposition 4.4.13, there exists $I_1 \subseteq k$ with $|I_1| = k - n_1$ such that $\{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i \in I_1\}$ is almost mutually indiscernible over $A \cup \{a\}$. Again by Proposition 4.4.13, there exists $I_2 \subseteq I_1$ with $|I_2| = k - n_1 - n_2$ such that $\{\langle b_{i,j} : j \in \mathbb{Z}\rangle : i \in I_2\}$ is almost mutually indiscernible over $A \cup \{a, b\}$.

By Proposition 4.4.13 in the other direction (since $\mathcal{J}$ was arbitrary), $\mathrm{opD}(a, b/A) \leq n_1 + n_2$. □

If X is a (type-)definable set, then define $\mathrm{opD}(X) = \mathrm{opD}(x \in X)$ (i.e., the op-dimension of X is the op-dimension of the partial type $x \in X$).

EXERCISE 4.4.15. Suppose X and Y are definable sets. Prove the following:

(1) If $\sigma : X \to Y$ is an injection, then $\mathrm{opD}(X) \leq \mathrm{opD}(Y)$.
(2) If X and Y are of the same sort, then
$$\mathrm{opD}(X \cup Y) = \max\{\mathrm{opD}(X), \mathrm{opD}(Y)\}.$$

COROLLARY 4.4.16. *If T is o-minimal (expanding a dense linear order) and X is any definable set, then the op-dimension of X equals the o-minimal dimension of X.*

PROOF. It is a well known fact that o-minimal theories admit cell decomposition. In particular, X has o-minimal dimension n if and only if $X = \bigcup_{i<k} Y_i$ for definable sets Y_i and there exists, for each $i < k$, an injective projection $\pi_i : Y_i \to \mathcal{U}^{m_i}$ such that $\pi_i(Y)$ is open (in the o-minimal topology) and $n = \max\{m_i : i < k\}$. On the one hand, by Theorem 4.4.10, $\mathrm{opD}(\mathcal{U}^n) \leq n$. On the other hand, if $m_i = n$, then the formulas $\psi_j(x; y) = \pi_i(x) <_j \pi_i(y)$ for $j < n$ form an IRD-pattern of depth n in Y_i (since $\pi_i(Y_i)$ is open, it contains an open box B). Therefore, $n \leq \mathrm{opD}(Y_i) \leq \mathrm{opD}(\mathcal{U}^n) \leq n$ (by Exercise 4.4.15 (1)). Hence, $\mathrm{opD}(X) = n$ (by Exercise 4.4.15 (2)). □

EXERCISE 4.4.17. Modifying the proof above, show that if T is o-minimal (expanding a linear order) and X is any definable set, then the op-dimension of X, the o-minimal dimension of X, and the dp-rank of X coincide.

5. The Splitting Conjecture

In this section, we draw attention to the relationship between independence dimension and UDTFS-rank. We culimate in a conjecture about UDTFS.

Let $\varphi(x; y)$ be a formula and let $\Delta(y; z)$ be a set of formulas. Let $B \subseteq \mathcal{U}_y$, $p \in S_\varphi(B)$, and $C \subseteq \mathcal{U}_z$.

DEFINITION 4.5.1. We say that $p(x)$ *does not Δ-split over C* if, for all $b_1, b_2 \in B$, if
$$\models \delta(b_1; c) \leftrightarrow \delta(b_2; c)$$
for all $c \in C$ and $\delta(y; z) \in \Delta(y; z)$, then $\varphi(x; b_1) \in p(x)$ if and only if $\varphi(x; b_2) \in p(x)$.

Notice the relationship to splitting as defined in Remark 3.2.2 above. First thing to note is that UDTFS is equivalent to non-splitting in the following sense:

PROPOSITION 4.5.2. *Fix a formula $\varphi(x;y)$. Then φ has UDTFS if and only if there exists a finite $\Delta(y;z_1,...,z_n)$ and $K<\omega$ such that, for all finite $B\subseteq\mathcal{U}_y$ and all $p\in S_\varphi(B)$, there exists $C\subseteq B^n$ with $|C|\leq K$ so that p does not Δ-split over C.*

Moreover, if the right-hand condition holds, the UDTFS-rank of φ is $\leq Kn$.

PROOF. ($\Rightarrow$): Suppose $\varphi(x;y)$ has UDTFS witnessed by

$$\Psi(y;z_1,...,z_n)=\{\psi_j(y;z_1,...,z_n):j<N\}.$$

Then, for all finite $B\subseteq\mathcal{U}_y$ and all $p\in S_\varphi(B)$, there exists $j<N$ and $c\in B^n$ so that $\psi_j(y;c)$ defines $p(x)$. In particular, p does not Ψ-split over $\{c\}$ (so $K=1$).

($\Leftarrow$): Suppose $\Delta(y;z_1,...,z_n)$ and $K<\omega$ satisfy the right-hand side. $S\subseteq {}^{(\Delta\times K)}2$. For each such S, define

$$\psi_S(y;w_0,...,w_{K-1})=\bigvee\{\bigwedge_{\ell<K,\delta\in\Delta}\delta(y;w_\ell)^{s(\delta,\ell)}:s\in S\}.$$

Fix $B\subseteq\mathcal{U}_y$ finite and $p\in S_\varphi(B)$ and let $C\subseteq B^n$ with $|C|\leq K$ be so that p does not Δ-split over C. Let $C=\{c_\ell:\ell<K\}$ enumerate C (with possible repetitions). Notice that $c\in B^{Kn}$. Now define S to be

$$\left\{s\in {}^{(\Delta\times K)}2:(\forall b\in B)\left[\models\bigwedge_{\ell<K,\delta\in\Delta}\delta(b;c_\ell)^{s(\delta,\ell)}\Rightarrow\varphi(x;b)\in p(x)\right]\right\}.$$

It is clear that $\psi_S(y;c)$ defines $p(x)$. □

The splitting conjecture involves making a guess at what Δ and K should be. Fix a formula $\varphi(x;y)$ and we guess $K=1$ and, for some $n<\omega$,

$$\Delta_{n,\varphi}(y;z_1,...,z_n)=\left\{\exists x\left(\bigwedge_{i<n}\varphi(x;z_i)^{s(i)}\wedge\varphi(x;y)^{s(n)}\right):s\in {}^{n+1}2\right\}.\tag{4.5}$$

By Proposition 4.5.2, if, for all finite $B\subseteq\mathcal{U}_y$ and all $p\in S_\varphi(B)$, there exists $c\in B^n$ so that p does not $\Delta_{n,\varphi}$-split over $\{c\}$, then φ has UDTFS-rank $\leq n$.

EXERCISE 4.5.3. Suppose $\varphi(x;y)$ has independence dimension ≤ 1 (that is, for all $b,c\in\mathcal{U}_y$, one of the following holds: $\varphi(x;b)\vdash\varphi(x;c)$,

$\varphi(x;c) \vdash \varphi(x;b)$, $\varphi(x;b) \wedge \varphi(x;c)$ is inconsistent, or $\varphi(x;b) \vee \varphi(x;c)$ is equivalent to $x = x$). Show that, for all finite $B \subseteq \mathcal{U}_y$ and all $p \in S_\varphi(B)$, there exists $c \in B$ so that p does not $\Delta_{1,\varphi}$-split over $\{c\}$. This establishes, in particular, that all φ with independence dimension ≤ 1 have UDTFS-rank ≤ 1.

Extrapolating, here is a wild guess at the form of the definition for the UDTFS Conjecture.

Open Question 4.5.4 (The Splitting Conjecture). *Is it true that a formula $\varphi(x;y)$ has NIP if and only if, there exists $n < \omega$ such that, for all finite $B \subseteq \mathcal{U}_y$ and all $p \in S_\varphi(B)$, there exists $c \in B^n$ so that p does not $\Delta_{n,\varphi}$-split over $\{c\}$.*

Moreover, is it possible that n is simply the independence dimension of φ?

Certainly this holds for independence dimension 1. The splitting conjecture also holds for φ with VC-density < 2 (see the proof of Theorem 3.3.14 above). It also holds for stable formulas.

Theorem 4.5.5. *A formula $\varphi(x;y)$ is stable if and only if there exists $n < \omega$ such that, for all $B \subseteq \mathcal{U}_y$ and all $p \in S_\varphi(B)$, there exists $c \in B^n$ such that p does not $\Delta_{n,\varphi}$-split over $\{c\}$.*

For the proof of Theorem 4.5.5, consider the Stable Formula Theorem, Theorem 4.1.6 above. Notice that (4) is remarkably close to UDTFS; the only difference is that for UDTFS we demand that B is finite. This is some of the reason why we make the UDTFS conjecture. We think of the analogy "Stable" is to "all sets" as "NIP" is to "finite sets." Indeed condition (3) classifies NIP formulas when restrict to finite B (and allow a polynomial bound in $|B|$).

We are now ready to prove Theorem 4.5.5. Notice that Theorem 4.1.6 gives that φ is stable if and only if there is some Δ for which, for all $B \subseteq \mathcal{U}_y$ and $p \in S_\varphi(B)$, there exists $c \in B^n$ such that p does not Δ-split over $\{c\}$. The proof of this is exactly as the proof of Proposition 4.5.2. However, we will show that $\Delta = \Delta_{\varphi,n}$ suffices (for some n). Above we omitted the proof of Theorem 4.1.6 (2) $\Rightarrow$ (4). This is because this is an immediate consequence of the proof of Theorem 4.5.5 (a diligent reader will check to make sure that there are no circular arguments occuring here; there are not).

Proof of Theorem 4.5.5. As noted above, Theorem 4.1.6 (4) $\Rightarrow$ (1) provides a proof of one direction. So assume that $\varphi(x;y)$ is stable. By compactness and Theorem 4.1.6 (1) $\Rightarrow$ (2), there exists $N < \omega$ such that, there are no sequences $\langle a_i : i < N \rangle$ from $\mathcal{U}_x$ and

$\langle b_j : j < N\rangle$ from $\mathcal{U}_y$ such that, for all $i, j < N$, $\models \varphi(a_i; b_j)$ if and only if $i < j$. Moreover, since φ is stable if and only if $\neg\varphi$ is stable, we may also assume that there exists no sequences $\langle a'_i : i < N\rangle$ from $\mathcal{U}_x$ and $\langle b'_j : j < N\rangle$ from $\mathcal{U}_y$ such that, for all $i, j < N$, $\models \neg\varphi(a'_i; b'_j)$ if and only if $i < j$ (for the same N). We claim that $\Delta = \Delta_{2^{2N+1}-1,\varphi}$ works.

For each $n < 2N$, let

$$X_n = \{0, 1\} \times \mathcal{P}(\{0, 1, ..., n-1\})$$

and let $K_n = |X_n| = 2^{n+1}$. Let $f^n : K_n \to X_n$ be any bijection (to put an ordering on X_n). For any $k < K_n$, let $f^n(k) = \langle f^n_0(k), f^n_1(k)\rangle$ where $f^n_0(k) < 2$ and $f^n_1(k) \in \mathcal{P}(n)$. Set $K = \sum_{n<2N} K_n = 2^{2N+1} - 1$.

Fix $B \subseteq \mathcal{U}_y$ small and $p \in S_\varphi(B)$. For simplicity, choose $a \in \mathcal{U}_x$ such that $a \models p$. We may assume that there exists $b_0, b_1 \in B$ such that $\models \neg\varphi(a; b_0)$ and $\models \varphi(a; b_1)$ since, otherwise, p does not $\Delta_{K,\varphi}$-split over $\{c\}$ for any choice of $c \in B^K$. We now construct, for each $n < 2N$ and $k < K_n$,

(1) $I^n_k \subseteq K_n$,
(2) $q^n_k(x_0, ..., x_{n-1})$ a partial type over B, and
(3) $b^n_k \in B$ satisfying $\models \neg\varphi(a; b^n_k)^{f^n_0(m)}$

Moreover, we construct, for each $n < 2N$,

(4) $I^n \subseteq K_n$
(5) $q^n(x_0, ..., x_n)$ a partial type over B.

The construction proceeds as follows:

For $n = 0$ and $k < K_0 = 2$, let $b^0_k \in B$ be such that $\models \neg\varphi(a; b^0_k)^{f^0_0(m)}$ (which exists by hypothesis). Let $I^0_0 = I^0_1 = \{0, 1\}$ and let $q^0_0 = q^0_1 = \emptyset$. Suppose that $n \geq 0$ and, for all $n' \leq n$, $b^{n'}_k$, $I^{n'}_k$, and $q^{n'}_k(x_0, ..., x_{n-1})$ have been constructed for all $k < K_{n'}$. Define

$$q^n(x_0, ..., x_n) = q^n_{K_n-1}(x_0, ..., x_{n-1}) \cup \left\{\neg\varphi(x_n; b^{n'}_k)^{f^{n'}_0(k)} : n' \leq n, k < K_{n'}\right\}.$$

That is, q^n is $q^n_{K_n-1}(x_0, ..., x_{n-1})$ (which is consistent) together with the type

$$\mathrm{tp}_\varphi\left(a/\left\{b^{n'}_k : n' \leq n, k < K_{n'}\right\}\right)(x_n).$$

Thus, q^n is clearly consistent. Define $I^n = I^n_{K_n-1}$.

Suppose $n \geq 1$, $k < K_n$, q^{n-1} is defined, and $I^n_{k'}$, $q^n_{k'}$, and $b^n_{k'}$ are defined for all $k' < k$. In the case where $k = 0$, let $I^n_{-1} = I^{n-1}$ and $q^n_{-1} = q^{n-1}$ (so we set them equal to the final objects from the previous level of the construction). Let $t = f^n_0(k)$ and $W = f^n_1(k)$, so $f^n(k) = \langle t, W\rangle \in X_n$.

Case 1. There exists $b \in B$ such that

(i) $\models \neg\varphi(a;b)^t$, and
(ii) $q^n_{k-1}(x_0, ..., x_{n-1}) \cup \{\varphi(x_i;b)^t : i \in W\}$ is consistent.

In this case, let $b^n_k = b$ be such a witness, let $I^n_k = I^n_{k-1} \cup \{k\}$, and let

$$q^n_k(x_0, ..., x_{n-1}) = q^n_{k-1}(x_0, ..., x_{n-1}) \cup \left\{\varphi(x_i;b)^t : i \in W\right\}.$$

Case 2. There is no $b \in B$ satisfying (i) and (ii). In this case, let $I^n_k = I^n_{k-1}$, $q^n_k = q^n_{k-1}$, and choose $b^n_k = b \in B$ arbitrary so that (i) holds. By assumption, this is always possible.

In either case, notice that we have (1), (2), and (3) of our construction. Finally, let

$$q(x_0, ..., x_{2N-1}) = q^{2N-1}(x_0, ..., x_{2N-1})$$

and set $c = \langle b^n_k : n < 2N, k < K_n\rangle$, so $c \in B^K$. This completes our construction. We now show that p does not $\Delta_{K,\varphi}$-split over $\{c\}$, as desired.

Claim 3. For any $W \subseteq 2N$ and any $t < 2$, if there exists $b \in B$ such that

$$q(x_0, ..., x_{2N-1}) \cup \{\varphi(x_i;b)^t : i \in W\}$$

is consistent and $\models \neg\varphi(a;b)^t$, then, for any realization $\langle a_0, ..., a_{2N-1}\rangle \models q$, there exists $k_n < K_n$ for each $n \in W$ such that, for all $i, j \in W$,

$$\models \varphi(a_i; b^j_{k_j})^t \text{ if and only if } i < j.$$

That is, $\langle a_i : i \in W\rangle$ and $\langle b^j_{k_j} : j \in W\rangle$ witness that φ^t has the order property of length $|W|$.

Proof of Claim 3. Fix $\langle a_0, ..., a_{2N-1}\rangle \models q$, any $j \in W$, and let $W' = W \cap j$. Choose $k = k_j < K_j$ such that $f^j_0(k) = t$ and $f^j_1(k) = W'$. Since

$$q(x_0, ..., x_{2N-1}) \cup \{\varphi(x_i;b)^t : i \in W\}$$

is consistent, so is the subtype

$$q^j_k(x_0, ..., x_{j-1}) \cup \{\varphi(x_i;b)^t : i \in W'\}.$$

Hence, b is a witness to the fact that conditions (i) and (ii) of Case 1 are met. Therefore, $k \in I^j$ and, furthermore,

$$\{\varphi(x_i, b^j_k)^t : i \in W'\} \subseteq q(x_0, ..., x_{2N-1}).$$

Therefore, by construction, for all $i \in W$ with $i < j$ (i.e., all $i \in W'$), $\varphi(x_i, b^j_k)^t \in q$. That is, $\models \varphi(a_i, b^j_k)^t$.

On the other hand, for all $i \geq j$, the variables x_i in q match the value of a over b^j_k. That is, $\neg\varphi(x_i; b^j_k)^t \in q$. Hence, $\models \neg\varphi(a_i, b^j_k)^t$. Therefore, we see that, for all $i \in W$,

$$\models \varphi(a_i; b^j_{k_j}) \text{ if and only if } i < j.$$

□

Therefore, by our choice of N, for any $W \subseteq 2N$ with $|W| = N$, for any $b \in B$, and any $t < 2$, if

$$q(x_0, ..., x_{2N-1}) \cup \{\varphi(x_i; b)^t : i \in W\}$$

is consistent, then $\models \varphi(a; b)^t$. However, by pigeon-hole principle, for any $b \in B$, there exists $t < 2$ and $W \subseteq 2N$ with $|W| = N$ such that

$$q(x_0, ..., x_{2N-1}) \cup \{\varphi(x_i; b)^t : i \in W\}$$

is consistent. Therefore, for all $b \in B$, $\varphi(x; b) \in p(x)$ if and only if there exists $W \subseteq 2N$ with $|W| = N$ such that $q \cup \{\varphi(x_i; b) : i \in W\}$ is consistent. This suffices to complete the proof.

Fix $b_0, b_1 \in B$ so that $\neg\varphi(x; b_0), \varphi(x; b_1) \in p(x)$. By the characterization above, there exists $W \subseteq 2N$ with $|W| = N$ such that, for each $t < 2$,

$$q(x_0, ..., x_{2N-1}) \cup \{\varphi(x_i; b_t) : i \in W\}$$

is consistent if and only if $t = 1$. Therefore,

$$\theta(y) := \exists x_0 ... x_{2N-1} \left(\bigwedge q(x_0, ..., x_{2N-1}) \wedge \bigwedge_{i \in W} \varphi(x_i; y) \right)$$

holds of b_1 and fails of b_0. However, this is clearly equivalent to a boolean combination of $\Delta_{K,\varphi}$-formulas over c. Hence,

$$\mathrm{tp}_{\Delta_{K,\varphi}}(b_0/c) \neq \mathrm{tp}_{\Delta_{K,\varphi}}(b_1/c).$$

Hence p does not $\Delta_{K,\varphi}$-split over $\{c\}$, as desired. □

Exercise 4.5.6. Check the proof of Theorem 3.3.14 to verify the stronger result: If $\varphi(x; y)$ has VC-density < 2, then there exists $n < \omega$ such that, for all finite $B \subseteq \mathcal{U}_y$ and all $p \in S_\varphi(B)$, there exists $c \in B^n$ such that p does not $\Delta_{n,\varphi}$-split over $\{c\}$.

In particular, this holds for formulas in one variable in a weakly o-minimal theory.

Exercise 4.5.7. Show that, if $\varphi(x; y)$ is a formula with $|x| = 1$ in a dp-minimal theory, then there exists $n < \omega$ such that, for all finite $B \subseteq \mathcal{U}_y$ and all $p \in S_\varphi(B)$, there exists $c \in B^n$ such that p does not $\Delta_{n,\varphi}$-split over $\{c\}$. Hint: Another simple modification of the proof of Theorem 3.3.14.

What about formulas in multiple variables? Does subadditivity hold?

OPEN QUESTION 4.5.8. *Fix variables* x_1 *and* x_2 *and suppose that, for all formulas* $\varphi(x_i, y)$ *for* $i = 1, 2$*, there exists* $n_i < \omega$ *such that, for all finite* $B \subseteq \mathcal{U}_y$ *and all* $p \in S_\varphi(B)$*, there exists* $c \in B^{n_i}$ *such that* p *does not* $\Delta_{n_i,\varphi}$*-split over* $\{c\}$*. Then, for a formula* $\varphi(x_1, x_2; y)$*, does there exists* $n < \omega$ *such that, for all finite* $B \subseteq \mathcal{U}_y$ *and all* $p \in S_\varphi(B)$*, there exists* $c \in B^n$ *such that* p *does not* $\Delta_{n,\varphi}$*-split over* $\{c\}$*? Is it that* $n \leq n_1 + n_2$ *in this case?*

Bibliography

[1] M. Aschenbrenner, A. Dolich, D. Haskell, D. MacPherson, and S. Starchenko, *Vapnik-Chervonenkis density in some theories without the independence property, I*, Preprint (2011).

[2] ———, *Vapnik-Chervonenkis density in some theories without the independence property, II*, Preprint (2011).

[3] I. Barany and J. Matousek, *A fractional Helly theorem for convex lattice sets*, Advances in Mathematics **174** (2003), no. 2, 227–235.

[4] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth, *Learnability and Vapnik-Chervonenkis Dimension*, Journal of the Association for Computing Machinery **36** (1989), no. 4, 929–965.

[5] A. Chernikov and P. Simon, *Externally definable sets and dependent pairs*, Israel Journal of Mathematics (2012). to appear.

[6] ———, *Externally definable sets and dependent pairs, II* (2012). preprint.

[7] S. Floyd and M. Warmuth, *Sample compression, learnability, and the Vapnik-Chervonenkis dimension*, Machine Learning (1995), 269–304.

[8] V. Guingona and C. Hill, *On VC-density over indiscernible sequences* (2012). preprint.

[9] ———, *On a common generalization of Shelah's 2-rank, dp-rank, and o-minimal dimension* (2013). preprint.

[10] V. Guingona, *Dependence and isolated extensions*, Proceedings of the American Mathematical Society **139** (2011), 3349–3357.

[11] ———, *On uniform definabiliy of types over finite sets*, J. Symbolic Logic **77** (2012), no. 2, 499–514.

[12] H. R. Johnson, *dp-Rank and forbidden configurations*, Notre Dame Journal of Formal Logic **54** (2013), no. 1.

[13] ———, *Vapnik-Chervonenkis density on indiscernible sequences, stability, and the maximum property* (2013). preprint.

[14] H. R. Johnson and M. C. Laskowski, *Compression schemes, stable definable families, and o-minimal structures*, Discrete Comput. Geom. **43** (2010), no. 4, 914–926.

[15] I. Kaplan, A. Onshuus, and A. Usvyatsov, *Additivity of the dp-rank*, Trans. Amer. Math. Soc. (to appear).

[16] M. C. Laskowski, *Vapnik-Chervonenkis classes of definable sets*, J. London Math. Soc. **245** (1992), 377–384.

[17] J. Matousek, *Lectures on Discrete Geometry*, Springer-Verlag New York Inc., 2002.

[18] ———, *Bounded VC-dimension implies a fractional Helly theorem*, Discrete Comput. Geom. **31** (2004), 251–255.

[19] S. Shelah, *Dependent first order theories, continued*, Israel Journal of Mathematics **173** (2009), 1–60.
[20] ———, *Classification Theory and the Number of Non-Isomorphic Models.*
[21] P. Simon, *Distal and non-distal NIP theories*, Annals of Pure and Applied Logic **164** (2013), no. 3, 294–318.
[22] ———, *On dp-minimal ordered structures*, J. Symbolic Logic **76** (2011), 448–460.
[23] ———, *Invariant types in dp-minimal theories* (2012). preprint.
[24] ———, *Lecture notes on NIP theories*, 2012. preprint.