TOPICS IN COMBINATORICS (MATH 285N, UCLA, WINTER 2016)

ARTEM CHERNIKOV

1. Vapnik–Chervonenkis dimension: basic properties and examples

Let X be a set (finite or infinite), and let $\mathcal{F}$ be a family of subsets of X. A pair $(X, \mathcal{F})$ is called a *set system.*

Given $A \subseteq X$, we say that the family $\mathcal{F}$ *shatters* A if for every $A' \subseteq A$, there is a set $S \in \mathcal{F}$ such that $S \cap A = A'$.

The family $\mathcal{F}$ has *VC-dimension* at most n (written as $\mathrm{VC}(\mathcal{F}) \leq n$), if there is no $A \subseteq X$ of cardinality $n+1$ such that $\mathcal{F}$ shatters A. We say that $\mathcal{F}$ is of VC-dimension n if it is of VC-dimension at most n and shatters some subset of size n.

If for every $n \in \mathbb{N}$ we can find a subset of X of cardinality n shattered by $\mathcal{F}$, then we say that $\mathcal{F}$ has infinite VC-dimension ($\mathrm{VC}(\mathcal{F}) = \infty$). If $\mathrm{VC}(\mathcal{F})$ is finite, we say that $\mathcal{F}$ is a *VC-family.* Note that if $\mathcal{F}' \subseteq \mathcal{F}$ then $\mathrm{VC}(\mathcal{F}') \leq \mathrm{VC}(\mathcal{F})$.

Example 1.1. Let $X = \mathbb{R}$ and let $\mathcal{F}$ be the family of all unbounded intervals. Then $\mathcal{F}$ has VC-dimension 2. Clearly any two-element set can be shattered by $\mathcal{F}$. However, if we take any $a < b < c$, then $\{a, b, c\}$ cannot be shattered by $\mathcal{F}$.

Exercise 1.2. Let $X = \mathbb{R}^2$, and let $\mathcal{F}$ be the set of all half-spaces. Show that $\mathrm{VC}(\mathcal{F}) = 3$.

Exercise 1.3. Let $X = \mathbb{R}^2$ and let $\mathcal{F}$ be the set of all convex polygons. Show that $\mathrm{VC}(\mathcal{F}) = \infty$.

Exercise 1.4. Show that for every $d \in \mathbb{N}$, there is a convex set on the real plane such that the family of all of its isometric copies has VC-dimension at least d.

For a finer analysis, we define the *shatter function* $\pi_{\mathcal{F}} : \mathbb{N} \to \mathbb{N}$ associated to the family $\mathcal{F}$ as follows. For a set $A \subseteq X$ we let $\mathcal{F} \cap A := \{S \cap A : S \in \mathcal{F}\}$. Then $\pi_{\mathcal{F}}(n)$ is the maximum over all $A \subseteq X$ of $|A \cap \mathcal{F}|$.

Note that $\pi_{\mathcal{F}}(n) \leq 2^n$, and that $\mathrm{VC}(\mathcal{F}) < n \iff \pi_{\mathcal{F}}(m) < 2^m$ for all $m \geq n$. The following fundamental lemma states that either $\pi_{\mathcal{F}}(n) = 2^n$ for all $n \in \mathbb{N}$, or $\pi_{\mathcal{F}}(n)$ has polynomial growth.

Lemma 1.5. *(Sauer-Shelah lemma) Let $(X, \mathcal{F})$ be a set system of VC-dimension at most k. Then, for all $n \geq k$, we have $\pi_{\mathcal{F}}(n) \leq \sum_{i=0}^{k} \binom{n}{i}$.*

In particular, $\pi_{\mathcal{F}}(n) = O\left(n^k\right)$.

There are numerous proofs and generalizations, we give a proof using the so-called "shifting" technique. Notice that the bound is tight: take $\mathcal{F}$ to be the family of all subsets of X of cardinality $\leq k$, then $\mathcal{F}$ has VC-dimension exactly k and its shatter function is equal to the bound in the statement of the lemma. The idea of the proof is to reduce the general situation to this case by modifying the elements of $\mathcal{F}$ making them as small as possible without changing the cardinality or VC-dimension of $\mathcal{F}$.

Proof. Fix an integer $n \geq k$. If $\mathcal{F}$ contradicts the bound, then this is also true for some finite subfamily of $\mathcal{F}$, so for the proof we may assume that $\mathcal{F}$ is finite. Similarly we may assume that X is finite, say, $X = \{x_1, \ldots, x_n\}$, and $\pi_{\mathcal{F}}(n) = |\mathcal{F}|$.

We define recursively families $\mathcal{F}_0, \ldots, \mathcal{F}_n$ of subsets of X. Set $\mathcal{F}_0 = \mathcal{F}$.

Let $l < n$ and assume that $\mathcal{F}_l$ has been defined. Go through the sets in $\mathcal{F}_l$ one by one. For each $S \in \mathcal{F}_l$, if $x_{l+1} \in S$ and $S \setminus \{x_{l+1}\}$ is not in $\mathcal{F}_l$, replace S by $S \setminus \{x_{l+1}\}$. If not, leave S as it is. Let $\mathcal{F}_{l+1}$ be the resulting family.

The following is straightforward by construction:

(1) for each l, $|\mathcal{F}_{l+1}| = |\mathcal{F}_l|$;
(2) let $S \in \mathcal{F}_l$ and $A = S \cap \{x_1, \ldots, x_l\}$, then for every $A_0 \subseteq A$, the set $A_0 \cup (S \setminus A)$ is in $\mathcal{F}_l$;
(3) any $A \subseteq X$ shattered by $\mathcal{F}_{l+1}$ is also shattered by $\mathcal{F}_l$.

It follows from (2) that if $S \in \mathcal{F}_n$, then S is shattered by $\mathcal{F}_n$. It follows from (3) that the $\mathrm{VC}(\mathcal{F}_n) \leq \mathrm{VC}(\mathcal{F})$. Therefore no set in $\mathcal{F}_n$ can have cardinality greater than k. Hence, by (3) we have $\sum_{i=0}^{k} \binom{n}{i} \geq |\mathcal{F}_n| = |\mathcal{F}| = \pi_{\mathcal{F}}(n)$. $\square$

Exercise 1.6. Prove a more general fact, due to Pajor: every finite set system $(X, \mathcal{F})$ shatters at least $|\mathcal{F}|$ subsets of X. Deduce Sauer-Shelah lemma from it.

We consider some general ways of producing VC-families.

Exercise 1.7. (Boolean operations preserve finite VC-dimension) Let $\mathcal{F}_1, \mathcal{F}_2$ be two families of subsets of X of finite VC-dimension. Show that all of the following families have finite VC-dimension:

(1) $\mathcal{F} := \mathcal{F}_1 \cup \mathcal{F}_2$,
(2) $\mathcal{F}_\cap := \{S_1 \cap S_2 : S_i \in \mathcal{F}_i, i = 1, 2\}$,
(3) $\mathcal{F}_\cup := \{S_1 \cup S_2 : S_i \in \mathcal{F}_i, i = 1, 2\}, \mathcal{F}_1^c := \{X \setminus S_1 : S_1 \in \mathcal{F}_1\}$,
(4) $\mathcal{F}_1 \times \mathcal{F}_2 := \{S_1 \times S_2 : S_1 \in \mathcal{F}_1, S_2 \in \mathcal{F}_2\}$ — a family of subsets of $X \times X$.
(5) Besides, if X' is an infinite set and $f : X' \to X$ is a map, let $f^{-1}(\mathcal{F}_1) := \left\{f^{-1}(S) : S \in \mathcal{F}_1\right\}$. Then $\mathrm{VC}\left(f^{-1}(\mathcal{F}_1)\right) \leq \mathrm{VC}(\mathcal{F}_1)$.

(Hint: bound the shattering functions of the corresponding families in terms of $\pi_{\mathcal{F}_1}, \pi_{\mathcal{F}_2}$ and use Lemma 1.5.)

Definition 1.8. Given a set system $(X, \mathcal{F})$, we define the *dual set system* $(X^*, \mathcal{F}^*)$, where $X^* = \mathcal{F}$ and $\mathcal{F}^* = \{\mathcal{F}_a : a \in X\}$ with $\mathcal{F}_a = \{S \in \mathcal{F} : a \in S\}$. We then define the *dual VC-dimension* of $\mathcal{F}$ (written as $\mathrm{VC}^*(\mathcal{F})$) as the VC-dimension of $\mathcal{F}^*$, and the *dual shatter function* $\pi^*_{\mathcal{F}}$ as the shatter function of $\mathcal{F}^*$.

Given a set system $(X, \mathcal{F})$, we can consider its *incidence matrix* M defined as an $|X| \times |\mathcal{F}|$-matrix such that rows are identified with the elements of x, columns with the elements of $\mathcal{F}$ and for any $x \in X, S \in \mathcal{F}$ the corresponding entry $M_{x,S}$ is 1 if $x \in S$ and 0 otherwise. Note that if M is the incidence matrix for $(X, \mathcal{F})$ then the transposed matrix M^{T} is the incidence matrix of the dual system $(X^*, \mathcal{F}^*)$.

Exercise 1.9. Let $X = \mathbb{R}^2$ and let $\mathcal{F}$ be the family of all open half-planes. Then $\pi^*_{\mathcal{F}}(n)$ is the maximal number of regions into which n lines can partition the plane. Show by induction that this number is equal to $\frac{n(n+1)}{2} + 1$.

Lemma 1.10. *(VC duality) We have* $\mathrm{VC}^*(\mathcal{F}) < 2^{\mathrm{VC}(\mathcal{F})+1}$ *and* $\mathrm{VC}(\mathcal{F}) < 2^{\mathrm{VC}^*(\mathcal{F})+1}$.

Proof. Assume that $\mathrm{VC}(\mathcal{F}) \geq 2^n$. Then there is some subset $B \subseteq X$ of size 2^n shattered by $\mathcal{F}$. Write $B = \{b_J : J \subseteq n\}$, and we have $\{S_I : I \subseteq \mathcal{P}(n)\} \subseteq \mathcal{F}$ such that $b_J \in S_I \iff J \in I$. For each $k < n$, let $I_k = \{J \subseteq n : k \in J\}$. Then $J \in I_k \iff k \in J$, hence we have $b_J \in S_{I_k} \iff k \in J$. This shows that the set $\{S_{I_k} : k < n\} \subseteq X^*$ is shattered by $\{\mathcal{F}_{b_J} : J \subseteq n\} \subseteq \mathcal{F}^*$, so $\mathrm{VC}^*(\mathcal{F}) \geq n$.

This proves the second inequality, the first one is proved similarly. □

Exercise 1.11. Show that this bound is optimal.

We describe a large source of examples of VC-families coming from model theory.

A *(first-order) structure* $\mathcal{M} = (M, R_1, R_2, \dots, f_1, f_2, \dots, c_1, c_2, \dots)$ consists of an underlying set M, together with some distinguished relations R_i (subsets of M^{n_i}, $n_i \in \mathbb{N}$), functions $f_i : M^{n_i} \to M$, and constants c_i (distinguished elements of M). We refer to the collection of all these relations, function symbols and constants as *the signature* of $\mathcal{M}$. For example, a group is naturally viewed as a structure $\left(G, \cdot, ^{-1}, 1\right)$, as well as a ring $(R, +, \cdot, 0, 1)$, ordered set $(X, <)$, graph (X, E), etc. A *formula* is an expression of the form $\psi(y_1, \dots, y_m) = Q_1 x_1 \dots Q_n x_n \phi(x_1, \dots, x_n; y_1, \dots, y_n)$, where $Q_i \in \{\forall, \exists\}$ and ϕ is given by a boolean combination of (superpositions of) the basic relations and functions (and $y_1, \dots, y_n$ are the *free variables* of ψ). We denote the set of all formulas by L. By a partitioned formula $\phi(\bar{x}, \bar{y})$ we mean a formula with its free variables partitioned into two groups $\bar{x}$ (object variables) and $\bar{y}$ (parameter variables). Given a partitioned formula $\phi(\bar{x}, \bar{y})$ and $\bar{b} \in M^{|\bar{y}|}$, we let $\phi\left(M^{|\bar{x}|}, \bar{b}\right)$ be the set of all $\bar{a} \in M^{|\bar{x}|}$ such that $\mathcal{M} \models \phi\left(\bar{a}, \bar{b}\right)$. Sets of this form are called *definable* (or ϕ*-definable*, in this case). We consider the family $\mathcal{F}_{\phi(\bar{x},\bar{y})}$ of subsets of $M^{|\bar{x}|}$ defined by $\mathcal{F}_{\phi(\bar{x},\bar{y})} = \left\{\phi\left(M^{|\bar{x}|}, \bar{b}\right) : \bar{b} \in M^{|\bar{y}|}\right\}$.

Example 1.12. Let $G = (V, E)$ be a graph. Then we can consider the formula $E(x, y)$, and for every $v \in V$, then set $E(V, v)$ is the set of all elements connected to v, and the family $\mathcal{F}_{E(x,y)}$ is the family of all neighborhoods of vertices in G. Similarly, for any $k \in \mathbb{N}$ let

$$d_k(x, y) := \exists z_0 \dots \exists z_{k-1} \left(z_0 = x \wedge z_{k-1} = y \wedge \bigwedge_{1 \leq i \leq k} E(z_{i-1}, z_i) \right).$$

Let $D_k(x, y) := \bigvee_{l \leq k} d_l(x, y)$. Then for every $v \in V$, $D_k(V, v)$ is the set of all vertices at distance at most k from v. But there is no first-order formula expressing that x is in the connected component of y (Exercise).

Theorem 1.13. *(Reduction to formulas in a single variable, Shelah) Let M be a first-order structure. Assume that for every partitioned formula $\phi(x, \bar{y})$ with x a* ***singleton****, the family $\mathcal{F}_\phi$ has finite VC dimension. Then for any $\phi(\bar{x}, \bar{y}) \in L$, the corresponding family $\mathcal{F}_\phi$ has finite VC dimension.*

Before proving it, we point out some examples of VC-families that it easily applies to.

Example 1.14. (Semialgebraic sets of bounded complexity) Recall that a set $X \subseteq \mathbb{R}^n$ is *semialgebraic* if it is given by a Boolean combination of polynomial equalities and inequalities.

We say that the *description complexity* of a semialgebraic set $X \subseteq \mathbb{R}^d$ is bounded by $t \in \mathbb{N}$ if $d \leq t$ and X can be defined as a Boolean combination of at most t

polynomial equalities and inequalities, such that all of the polynomials involved have degree at most t. For example, consider the family of all spheres in $\mathbb{R}^n$, or all cubes in $\mathbb{R}^n$, etc.

We claim that for any t, the family $\mathcal{F}_t$ of all semialgebraic sets of complexity $\leq t$ has finite VC-dimension. To see this, consider the field of real numbers as a first-order structure $\mathcal{M} = (\mathbb{R}, +, \times, 0, 1, <)$. Note that $\mathcal{F}_t$ is contained in the union of finitely many families of the form $\left\{\mathcal{F}_{\phi_i(\bar{x},\bar{y})} : i < t'\right\}$ where t' only depends on t (since there are only finitely many different polynomials of degree $\leq t$, up to varying coefficients, and only finitely many different Boolean combinations of size $\leq t$). So it is enough to show that every such family has finite VC-dimension (by Exercise 1.7).

By the classical result of Tarski, this structure $\mathcal{M}$ eliminates quantifiers, and so definable sets are precisely the semialgebraic ones. In particular, if we are given a formula of the form $\phi(x, \bar{y})$, for every $b \in M^{|\bar{y}|}$ the set $\phi(M, b)$ is just a union of at most n_ϕ intervals and points, where n_ϕ only depends on ϕ. As the collection of all intervals has finite VC-dimension, in view of Exercise 1.7 we have that for all formulas $\phi(x, \bar{y})$ with $|x| = 1$, $\mathcal{F}_\phi$ has finite VC-dimension. By Theorem 1.13 this implies that the same is true for all formulas.

Remark 1.15. In particular, let $X = \mathbb{R}$ and let $\mathcal{F}$ be the family of all convex n-gons. Then $\mathcal{F}$ has finite VC-dimension (one can verify that $\mathrm{VC}(\mathcal{F}) = 2n+1$) — compare this to Exercise 1.3.

Example 1.16. More generally, definable families in arbitrary *o*-minimal structures have finite VC-dimension.

A structure $\mathcal{M} = (M, <, \ldots)$ is *o*-minimal if every definable subset of M is a finite union of singletons and intervals (with endpoints in $M \cup \{\pm\infty\}$). From this assumption one obtains cell decomposition for definable subsets of M^n, for all (see [8] for a detailed treatment of o-minimality, or [6, Section 3] and references there for a quick introduction). Examples of o-minimal structures include (in each of these cases it is a highly non-trivial theorem): $(\mathbb{R}, +, \times, 0, 1, <)$, $\mathbb{R}_{\exp} = (\mathbb{R}, +, \times, e^x)$, $\mathbb{R}_{\mathrm{an}} = \left(R, +, \times, f\restriction_{[0,1]^k}\right)$ for f ranging over all functions real-analytic on some neighborhood of $[0,1]^k$, or the combination of both $\mathbb{R}_{\mathrm{an},\exp}$.

The same argument as in the previous example shows that all definable families in *o*-minimal structures have finite VC-dimension.

Exercise 1.17. Show that the family $\mathcal{F} = \{X_\lambda : \lambda \in \mathbb{R}\}$ where

$$X_\lambda := \left\{x \in \mathbb{R} : x \geq 0, 0 \leq \frac{x^\lambda - 1}{\lambda}\right\}$$

has finite VC-dimension.

Example 1.18. Definable families in stable structures.

The class of stable structures is well studied in model theory, originating from Morley's theorem and Shelah's work on classification theory. See e.g. [3] for more details. Examples of stable structures:

- $(\mathbb{C}, \times, +, 0, 1)$ (definable sets correspond to the constructible sets, i.e. Boolean combinations of algebraic sets),
- separably closed and differentially closed fields,
- arbitrary planar graphs $G = (V, E)$,

- abelian groups (viewed as structures in the pure group language $(G, \cdot, 1)$),
- [Z. Sela] non-commutative free groups (in the pure group language).

Example 1.19. [4] Let $(G, \cdot, <)$ be an arbitrary ordered abelian group. Then definable families of sets have finite VC-dimension. In particular, all definable families in Presburger arithmetic $(\mathbb{Z}, +, <)$ have finite VC-dimension.

Example 1.20. Let $(\mathbb{Q}_p, \times, +, 0, 1)$ be the field of p-adics. Using quantifier elimination results of Macintyre in this setting, one can show that again all definable families have finite VC-dimension.

Now we work towards a proof of Theorem 1.13. First recall a classical theorem of Ramsey generalizing the pigeon-hole principle.

Theorem 1.21. *(Ramsey) For any $l, k, n \in \mathbb{N}$ there is some $N \in \mathbb{N}$ such that $N \to (n)^k_l$ holds, i.e. for any set A with $|A| \geq N$ and any coloring of k-subsets of A in l colors, $f : \binom{A}{k} \to l$, there is a homogeneous subset $B \subseteq A$ of size $\geq n$, where* homogeneous *means that the value of f is constant on all k-subsets of B.*

Let us fix a first-order structure $\mathcal{M}$. Given a formula $\phi(\bar{x}_1, \ldots, \bar{x}_k)$ with $|\bar{x}_i| = m$ and a sequence $(\bar{a}_i : i \in I)$ of elements of M^m, where I is an arbitrary linearly ordered set, we say that this sequence is *ϕ-indiscernible* if for any $i_1 < \ldots < i_k$ and $j_0 < \ldots < j_k$ from I we have that $\phi(\bar{a}_{i_1}, \ldots, \bar{a}_{i_k}) \iff \phi(\bar{a}_{j_1}, \ldots, \bar{a}_{j_k})$ holds in $\mathcal{M}$. Given a set of such formulas Δ, we say that $(\bar{a}_i : i \in I)$ is *Δ-indiscernible* if it is ϕ-indiscernible for every $\phi \in \Delta$ such that the arities of its variables are compatible with the arities of the tuples in the sequence.

Note that if Δ is a finite set of formulas, then its closure under arbitrary repartitions of the variables of formulas in it is also finite. Thus, in all the arguments below we may assume that our finite sets of formulas are closed under choosing a different partition of the variables.

Lemma 1.22. *For every **finite** set of formulas Δ and every $n \in \mathbb{N}$ there is some $N \in \mathbb{N}$ such that every sequence $(\bar{a}_i : i < N)$ of m-tuples from M of length $\geq N$ contains a Δ-indiscernible subsequence of length $\geq n$.*

Proof. Follows by an iterated application of Ramsey theorem. Let n be fixed. Let's say $\Delta = \{\phi_1, \ldots, \phi_r\}$ with $\phi_i(\bar{x}_1, \ldots, \bar{x}_{k_i})$ for $i = 1, \ldots, r$. Let $k = \max\{k_i : 1 \leq i \leq r\}$. We find an N as wanted by induction on r.

Assume first that $\Delta = \{\phi_1\}$. Let N_0 be as given by Theorem 1.21 such that $N_0 \to (n)^k_2$ holds. Let $(\bar{a}_i : i < N_0)$ from M^m be arbitrary. Consider a coloring of all increasing k-tuples from this sequence in two colors, such that the color of the tuple $(\bar{a}_{i_1}, \ldots, \bar{a}_{i_k})$ corresponds to the truth value of $\phi(\bar{a}_{i_1}, \ldots, \bar{a}_{i_k})$. Then by Theorem 1.21 there is a homogeneous subset of size $\geq n$, which corresponds to a ϕ-indiscernible subsequence of length $\geq n$.

Assume now that we have found a number N_r that works for $\{\phi_1, \ldots, \phi_r\}$, and let $\Delta = \{\phi_1, \ldots, \phi_r, \phi_{r+1}\}$ be given. Let N_{r+1} be as given by Theorem 1.21 such that $N_{r+1} \to (N_r)^k_2$ holds. Again, it follows that an arbitrary sequence $(\bar{a}_i : i < N_{t+1})$ contains a ϕ_{r+1}-indiscernible subsequence $(\bar{a}_i : i \in I)$ for some $I \subseteq N, |I| \geq N_r$. Applying the inductive assumption to the sequence $(\bar{a}_i : i \in I)$ and $\{\phi_1, \ldots \phi_r\}$, we find a Δ-indiscernible subsequence of length $\geq n$. $\square$

Remark 1.23. The dual set system for $\left(M^{|x|}, \mathcal{F}_{\phi(\bar{x},\bar{y})}\right)$ is given by $\left(M^{|y|}, \mathcal{F}_{\phi^*(\bar{y},\bar{x})}\right)$ and $\phi^*(\bar{y},\bar{x}) = \phi(\bar{x},\bar{y})$, i.e. we exchange the roles of the object variables and the parameter variables in the partitioned formula (see Definition 1.8).

From now on, by the *VC-dimension of a formula* $\mathrm{VC}(\phi)$ we mean $\mathrm{VC}(\mathcal{F}_\phi)$ (and the same for the dual VC-dimension, shatter function, etc.).

Lemma 1.24. *Let $\phi(\bar{x},\bar{y})$ be a partitioned formula, and assume that $\mathrm{VC}(\phi) = \infty$. Then for any finite set of formulas Δ and every $n \in \mathbb{N}$ there is a Δ-indiscernible sequence $(\bar{a}_i : i < n)$ from $M^{|\bar{y}|}$ and a tuple $\bar{b} \in M^{|x|}$ so that $M \models \phi\left(\bar{b}, \bar{a}_i\right)$ if and only if i is even.*

Proof. Note that if $\mathrm{VC}(\mathcal{F}_\phi) = \infty$, then also $\mathrm{VC}(\mathcal{F}_{\phi^*}) = \infty$ (see Lemma 1.10 and Remark 1.23). This means that for any $N \in \mathbb{N}$ we can find some set $A = (\bar{a}_i : i < N)$ in $M^{|\bar{y}|}$ of size $\geq N$ which is shattered by the family $\mathcal{F}_{\phi^*} = \left\{\phi\left(\bar{b}, M^{|\bar{y}|}\right) : \bar{b} \in M^{|\bar{x}|}\right\}$. Then taking N sufficiently large, Lemma 1.22 ensures that A contains a Δ-indiscernible subsequence $(\bar{a}_i : i < n)$ of length n, and this sequence is still shattered by $\mathcal{F}_{\phi^*}$, in particular there is some $\bar{b}$ such that $\models \phi\left(\bar{b}, \bar{a}_i\right)$ iff i is even. □

Now we give a converse to Lemma 1.24. The point is that a formula of finite VC-dimension cannot cut out the set of even members from a *sufficiently* long and indiscernible sequence. What does "sufficiently" mean here?

Given $\phi(\bar{x},\bar{y})$, $n \in \mathbb{N}$ and $w \subseteq n$, let

$$\theta^\phi_w(\bar{x}, \bar{y}_0, \ldots, \bar{y}_{n-1}) := \bigwedge_{i \in w} \phi(\bar{x}, \bar{y}_i) \wedge \bigwedge_{i \in n \setminus w} \neg\phi(\bar{x}, \bar{y}_i),$$

$$\rho^\phi_w(\bar{y}_0, \ldots, \bar{y}_{n-1}) := \exists \bar{x} \theta^\phi_w(\bar{x}, \bar{y}_0, \ldots, \bar{y}_{n-1}),$$

$$\Delta^\phi_n := \left\{\rho^\phi_w : w \subseteq n\right\}.$$

Lemma 1.25. *Let $\phi(\bar{x},\bar{y})$ be a formula with $\mathrm{VC}^*(\phi) \leq d$. Then if $(\bar{a}_i : i < N)$ is a Δ^ϕ_d-indiscernible sequence from $M^{|\bar{y}|}$ and $\bar{b} \in M^{|\bar{x}|}$, there do not exist $i_0 < \ldots < i_{2d-1} < N$ so that $\phi\left(\bar{b}, \bar{a}_{i_j}\right)$ holds if and only if j is even.*

Proof. Assume that $(\bar{a}_i : i < N)$ is a Δ^ϕ_d-indiscernible sequence from $M^{|\bar{y}|}$ and there are $i_0 < \ldots < i_{2d-1}$ and $\bar{b} \in M^{|\bar{x}|}$ such that $M \models \phi\left(\bar{b}, \bar{a}_{i_j}\right)$ iff j is even. We claim that the sequence $(\bar{a}_i : i < d)$ is shattered by the family $\left\{\phi\left(\bar{b}, M^k\right) : \bar{b} \in M^{|\bar{x}|}\right\}$, which would give a contradiction.

Indeed, for any $w \subseteq d$, define a function $f_w : \{0, \ldots, d-1\} \to \{i_0, \ldots, i_{2d-1}\}$ by $f_w(j) := i_{2j}$ if $j \in w$ and $f_w(j) = i_{2j+1}$ if $j \notin w$. Now $\theta^\phi_w\left(\bar{b}, \bar{a}_{f_w(0)}, \ldots, \bar{a}_{f_w(d-1)}\right)$ holds in $\mathcal{M}$, so $\rho^\phi_w\left(\bar{a}_{f_w(0)}, \ldots, \bar{a}_{f_w(d-1)}\right)$ holds in $\mathcal{M}$. However, as $f_w(0) < \ldots < f_w(d-1)$ and $(\bar{a}_i : i < N)$ is Δ^ϕ_n-indiscernible, so in particular ρ^ϕ_w-indiscernible, this implies that $\rho^\phi_w(\bar{a}_0, \ldots, \bar{a}_{d-1})$ holds as well, i.e. there is some $\bar{b}' \in M^{|\bar{x}|}$ such that for all $i < d$ we have $\phi\left(\bar{b}, \bar{a}_i\right)$ holds in $\mathcal{M}$ iff $i \in w$. Since this works for any $w \subseteq d$, we conclude. □

Now we amplify this by finding a large chunk of our sequence that is indiscernible "over $\bar{b}$".

Lemma 1.26. *Let Δ be a finite set of formulas such that for every $\phi(\bar{x},\bar{y}) \in \Delta$ with $|\bar{x}| \leq l$ we have $\mathrm{VC}^*(\phi) \leq d$. Then for any $n \in \mathbb{N}$ there is a finite set of* *formulas* Δ' *and* $N \in \mathbb{N}$ *such that if* $(\bar{a}_i : i < N)$ *is a* Δ'*-indiscernible sequence and* $\bar{b} \in M^{|x|}$ *is an arbitrary tuple of length* $\leq l$*, then for some* ***interval*** $I \subseteq N$ *with* $|I| \geq n$*, the sequence of* $(|x| + |y|)$*-tuples* $\left(\bar{a}_i \bar{b} : i \in I\right)$ *is* Δ*-indiscernible.*

Proof. Let us take $\Delta' = \bigcup_{\phi \in \Delta} \Delta_d^{\phi}$, let $N' = |\Delta| (2d - 1)$ and $N = N'n$.

Let $(\bar{a}_i : i < N)$ and $\bar{b} \in M^{|x|}$ be arbitrary.

Let us partition N into N'-many consecutive intervals $(I_j : j < N')$, each of length n. Assume that for each of those intervals, the conclusion of the lemma does not hold. This means that for each $j < N'$ there is some $\phi_j \in \Delta$ and some $i_0^j < \ldots < i_{m_j-1}^j, h_0^j < \ldots < h_{m_j-1}^j \in I_j$ such that $M \models \phi_j \left(\bar{b}, \bar{a}_{i_0^j}, \ldots, \bar{a}_{i_{m_j-1}^j}\right) \wedge \neg\phi_j \left(\bar{b}, \bar{a}_{h_0^j}, \ldots, \bar{a}_{h_{m_j-1}^j}\right)$.

By the choice of N', throwing away some intervals, we may assume that ϕ_j is constant for all $j < 2d - 1$, say $\phi_j = \phi'$ and $m_j = m'$.

For $j < 2d - 1$, we consider the sequence of $(m' |\bar{y}|)$-tuples $\bar{a}'_j$ defined by taking $\bar{a}'_j$ to be $\bar{a}_{i_0^j} \ldots \bar{a}_{i_{m'-1}^j}$ if j is even, and to be $\bar{a}_{h_0^j} \ldots \bar{a}_{h_{m'-1}^j}$ if j is odd.

Observe that the sequence $\left(\bar{a}'_j : j < 2d - 1\right)$ is still $\Delta_d^{\phi'}$-indiscernible since the original sequence was, and $\phi' \left(\bar{b}, \bar{a}'_j\right)$ holds iff j is even. By Lemma 1.25 this contradicts the assumption that $\mathrm{VC}^* (\phi') \leq d$. □

Iterating Lemma 1.26, we obtain:

Corollary 1.27. *Assume that all formulas in* M *in a single object variable have finite VC-dimension. Then for any* Δ *a finite set of formulas and* $n, l \in \mathbb{N}$ *there is a finite set of formulas* Δ' *and* $N \in \mathbb{N}$ *such that if* $(\bar{a}_i : i < N)$ *is a* Δ'*-indiscernible sequence and* $\bar{b} \in M^l$ *is an arbitrary* l*-tuple, then for some interval* $I \subseteq N$ *with* $|I| \geq n$*, the sequence of* $(l |\bar{a}_i|)$*-tuples* $\left(\bar{a}_i \bar{b} : i \in I\right)$ *is* Δ*-indiscernible.*

Proof. Define recursively $\Delta_0 = \Delta, N_0 = n$ and take N_{j+1}, Δ_{j+1} to be the N and Δ' given by Lemma 1.26 for $n = N_j, \Delta = \Delta_j$ and $d = \max \{\mathrm{VC}^* (\phi (x, \bar{y})) : \phi (x, \bar{y}) \in \Delta_j\}$. Let $\tilde{N} := N_{l-1}$ and $\tilde{\Delta} = \Delta_{l-1}$.

Let now $\left(\bar{a}_i : i < \tilde{N}\right)$ be $\tilde{\Delta}$-indiscernible and let $\bar{b} \in M^l$ be arbitrary, say $\bar{b} = b_0 \ldots b_{l-1}$. Applying Lemma 1.26 to $\left(\bar{a}_i : i < \tilde{N}\right)$ and b_0 we find some interval $I_{l-1} \subseteq \tilde{N}$ with $|I_{l-1}| \geq N_{l-1}$ such that the sequence $(\bar{a}_i b_0 : i \in I_{l-1})$ is Δ_{l-1}-indiscernible. Again applying Lemma 1.26 to this sequence of thicker tuples $(\bar{a}_i b_0 : i \in I_{l-1})$ and b_1, we find some interval $I_{l-2} \subseteq I_{l-1}$ with $|I_{l-2}| \geq N_{l-2}$ such that the sequence $(\bar{a}_i b_0 b_1 : i \in I_{l-2})$ is Δ_{l-2}-indiscernible. Continuing in this fashion we finally find an interval $I_0 \subseteq N$ with $|I_0| \geq n$ such that the sequence $(\bar{a}_i b_0 \ldots b_{l-1} : i \in I_0)$ is Δ-indiscernible, as wanted. □

Finally, to prove Theorem 1.13, assume that some formula $\phi (\bar{x}, \bar{y})$ has infinite VC-dimension. Let Δ be the finite collection of formulas given by taking arbitrary partitions of the variables in ϕ. Let Δ' and N be as given by Corollary 1.27 for Δ, $l = |\bar{x}|$ and $n = 2$.

Now on the one hand, by Lemma 1.24 we can find a Δ-indiscernible sequence $(\bar{a}_i : i < N)$ and $\bar{b}$ such that $\phi \left(\bar{b}, \bar{a}_i\right)$ holds iff i is even. On the other hand, by Corollary 1.27 we can find an interval $I \subseteq N$ of length ≥ 2 such that the sequence

$\left\{\bar{a}_i\bar{b} : i \in I\right\}$ is Δ-indiscernible, so for all $i, i' \in I$ we have $\phi\left(\bar{b}, a_i\right) \iff \phi\left(\bar{b}, a_{i'}\right)$. As I must contain both an even and an odd indices, we get a contradiction.

Remark 1.28. The bounds on the VC-dimension given by this proof are astronomical as we have used Ramsey's theorem iteratively. In most specific cases it is possible to obtain much stronger bounds.

E.g., let $\mathcal{F}_{k,m,n}$ be the family of all semialgebraic subsets of $\mathbb{R}^n$ that can be represented as a Boolean combination of at most k sets of the form $\{\bar{x} \in \mathbb{R}^n : f_j(\bar{x}) > 0\}$ where the functions f_j are real polynomials of maximum degree m. Then $\mathrm{VC}(\mathcal{F}_{k,m,n}) \leq 2k\binom{m+n}{m}\log\left(k(k+1)\binom{m+n}{m}\right)$, and this differs only by a logarithmic factor from the known lower bound (see [2] for the details).

Historic remarks. In model theory, a partitioned formula $\phi(\bar{x}, \bar{y})$ is called *NIP* (No Independence Property) if the family $\mathcal{F}_\phi$ has finite VC-dimension. A structure $\mathcal{M}$ is NIP if all definable families in it are NIP. Such structures were defined by Shelah around the same time as Vapnik and Chervonenkis have defined their dimension for entirely different purposes, and are currently being actively studied in model theory (see [7] for a survey). The original proof of Lemma 1.13 by Shelah used forcing and absoluteness (see [3] for some more details). It was first finitized by Laskowski [5], and further simplified by Poizat, Adler and others [1]. We avoid the use of compactness and give a purely combinatorial proof which in principle gives explicit bounds, using compactness we could have avoided micromanaging all the numerical parameters involved.

2. VC-density

Let $(X, \mathcal{F})$ be a family of finite VC-dimension. By Lemma 1.5 we know that $\pi_{\mathcal{F}}(n) = O\left(n^d\right)$, but how exactly can $\pi_{\mathcal{F}}(n)$ grow?

Definition 2.1. We define the *VC-density* of $\mathcal{F}$ to be $\mathrm{vc}(\mathcal{F}) = \limsup_{n\to\infty} \frac{\log(\pi_{\mathcal{F}}(n))}{\log n}$. In other words, $\mathrm{vc}(\mathcal{F})$ is the infimum over all real numbers $r \geq 0$ for which we have $\pi_{\mathcal{F}}(n) = O(n^r)$. Similarly, we define the *VC-codensity* as $\mathrm{vc}^*(\mathcal{F}) = \mathrm{vc}(\mathcal{F}^*)$.

We have $\mathrm{vc}(\mathcal{F}) < \infty \iff \mathrm{VC}(\mathcal{F}) < \infty$ and by Lemma 1.5 we have $\mathrm{vc}(\mathcal{F}) \leq \mathrm{VC}(\mathcal{F})$. Often they coincide.

Exercise 2.2.
1. Let $\mathcal{F} = \binom{X}{\leq d}$. Show that $\mathrm{VC}(\mathcal{F}) = \mathrm{vc}(\mathcal{F}) = d$.
2. Let $X = \mathbb{R}$, $k \geq 1$ and let $\mathcal{F}$ be the collection whose members are the unions of k disjoint open intervals in $\mathbb{R}$. Show that $\mathrm{VC}(\mathcal{F}) = \mathrm{vc}(\mathcal{F}) = 2k$, and in fact $\pi_{\mathcal{F}}(n) = \binom{n}{\leq 2k}$ for each n.

In some sense, the VC-density is a more rigid version of the VC-dimension ignoring small noise.

Example 2.3. Let $(X, \mathcal{F})$ be a set system with $\mathrm{vc}(\mathcal{F}) < k$. Let $X' = X \cup Y$ where Y is a set of size k disjoint from X. Let $\mathcal{F}' = \mathcal{F} \cup \mathcal{P}(Y)$. Then $\mathrm{VC}(\mathcal{F}') = k$, but $\mathrm{vc}(\mathcal{F}') = \mathrm{vc}(\mathcal{F}) < k$.

Exercise 2.4. (Some properties of VC-density)
1. Let $(X, \mathcal{F})$ be a set system, let X' be an infinite set and $f : X' \to X$ be a map. Let $f^{-1}(\mathcal{F}) := \left\{f^{-1}(S) : S \in \mathcal{F}\right\}$. Then $\pi_{f^{-1}(\mathcal{F})} \leq \pi_F$ for all n, with equality if f is surjective. In particular, $\mathrm{vc}\left(f^{-1}(\mathcal{F})\right) \leq \mathrm{vc}(\mathcal{F})$.

(2) If $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2$ then $\mathrm{vc}(\mathcal{F}) = \max\{\mathrm{vc}(\mathcal{F}_1), \mathrm{vc}(\mathcal{F}_2)\}$, and $\mathrm{vc}(X \setminus \mathcal{F}) = \mathrm{vc}(\mathcal{F})$ where $X \setminus \mathcal{F} := \{X \setminus S : S \in \mathcal{F}\}$.
(3) Given two set systems $(X_1, \mathcal{F}_1), (X_2, \mathcal{F}_2)$, consider the set system $(X_1 \times X_2, \mathcal{F})$ with $\mathcal{F} := (S_1 \times S_2 : S_i \in \mathcal{F}_i)$. Then $\mathrm{vc}(\mathcal{F}) \leq \mathrm{vc}(\mathcal{F}_1) + \mathrm{vc}(\mathcal{F}_2)$.
(4) $\mathcal{F}$ is finite if and only if $\mathrm{vc}(\mathcal{F}) = 0$.

In fact, the last claim can be strengthened.

Proposition 2.5. *If $\mathrm{vc}(\mathcal{F}) < 1$, then $\mathcal{F}$ is finite (and so VC-density never takes values in the interval $(0,1)$).*

Proof. Assume that we have a set system $(X, \mathcal{F})$ with $\mathrm{vc}(\mathcal{F}) < 1$. Let $\mathcal{F}_1 = \mathcal{F} \cup (X \setminus \mathcal{F})$. By Exercise 2.4, $\mathrm{vc}(\mathcal{F}) = \mathrm{vc}(\mathcal{F}_1) < 1$. This means that there is some $c \in \mathbb{N}$ and $s \in (0,1)$ such that $|A \cap \mathcal{F}_1| \leq c|A|^s$ for all $A \subseteq X$. In particular, for $m \in \mathbb{N}$ large enough we have that if $|A| = m$ then $|A \cap \mathcal{F}_1| < m$. Fix some m like that.

Claim. Fix $k \in \mathbb{N}$, and let $\mathcal{F}_k := \left\{\bigcup_{i<k} S_i : S_i \in \mathcal{F}_1\right\}$. Now for any $\mathcal{F}' \subseteq \mathcal{F}_k$ there is some $\mathcal{S} \subseteq \mathcal{F}'$ with $|\mathcal{S}| < m^k$ such that $\bigcap \mathcal{S} \subseteq \bigcap \mathcal{F}'$.

Proof. Assume that there is no such $\mathcal{S}$ for some $\mathcal{F}'$. Then there are some sets $T_i \in \mathcal{F}', i = 1, \ldots, m^k$, such that $\left(\bigcap_{j<i} T_j\right) \setminus T_i \neq \emptyset$ for all $i = 1, \ldots, m^k$. Take $A = \{x_1, \ldots, x_{m^k}\}$ with $x_i \in \left(\bigcap_{j<i} T_j\right) \setminus T_i$. We then have $|A \cap \mathcal{F}_k| \geq m^k > |A \cap \mathcal{F}_1|^k$ — which is impossible as every element in $\mathcal{F}_k$ is determined by k elements in $\mathcal{F}_1$.

Let now $\mathcal{G}$ be an arbitrary finite subfamily of $\mathcal{F}_1$. Let $C \subseteq X$ be an atom of the Boolean algebra generated by $\mathcal{G}$ (i.e. $C = \bigcap_{S \in \mathcal{G}} S^f$ for some $f : \mathcal{G} \to \{0,1\}$, where $S^f = S$ if $f(S) = 1$ and $S^f = X \setminus S$ if $f(S) = 0$).

Let $\mathcal{G}_C$ be the collection of all elements of $\mathcal{G} \cup (X \setminus \mathcal{G})$ that contain C. Applying the claim with $k = 1$ to $\mathcal{G}_C$, we see that C is given by the intersection of some $(m-1)$ elements of $\mathcal{F}_1$.

Let now $\mathcal{H}$ be the collection of all sets of the form $X \setminus C$ with C an atom of the Boolean algebra generated by $\mathcal{G}$. We have $\mathcal{H} \subseteq \mathcal{F}_{m-1}$ by the previous remark.

Note that $\bigcap \mathcal{H} = \emptyset$. Applying the claim again with $k = m-1$ to $\mathcal{H}$, it follows that $\mathcal{H}$ contains some subset $\mathcal{H}'$ such that $|\mathcal{H}'| \leq N = m^{m-1} - 1$ and $\bigcap \mathcal{H}' = \emptyset$.

It follows that $\mathcal{G}$ has at most N atoms, thus $|\mathcal{G}| \leq 2^N$. Since $\mathcal{G}$ was an arbitrary finite subfamily of $\mathcal{F}_1$, it follows that $|\mathcal{F}_1| \leq 2^N$ as well. □

References

[1] H. Adler. An introduction to theories without the independence property. *Archive for Mathematical Logic*, 2008.
[2] Shai Ben-David and Michael Lindenbaum. Localization vs. identification of semi-algebraic sets. *Machine Learning*, 32(3):207–224, 1998.
[3] Artem Chernikov. Lecture notes on stability theory (math 285D). `http://www.math.ucla.edu/~chernikov/teaching/StabilityTheory285D/StabilityNotes.pdf`.
[4] Yuri Gurevich and Peter H Schmitt. The theory of ordered abelian groups does not have the independence property. *Transactions of the American Mathematical Society*, 284(1):171–182, 1984.
[5] Michael C Laskowski. Vapnik-chervonenkis classes of definable sets. *Journal of the London Mathematical Society*, 2(2):377–384, 1992.
[6] Thomas Scanlon. o-minimality as an approach to the andré-Oort conjecture. *Panoramas et Synthèses, to appear.*
[7] Pierre Simon. *A guide to NIP theories*, volume 44. Cambridge University Press, 2015.

[8] Lou Van den Dries. *Tame topology and o-minimal structures*, volume 248. Cambridge university press, 1998.