1. Computational Learning

Abstract. I discuss the basics of computational learning theory, including concept classes, VC-dimension, and VC-density. Next, I touch on the VC-Theorem, ϵ-nets, and the (p,q)-Theorem for VC-classes.

Definition 1.1. A *concept class* $\mathcal{C}$ on a set X is a subset of the powerset of X. Elements of $\mathcal{C}$ are known as *concepts*.

Definition 1.2. Let $\mathcal{C}$ be a concept class on a set X:

(1) for some $Y \subseteq X$, $\mathcal{C}$ *shatters* Y if

$$\mathcal{P}(Y) = \mathcal{C}|_Y := \{Y \cap A : A \in \mathcal{C}\}.$$

(2) $\mathcal{C}$ has *VC-dimension* $\geq d$ if there exists $Y \subseteq X$ with $|Y| = d$ such that $\mathcal{C}$ shatters Y.

(3) The *shatter function* of $\mathcal{C}$ is a function $\pi_{\mathcal{C}} : \omega \to \omega$ given by

$$\pi_{\mathcal{C}}(m) := \max\{|\mathcal{C}|_Y| : Y \subseteq X, |Y| = m\}.$$

Hence $\mathcal{C}$ has VC-dimension $\geq d$ iff. $\pi_{\mathcal{C}}(d) = 2^d$.

(4) $\mathcal{C}$ has *VC-density* $\leq \ell$ if there exists $K \in \mathbb{R}$ such that, for all $m < \omega$, $\pi_{\mathcal{C}}(m) \leq Km^\ell$.

(5) $\mathcal{C}$ is a *VC-class* if it has finite VC-dimension.

(6) The *dual concept class* (on the set $\mathcal{C}$) is given by

$$\mathcal{C}^* = \{\mathcal{D} \subseteq \mathcal{C} : (\exists x \in X)(\forall A \in \mathcal{C})(x \in A \Leftrightarrow A \in \mathcal{D})\}.$$

(7) The *dual shatter function* of $\mathcal{C}$ is $\pi^*_{\mathcal{C}} = \pi_{\mathcal{C}^*}$.

(8) the *VC-codimension* of $\mathcal{C}$ is the VC-dimension of $\mathcal{C}^*$.

(9) the *VC-codensity* of $\mathcal{C}$ is the VC-density of $\mathcal{C}^*$.

Example 1.3. Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the collection of open intervals in $\mathbb{R}$. Any two-element subset of X is shattered but no three-element subset of X is shattered. So, the VC-dimension of $\mathcal{C}$ is 2. For any $Y \subseteq \mathbb{R}$ with $|Y| = m$, there are $\binom{m}{2}$ elements of $\mathcal{C}|_Y$ with at least two elements, m elements with exactly one element, and 1 element with none. Hence, the shatter function is given by

$$\pi_{\mathcal{C}}(m) = \frac{1}{2}m^2 + \frac{1}{2}m + 1, \text{ for } m \geq 2.$$

Therefore, $\pi_{\mathcal{C}}(m) = \mathcal{O}(m^2)$, so $\mathcal{C}$ has VC-density 2. On can show that $\mathcal{C}$ has VC-codimension 2 and VC-codensity 1.

Theorem 1.4 ([11])**.** *A concept class $\mathcal{C}$ is a VC-class if and only if $\mathcal{C}^*$ is.*

Theorem 1.5 (Sauer's Lemma)**.** *If $\mathcal{C}$ has VC-dimension n and $m > n$, then*

$$\pi_{\mathcal{C}}(m) \leq \sum_{i=0}^{n} \binom{m}{i}.$$

In particular, the VC-density is bounded above by the VC-dimension.

Therefore, the following are equivalent for a concept class $\mathcal{C}$:

- $\mathcal{C}$ is a VC-class,
- $\mathcal{C}$ has finite VC-dimension,
- $\mathcal{C}$ has finite VC-density,
- $\mathcal{C}$ has finite VC-codimension,
- $\mathcal{C}$ has finite VC-codensity.

Definition 1.6. Let $\mathcal{C}$ be a concept class on a finite set X:

(1) A subset $T \subseteq X$ is a *transversal* of $\mathcal{C}$ if, for all $A \in \mathcal{C}$, $A \cap T \neq \emptyset$.
(2) The *transversal number* of $\mathcal{C}$, denoted $\tau(\mathcal{C})$, is the minimal cardinality of a transversal of $\mathcal{C}$.
(3) A *fractional transversal* of $\mathcal{C}$ is a map $\sigma : X \to [0, 1]$ such that, for each $A \in \mathcal{C}$, $\sum_{x \in A} \sigma(x) \geq 1$.
(4) The *fractional transversal number* of $\mathcal{C}$, denoted $\tau'(\mathcal{C})$, is the infimum of $\sum_{x \in X} \sigma(x)$ over all fractional transversals σ of $\mathcal{C}$.
(5) A subset $\mathcal{D} \subseteq \mathcal{C}$ is a *packing* of $\mathcal{C}$ if, for all $A, B \in \mathcal{D}$, $A \cap B = \emptyset$.
(6) The *packing number* of $\mathcal{C}$, denoted $\nu(\mathcal{C})$, is the maximal cardinality of a packing of $\mathcal{C}$.
(7) A *fractional packing* of $\mathcal{C}$ is a map $\sigma : \mathcal{C} \to [0, 1]$ such that, for each $x \in X$, $\sum_{A \in \mathcal{C}, A \ni x} \sigma(A) \leq 1$.
(8) The *fractional packing number* of $\mathcal{C}$, denoted $\nu'(\mathcal{C})$, is the supremum of $\sum_{A \in \mathcal{C}} \sigma(A)$ over all fractional packings σ of $\mathcal{C}$.

Remark 1.7. For any concept class $\mathcal{C}$ on a finite set X,

$$\nu(\mathcal{C}) \leq \tau(\mathcal{C}), \nu(\mathcal{C}) \leq \nu'(\mathcal{C}), \text{ and } \tau'(\mathcal{C}) \leq \tau(\mathcal{C}).$$

Example 1.8. Let $X = \{0, 1, 2\}$, $\mathcal{C} = \{\{0, 1\}, \{1, 2\}, \{0, 2\}\}$. Then we have that

- $\nu(\mathcal{C}) = 1$,
- $\tau(\mathcal{C}) = 2$,
- $\nu'(\mathcal{C}) = \frac{3}{2}$, and
- $\tau'(\mathcal{C}) = \frac{3}{2}$.

Theorem 1.9. *For every concept class $\mathcal{C}$ on a finite set X,*

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

Proof. This is proved via the Duality of Linear Programming. That is, if $\mathbf{A}$ is an $m \times n$ real matrix, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$, if

- $P = \{x \in \mathbb{R}^n : x \geq 0 \text{ and } \mathbf{A}x \geq b\}$, and
- $D = \{y \in \mathbb{R}^m : y \geq 0 \text{ and } y^T\mathbf{A} \leq c^T\}$

are both non-empty, then

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

Let $\mathbf{A}$ be the incidence matrix for X and $\mathcal{C}$. So, $\mathbf{A} = (\chi_A(x))_{A \in \mathcal{C}, x \in X}$, where $\chi_A : X \to \{0,1\}$ is the characteristic function of A. Set $b = \mathbf{1}_{\mathcal{C}}$ and $c = \mathbf{1}_X$ (tuples whose entries are 1). Check that

- $\tau'(\mathcal{C}) = \min\{c^T x : x \geq 0 \text{ and } \mathbf{A}x \geq b\}$, and
- $\nu'(\mathcal{C}) = \max\{y^T b : y \geq 0 \text{ and } y^T\mathbf{A} \leq c^T\}$.

□

The goal is now to find a condition to bound $\tau(\mathcal{C})$.

Definition 1.10. For $\epsilon > 0$, μ a probability measure on X, and $\mathcal{C}$ a concept class of μ-measurable subsets of X, an ϵ-net of $\mathcal{C}$ is a transversal for the set

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

Theorem 1.11 (The VC-Theorem)**.** *Fix $d, n < \omega$ and $\epsilon > 0$. Suppose that μ is a probability measure on X and $\mathcal{C}$ is a concept class on X (of μ-measurable sets) with VC-dimension $\leq d$. Then,*

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

Proof uses "double counting" method and Chebyshev's Inequality.

Corollary 1.12. *Fix $d < \omega$ and $r \in \mathbb{R}$. Then, there exists $n \in \mathbb{R}$ such that, for all finite concept classes $\mathcal{C}$ with VC-dimension $\leq d$ and $\nu'(\mathcal{C}) \leq r$, we have $\tau(\mathcal{C}) \leq n$.*

Proof. Fix $\epsilon = 1/r$ and n large enough such that $2(2n)^d 2^{-\epsilon n/2} < 1$. As $\tau'(\mathcal{C}) = \nu'(\mathcal{C}) \leq r$, $1/\tau'(\mathcal{C}) \geq \epsilon$. Let $\sigma : X \to [0,1]$ be an optimal fractional transversal and define a probability measure μ on X by letting $\mu(A) = \frac{1}{\tau'(\mathcal{C})} \sum_{x \in A} \sigma(x)$. For all $A \in \mathcal{C}$, $\mu(A) \geq \epsilon$ by definition, hence $\mathcal{C} = \mathcal{C}_\epsilon$. Then, the probability that a random n-element subset of X is not an ϵ-net (i.e., a transversal) for $\mathcal{C}$ is < 1. Hence, a n-element transversal exists. □

Definition 1.13. A concept class $\mathcal{C}$ has the (p,q)*-property* if, for all $\mathcal{D} \subseteq \mathcal{C}$ with $|\mathcal{D}| = p$, there exists $a \in X$ such that

$$|\{A \in \mathcal{D} : a \in A\}| \geq q.$$

In other words, every p-element subset of $\mathcal{C}$ has at least q elements that have a common intersection.

Proposition 1.14 ([13]). *For all $0 < q \leq p$ and $K < \omega$, there exists $r \in \mathbb{R}$ such that, for all finite concept classes $\mathcal{C}$ such that, $\mathcal{C}$ has the (p,q)-property and $(\forall m \geq q)(\pi^*_{\mathcal{C}}(m) < K\binom{m}{q})$, we have*

$$\nu'(\mathcal{C}) \leq r.$$

The proof of this uses the Fractional Helly Property.

Theorem 1.15 ((p,q)-Theorem for VC-classes). *If $0 < q \leq p < \omega$ and $\mathcal{C}$ is a VC-class on X with VC-codensity $< q$, then there exists $n < \omega$ such that, for all finite $\mathcal{D} \subseteq \mathcal{C}$, if $\mathcal{D}$ has the (p,q)-property, then $\tau(\mathcal{D}) \leq n$. (Moreover, n only depends on q, p, the VC-dimension of $\mathcal{C}$, and the behavior of $\pi^*_{\mathcal{C}}$.)*

Proof. As $\mathcal{C}$ is a VC-class, choose d such that $\mathcal{C}$ has VC-dimension $\leq d$ and, since $\mathcal{C}$ has VC-codensity $< q$, choose K such that $(\forall m \geq q)(\pi^*_{\mathcal{C}}(m) < K\binom{m}{q})$. Let r be given by Proposition 1.14 for our q, p, and K. Then, let n be given by Corollary 1.12 for r and d. Proposition 1.14 gives us that $\nu'(\mathcal{D}) \leq r$ and Corollary 1.12 gives us that $\tau(\mathcal{D}) \leq n$. □

Suppose $p = q$. Then the (q,q)-Theorem is, in effect, a "finite version of compactness" for NIP Theories. It says that, for a finite VC-class $\mathcal{C}$, if every q element subset has a common intersection, then there is a uniform bound on the number of elements needed to cover $\mathcal{C}$.

2. Introduction to NIP Theories

Abstract. First, I conclude computational learning by discussing two notions of learning. Then, I discuss the notion of UDTFS and how it relates to sequence compression schemes. Finally, I begin talking about NIP Theories.

Let X be any set and $\mathcal{C}$ a concept class on X. Let $\mathcal{P}_{\text{fin}}(X)$ be the finite subsets of X.

Definition 2.1. We say that $\mathcal{C}$ is *probably approximately correctly learnable* (*PAC-learnable*) if there exists $\langle H_Y : Y \in \mathcal{P}_{\text{fin}}(X)\rangle$ a sequence of *hypothesis* functions $H_Y : \mathcal{C}|_Y \to \mathcal{P}(X)$ 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 $A \in \mathcal{C}$, and all probability measures μ on X, the μ-probability that a randomly chosen $a_1, ..., a_n \in X$ satisfies

$$\mu(A \triangle H_{\{a_1,...,a_n\}}(A \cap \{a_1, ..., a_n\})) > \epsilon$$

is $< \delta$.

Theorem 2.2 (PAC-Learning Theorem, [3])**.** *For any concept class $\mathcal{C}$, the following are equivalent:*

- *$\mathcal{C}$ is a VC-class,*
- *$\mathcal{C}$ is PAC-learnable.*

This is proved using the VC-Theorem from last talk.

Example 2.3 (Example 1.3, Revisited). Consider $X = \mathbb{R}$ and $\mathcal{C}$ the concept class of all open intervals. For every finite $Y \subseteq \mathbb{R}$ and every $A \subseteq Y$, let

$$H_Y(A) = \{x \in \mathbb{R} : (\exists a, b \in A)(a \leq x \leq b)\}.$$

Then, one can show that this witnesses that $\mathcal{C}$ is PAC-learnable with

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

The function $\langle \epsilon, \delta\rangle \mapsto N_{\epsilon,\delta}$ is called the *sample complexity* of the hypothesis function $\overline{H} = \langle H_Y : Y \in \mathcal{P}_{\text{fin}}(X)\rangle$.

Definition 2.4. We say that $\mathcal{C}$ has a *d-dimensional sequence compression scheme* if there exists a sequence of *compression* functions $\langle \kappa_Y : Y \in \mathcal{P}_{\text{fin}}(X)\rangle$ (with $\kappa_Y : \mathcal{C}|_Y \to Y^d$) and a finite number of *recovery* functions $\langle \rho_i : i < K\rangle$ (with $\rho_i : X^d \to \mathcal{P}(X)$) such that, for each finite non-empty $Y \subseteq X$ and $A \in \mathcal{C}|_Y$, there exists $i < K$ such that

$$\rho_i(\kappa_Y(A)) \cap Y = A.$$

Example 2.5 (Example 1.3, Revisited). Consider $X = \mathbb{R}$ and $\mathcal{C}$ the concept class of all open intervals. Consider the compression $\kappa_Y(A) = \langle \min A, \max A \rangle$ for $A \neq \emptyset$ and $\kappa_Y(\emptyset) = \langle \min Y, \max Y \rangle$. Consider the recovery functions $\rho_1(a,b) = \{x \in \mathbb{R} : a \leq x \leq b\}$ and $\rho_2(a,b) = \emptyset$. Check that, for any non-empty finite $Y \subseteq \mathbb{R}$ and any $A \in \mathcal{C}|_Y$,

- $\rho_1(\kappa_Y(A)) \cap Y = A$ if $A \neq \emptyset$, and
- $\rho_2(\kappa_Y(\emptyset)) \cap Y = \emptyset$.

There is no 1-dimensional compression scheme (by the following proposition).

Proposition 2.6. *If $\mathcal{C}$ has a d-dimensional sequence compression scheme, then $\mathcal{C}$ has VC-density $\leq d$. In particular, $\mathcal{C}$ is a VC-class.*

Proof. For any finite non-empty $Y \subseteq X$, an element $A \in \mathcal{C}|_Y$ is determined by an element in Y^d (namely, $\kappa_Y(A)$) and ρ_i for some $i < K$. Hence,

$$|\mathcal{C}|_Y| \leq K|Y|^d.$$

□

Conjecture 2.7 (Warmuth Conjecture)**.** *If $\mathcal{C}$ is a VC-class, then $\mathcal{C}$ has a sequence compression scheme.*

Proposition 2.8 ([7])**.** *The following hold for a concept class $\mathcal{C}$:*

- *If $\mathcal{C}$ has VC-density < 2, then $\mathcal{C}$ has a sequence compression scheme.*
- *If $\mathcal{C}$ has VC-dimension ≤ 1, then $\mathcal{C}$ has a 1-dimensional sequence compression scheme.*

Let T be a complete, first-order theory in a language L with monster model $\mathcal{U}$. We will bring all the definitions and results from the previous talk into model theory with the following convention:

Definition 2.9. Fix a partitioned formula $\varphi(x; y)$. For $a \in \mathcal{U}_x$ and $B \subseteq \mathcal{U}_y$, let

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

Now consider the following concept class on $\mathcal{U}_y$,

$$\mathcal{C}_\varphi^{\mathcal{U}} := \{\varphi(a; \mathcal{U}_y) : a \in \mathcal{U}_x\}.$$

From this, we get:

- Say φ has *VC-dimension n* if $\mathcal{C}_\varphi^{\mathcal{U}}$ does.
- Say φ has *VC-density ℓ* if $\mathcal{C}_\varphi^{\mathcal{U}}$ does.
- Say φ has *NIP* if $\mathcal{C}_\varphi^{\mathcal{U}}$ is a VC-class.
- Say T has *NIP* if every partitioned formula $\varphi(x; y)$ does.

Example 2.10. The following theories have NIP:

- $T = \mathrm{Th}(\mathbb{C}; +, \cdot, 0, 1)$, the theory of algebraically closed fields (of characteristic 0).
- $T = \mathrm{Th}(\mathbb{R}; +, \cdot, <, 0, 1)$, the theory of real closed fields.
- $T = \mathrm{Th}(\mathbb{Q}_p; +, \cdot, 0, 1)$, the theory of the p-adic field.
- $T = \mathrm{Th}(\mathbb{Q}; +, <)$, the theory of densely ordered abelian groups.

The theory of Rado's random graph does not have NIP.

Definition 2.11. A formula $\varphi(x; y)$ has *uniform definability of types over finite sets* (*UDTFS*) (*of rank* $\leq d$) if there exists finitely many formulas $\langle \psi_i(y; z_1, ..., z_d) : i < K \rangle$ such that, for all finite $B \subseteq \mathcal{U}_y$ and all $a \in \mathcal{U}_x$, there exists $c_1, ..., c_d \in B$ and $i < K$ such that

$$\varphi(a; B) = \psi_i(B; c_1, ..., c_n).$$

Proposition 2.12 ([1])**.** *If T is weakly o-minimal, then $\varphi(x; y)$ had UDTFS-rank $\leq |x|$.*

In particular, this holds for real closed fields and densely ordered abelian groups, as these theories are weakly o-minimal.

Theorem 2.13 ([10])**.** *If φ has UDTFS-rank d, then $\mathcal{C}_\varphi^{\mathcal{U}}$ has a d-dimensional sequence compression scheme.*

Proof. Let $\langle \psi_i(y; z_1, ..., z_d) : i < K \rangle$ witness that $\varphi(x; y)$ has UDTFS-rank d. For each $i < K$, consider the recovery function $\rho_i : \mathcal{U}_y^d \to \mathcal{P}(\mathcal{U}_y)$ given by

$$\rho_i(c_1, ..., c_d) = \psi_i(\mathcal{U}_y; c_1, ..., c_d).$$

For any finite non-empty $B \subseteq \mathcal{U}_y$, consider the compression function $\kappa_B : \mathcal{C}_\varphi^{\mathcal{U}}|_B \to B^d$ given by: For $a \in \mathcal{U}_x$, let $c_1, ..., c_d \in B$ and $i < K$ be such that $\varphi(a; B) = \psi_i(B; c_1, ..., c_d)$. Then, set

$$\kappa_B(\varphi(a; B)) = \langle c_1, ..., c_d \rangle.$$

Hence,

$$\rho_i(\kappa_B(\varphi(a; B))) \cap B = \psi_i(B; c_1, ..., c_d) = \varphi(a; B).$$

Therefore, this is a d-dimensional sequence compression scheme for $\mathcal{C}_\varphi^{\mathcal{U}}$. □

Example 2.14. Consider $\varphi(x_1, x_2, x_3; y_1, y_2) := (x_1 - y_1)^2 + (x_2 - y_2)^2 \leq x_3^2$ in real closed fields. Then, $\mathcal{C}_\varphi^{\mathbb{R}}$ is the set of all closed disks in $\mathbb{R}^2$. By Proposition 2.12 and Theorem 2.13, it has a 3-dimensional sequence compression scheme.

Corollary 2.15. *If φ has UDTFS-rank d, then φ has VC-density $\leq d$. In particular, if φ has UDTFS, then it has NIP.*

What about the converse?

Conjecture 2.16 (UDTFS Conjecture, [7]). *If $\varphi(x;y)$ has NIP, then φ has UDTFS.*

By Theorem 2.13, the UDTFS Conjecture implies the Warmuth Conjecture.

Proposition 2.17 ([7]). *Fix a partitioned formula, $\varphi(x;y)$.*

(1) *If φ has VC-density < 2, then φ has UDTFS.*
(2) *If φ had VC-dimension ≤ 1, then φ has UDTFS-rank ≤ 1.*

In particular, if $\mathcal{C}$ is a concept class with VC-density < 2, then $\mathcal{C}$ has a sequence compression scheme and, if $\mathcal{C}$ is a concept class with VC-dimension ≤ 1, then $\mathcal{C}$ has a 1-dimensional sequence compression scheme. So this proves Proposition 2.8.

Proof of Proposition 2.17 (2). Without loss of generality, we may suppose that, for all $b, c \in \mathcal{U}_y$, at least one of the following holds:

- $\models (\forall x)(\varphi(x;b) \to \varphi(x;c))$,
- $\models (\forall x)(\varphi(x;c) \to \varphi(x;b))$, or
- $\models \neg(\exists x)(\varphi(x;b) \wedge \varphi(x;c))$.

This puts a definable quasi-forest order $\trianglelefteq_\varphi$ on $\mathcal{U}_y$, namely

$$b \trianglelefteq_\varphi c \text{ if } \models (\forall x)(\varphi(x;b) \to \varphi(x;c)).$$

Let $\psi(y;z) = [z \trianglelefteq_\varphi y]$. Then, for any $a \in \mathcal{U}_x$ and finite $B \subseteq \mathcal{U}_y$, choose $c \in B$ such that $\models \varphi(a;c)$ and c is $\trianglelefteq_\varphi$-minimal such in B. Then, it is easy to check that

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

□

Theorem 2.18 ([7]). *If T is dp-minimal, then all formulas have UDTFS.*

Example 2.19. The p-adics are dp-minimal. Hence, the concept class of all closed balls in the p-adic field have a sequence compression scheme. Use

$$\varphi(x_1, x_2; y) := v_p(y - x_1) \geq v_p(x_2).$$

The following partial result to the UDTFS Conjecture is due to Artim Chernikov and Pierre Simon:

Theorem 2.20 ([5]). *If T has NIP, then all formulas have UDTFS.*

We will spend the final talk proving this theorem.

3. The UDTFS Conjecture

Abstract. Starting with "honest definitions," I discuss Chernikov and Simon's partial solution to the UDTFS Conjecture and the ramifications this has on computational learning theory.

Let T be a complete, first-order theory in a language L with monster model $\mathcal{U}$.

Recall the definition of UDTFS: A formula $\varphi(x; y)$ has UDTFS if there exists finitely many formulas $\langle \psi_i(y; z) : i < K \rangle$ such that, for all finite $B \subseteq \mathcal{U}_y$ and all $a \in \mathcal{U}_x$, there exists $c \in B^n$ and $i < K$ such that

$$\varphi(a; B) = \psi_i(B; c).$$

Recall also Theorem 2.20 by Chernikov and Simon: If T has NIP, then all formulas have UDTFS. We prove this now.

Definition 3.1. Fix $C \subseteq \mathcal{U}$ a set, $\langle I, < \rangle$ a linear order, and $b_i \in \mathcal{U}$ for each $i \in I$ (all of the same sort). We say that $\overline{b} := \langle b_i : i \in I \rangle$ is an *indiscernible sequence over* C if, for all $n < \omega$, for all $i_1 < ... < i_n$ and $j_1 < ... < j_n$ from I, for all $L(C)$-formulas $\delta(y_1, ..., y_n)$,

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

Proposition 3.2. *A formula $\varphi(x; y)$ has NIP if and only if there exists $n < \omega$ such that, for all indiscernible sequences $\overline{b} = \langle b_i : i \in I \rangle$ and all $a \in \mathcal{U}_x$, there do not exist $i_0 < ... < i_n$ such that*

$$\models \varphi(a; b_{i_\ell}) \leftrightarrow \neg\varphi(a; b_{i_{\ell+1}})$$

for all $\ell < n$.

Proof. ($\Leftarrow$): If $\varphi(x; y)$ does not have NIP, then by compactness and Ramsey's Theorem, there exists an indiscernible sequence $\langle b_i : i < \omega \rangle$ such that, for any $I \subseteq \omega$, $\{\varphi(x; b_i) : i \in I\} \cup \{\neg\varphi(x; b_i) : i \in \omega \setminus I\}$ is consistent. In particular, for I the even numbers, we have infinite alternation.

($\Rightarrow$): Fix n and suppose a and $\langle b_i : i < 2n \rangle$ are such that $\models \varphi(a; b_i)$ if and only if i is even. For any $s : n \to 2$, consider the formula

$$\exists x \left(\bigwedge_{i<n} \varphi(x; b_{2i+s(i)})^{s(i)} \right).$$

As this is true, by indiscernibility, the following is also true

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

As $s \in {}^n2$ was arbitrary, φ has VC-dimension $\geq n$. $\square$

Definition 3.3. Fix a global type $p(x) \in S_x(\mathcal{U})$ and a set $A \subseteq \mathcal{U}$.

- p is *invariant* over A if, for all $\sigma \in \mathrm{Aut}(\mathcal{U}/A)$, for all $\varphi(x; y)$, and all $b \in \mathcal{U}_y$, $\varphi(x; b) \in p(x)$ if and only if $\varphi(x; \sigma(b)) \in p(x)$.
- p is *definable* over A if, for all formulas $\varphi(x; y)$, there exists an $L(A)$-formula $\psi(y)$ such that, for all $b \in \mathcal{U}_y$, $\varphi(x; b) \in p(x)$ if and only if $\models \psi(b)$.
- p is *finitely satisfiable* over A if, for all formulas $\varphi(x; b)$ in $p(x)$, there exists $a \in A$ such that $\models \varphi(a; b)$.

Lemma 3.4. *If $p(x) \in S_x(\mathcal{U})$ is definable over A or finitely satisfiable over A, then p is invariant over A.*

Proof. (1): If $p(x)$ is definable over A, fix $\varphi(x; y)$ and suppose $\psi(y)$ is an $L(A)$-formula defining $\varphi(x; y)$ in p. Then, for any $b \in \mathcal{U}_y$ and $\sigma \in \mathrm{Aut}(\mathcal{U}/A)$, $\varphi(x; b) \in p(x)$ iff. $\models \psi(b)$ iff. $\models \psi(\sigma(b))$ iff. $\varphi(x; \sigma(b)) \in p(x)$.

(2): Suppose $p(x)$ is finitely satisfiable over A. If $p(x)$ is not invariant over A, then fix $\varphi(x; y)$, $b \in \mathcal{U}$, and $\sigma \in \mathrm{Aut}(\mathcal{U}/A)$ witnessing this. That is, $\varphi(x; b) \in p(x)$ but $\neg\varphi(x; \sigma(b)) \in p(x)$. Hence, there exists $a \in A$ such that $\models \varphi(a; b) \wedge \neg\varphi(a; \sigma(b))$. As $\varphi(a; y)$ is over A, this is a contradiction. □

Definition 3.5. For a set $A \subseteq \mathcal{U}$, define

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

Lemma 3.6. *For all A, each of $S_x^{\mathrm{inv}}(\mathcal{U}, A)$, $S_x^{\mathrm{fin}}(\mathcal{U}, A)$, and $S_x^{\mathrm{def}}(\mathcal{U}, A)$ are closed subsets of the compact space $S_x(\mathcal{U})$, hence each are compact.*

Proposition 3.7. *If $p(x) \in S_x(\mathcal{U})$ is invariant over A 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. We call such sequences Morley sequences *of p over A.*

Fix $M \models T$ and $B \subseteq M_y$ for some variable y. Define a new language $L_P := L \cup \{P(y)\}$. $\langle M, B \rangle$ is the obvious L_P-structure. Fix $\varphi(x; y)$ a formula and $a \in M_x$. As $\mathcal{U}$ is a monster model for T, all models (even L_P-structures) will have universes contained in $\mathcal{U}$.

Definition 3.8. 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

$\langle M', B' \rangle \succeq \langle M, B \rangle$ and $c \in (B')^n$ such that

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

In particular, $\varphi(a; B) = \psi(B; c)$, so $\psi(y; c)$ defines $\mathrm{tp}_\varphi(a/B)$.

Theorem 3.9 ([4])**.** *If $\varphi(x; y)$ has NIP, $a \in M_x$, $B \subseteq M_y$, then $\mathrm{tp}_\varphi(a/B)$ has an honest definition.*

To prove this, we choose a type q in the space $S_y^{\mathrm{fin}}(\mathcal{U}, B)$ and build a Morley sequence of q over B in B' enforcing alternation of $\varphi(a; y)$. By NIP, the alternation stops and, by compactness, we get $\theta_q(y) \in q(y)$ and $t_q < 2$ such that

$$\theta_q(y) \wedge P(y) \vdash \varphi(a; y)^{t_q}.$$

We use the compactness of $S_y^{\mathrm{fin}}(\mathcal{U}, B)$ to conclude.

Proposition 3.10. *We have $\psi(y; z)$ is an honest definition of* $\mathrm{tp}_\varphi(a/B)$ *if and only if, for all finite $B_0 \subseteq \varphi(a; B)$, there is $c \in B^n$ such that*

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

This is proved using the compactness of $S_y(B)$.

So, for any NIP formula $\varphi(x; y)$, any $B \subseteq \mathcal{U}_y$, and any $a \in \mathcal{U}_x$, there exists $\psi(y; z)$ such that, for any finite $B_0 \subseteq \varphi(a; B)$, there exists $c \in B^n$ so that

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

In particular, if B is finite and $B_0 = \varphi(a; B)$, we get

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

So we only need uniformity of ψ (i.e., independence from a and B).

By compactness, we get the following lemma.

Lemma 3.11. *Let $\varphi(x; y)$ be an L-formula. For any function η from L-formulas to ω, there are finitely many formulas $\psi_j(y; z_j)$ (for $j < m$) such that, for all $M \models T$, $B \subseteq M_y$, and $a \in M_x$, there exists $j < m$ such that, for all $B_0 \subseteq \varphi(a; B)$ with $|B_0| \leq \eta(\psi_j)$, there exists $c \in B^n$ such that*

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

Lemma 3.12. *Let $\varphi(x; y)$ be an L-formula. There exists 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 $a \in M_x$, for all $B_0 \subseteq \varphi(a; B)$ finite, there exists $c \in B^n$ and $j < m$ such that*

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

Proof of Lemma 3.12. For each $\psi(y; z)$, let $\eta(\psi)$ be the VC-density of ψ. Then there exists $\theta_0(y; z), ..., \theta_{m-1}(y; z)$ satisfying the conclusion of the previous lemma for η. Let $p = q = \eta(\theta_j)$ and we get K_j by the (p, q)-Theorem. Let $K = \max\{K_j : j < m\}$ and, for each $j < m$, define

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

This works.

Fix $M \models T$, $B \subseteq M_y$, $a \in M_x$, $B_0 \subseteq \varphi(a; B)$ finite. For $j < m$ in the previous lemma, define

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

Define a concept class on C,

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

This satisfies the (p, q)-property (with $p = q = \eta(\theta_j)$) by the previous lemma. [This is because, for each $B_1 \subseteq B_0$ (the index set for $\mathcal{D}$) with $|B_1| = q = \eta(\theta_j)$, there exists $c \in B^n$ so that $B_1 \subseteq \theta_j(B; c) \subseteq \varphi(a; B)$, hence $c \in C$ and $c \in \bigcap_{b \in B_1} \theta_j(b; C)$.] By the (p, q)-Theorem (and our choice of $K_j \leq K$), there exists a transversal $C_0 \subseteq C$ with $|C| \leq K$. If $\bar{c}_0$ enumerates C_0, then

$$B_0 \subseteq \psi_j(B, \bar{c}_0) \subseteq \varphi(a; B).$$

□

Setting $B_0 = \varphi(a; B)$ for finite B, this concludes the proof of Theorem 2.20.

References

[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] 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.

[4] A. Chernikov and P. Simon, *Externally definable sets and dependent pairs*, Israel Journal of Mathematics (2012). to appear.

[5] ———, *Externally definable sets and dependent pairs, II* (2012). preprint.

[6] S. Floyd and M. Warmuth, *Sample compression, learnability, and the Vapnik-Chervonenkis dimension*, Machine Learning (1995), 269–304.

[7] V. Guingona, *On uniform definabiliy of types over finite sets*, J. Symbolic Logic **77** (2012), no. 2, 499–514.

[8] H. R. Johnson, *dp-Rank and forbidden configurations*, Notre Dame Journal of Formal Logic **54** (2013), no. 1.

[9] ———, *Vapnik-Chervonenkis density on indiscernible sequences, stability, and the maximum property* (2013). preprint.

[10] 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.

[11] M. C. Laskowski, *Vapnik-Chervonenkis classes of definable sets*, J. London Math. Soc. **245** (1992), 377–384.

[12] J. Matousek, *Lectures on Discrete Geometry*, Springer-Verlag New York Inc., 2002.

[13] ———, *Bounded VC-dimension implies a fractional Helly theorem*, Discrete Comput. Geom. **31** (2004), 251–255.

[14] S. Shelah, *Dependent first order theories, continued*, Israel Journal of Mathematics **173** (2009), 1–60.

[15] ———, *Classification Theory and the Number of Non-Isomorphic Models.*

[16] P. Simon, *On dp-minimal ordered structures*, J. Symbolic Logic **76** (2011), 448–460.

[17] ———, *Invariant types in dp-minimal theories* (2012). preprint.

[18] ———, *Lecture notes on NIP theories*, 2012. preprint.

Ben-Gurion University of the Negev

URL: `http://www.math.bgu.ac.il/~guingona/`

E-mail address: `guingona@math.bgu.ac.il`