# Vapnik-Chervonenkis Density in Model Theory

Matthias Aschenbrenner
University of California, Los Angeles
(joint with A. Dolich, D. Haskell, D. Macpherson, and S. Starchenko)

UCLA

# VC dimension and VC density

Let $(X, \mathcal{S})$ be a set system, i.e., $X$ is a set (the **base set**), and $\mathcal{S}$ is a collection of subsets of $X$. (We sometimes also speak of a **set system $\mathcal{S}$ on $X$.**)

Given $A \subseteq X$, we let

$$\mathcal{S} \cap A := \{S \cap A : S \in \mathcal{S}\}$$

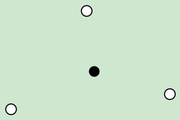
and call $(A, \mathcal{S} \cap A)$ **the set system on $A$ induced by $\mathcal{S}$.**

We say $A$ is **shattered by $\mathcal{S}$** if $\mathcal{S} \cap A = 2^A$.

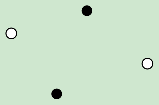
If $\mathcal{S} \neq \emptyset$, then we define the **VC dimension of $\mathcal{S}$,** denoted by $\mathrm{VC}(\mathcal{S})$, as the supremum (in $\mathbb{N} \cup \{\infty\}$) of the sizes of all finite subsets of $X$ shattered by $\mathcal{S}$. We also decree $\mathrm{VC}(\emptyset) := -\infty$.

## Examples

1. $X = \mathbb{R}$, $\mathcal{S} =$ all unbounded intervals. Then $\mathrm{VC}(\mathcal{S}) = 2$.
2. $X = \mathbb{R}^2$, $\mathcal{S} =$ all halfspaces. Then $\mathrm{VC}(\mathcal{S}) = 3$.



One point in the convex hull of the others

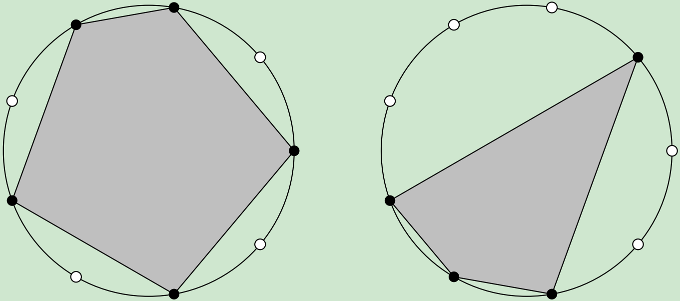


No point in the convex hull of the others

3. Let $\mathcal{S} =$ half spaces in $\mathbb{R}^d$. Then $\mathrm{VC}(\mathcal{S}) = d + 1$.
(The inequality $\leqslant$ follows from *Radon's Lemma*.)

## Examples (continued)

4. $X = \mathbb{R}^2$, $\mathcal{S} =$ *all* convex polygons. Then $\mathrm{VC}(\mathcal{S}) = \infty$.



(But $\mathrm{VC}(\{$convex $n$-gons in $\mathbb{R}^2\}) = 2n + 1$.)

# VC dimension and VC density

The function

$$n \mapsto \pi_{\mathcal{S}}(n) := \max\left\{ |\mathcal{S} \cap A| : A \in \binom{X}{n} \right\} : \mathbb{N} \to \mathbb{N}$$

is called the **shatter function of $\mathcal{S}$.** Then

$$\mathrm{VC}(\mathcal{S}) = \sup\left\{ n : \pi_{\mathcal{S}}(n) = 2^n \right\}.$$

One says that $\mathcal{S}$ is a **VC class** if $\mathrm{VC}(\mathcal{S}) < \infty$.

The notion of VC dimension was introduced by Vladimir Vapnik and Alexey Chervonenkis in the early 1970s, in the context of computational learning theory.

A surprising dichotomy holds for $\pi_{\mathcal{S}}$:

### The Sauer-Shelah dichotomy

*Either*

- $\pi_{\mathcal{S}}(n) = 2^n$ for every $n$ (if $\mathcal{S}$ is not a VC class),

*or*

- $\pi_{\mathcal{S}}(n) \leqslant \binom{n}{\leqslant d} := \binom{n}{0} + \cdots + \binom{n}{d}$ where $d = \mathrm{VC}(\mathcal{S}) < \infty$.

One may now define the **VC density of $\mathcal{S}$** as

$$\mathrm{vc}(\mathcal{S}) = \begin{cases} \inf\{r \in \mathbb{R}^{>0} : \pi_{\mathcal{S}}(n) = O(n^r)\} & \text{if } \mathrm{VC}(\mathcal{S}) < \infty \\ \infty & \text{otherwise.} \end{cases}$$

We also define $\mathrm{vc}(\emptyset) := -\infty$.

# VC dimension and VC density

## Examples

1. $\mathcal{S} = \binom{X}{\leqslant d}$. Then $\mathrm{VC}(\mathcal{S}) = \mathrm{vc}(\mathcal{S}) = d$; in fact $\pi_{\mathcal{S}}(n) = \binom{n}{\leqslant d}$.
2. $\mathcal{S} =$ half spaces in $\mathbb{R}^d$. Then $\mathrm{VC}(\mathcal{S}) = d + 1$, $\mathrm{vc}(\mathcal{S}) = d$.

VC density is often the right measure for the combinatorial complexity of a set system.

## Some basic properties:

- $\mathrm{vc}(\mathcal{S}) \leqslant \mathrm{VC}(\mathcal{S})$, and if one is finite then so is the other;
- $\mathrm{VC}(\mathcal{S}) = 0 \iff |\mathcal{S}| = 1$;
- $\mathcal{S}$ is finite $\iff \mathrm{vc}(\mathcal{S}) = 0 \iff \mathrm{vc}(\mathcal{S}) < 1$;
- $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \Rightarrow \mathrm{vc}(\mathcal{S}) = \max\{\mathrm{vc}(\mathcal{S}_1), \mathrm{vc}(\mathcal{S}_2)\}$.

Let $X$ be a set (possibly finite). Given $A_1, \ldots, A_n \subseteq X$, denote by $S(A_1, \ldots, A_n)$ the set of atoms of the Boolean subalgebra of $2^X$ generated by $A_1, \ldots, A_n$: those subsets of $X$ of the form

$$\bigcap_{i \in I} A_i \cap \bigcap_{i \notin I} X \setminus A_i \qquad \text{where } I \subseteq \{1, \ldots, n\}$$

which are *non-empty* (= "the non-empty sets in the Venn diagram of $A_1, \ldots, A_n$").

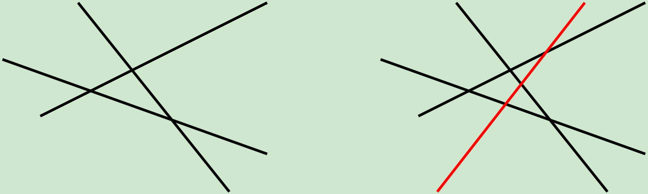
Suppose now that $\mathcal{S}$ is a set system on $X$. We define

$$n \mapsto \pi^*_{\mathcal{S}}(n) := \max \left\{ |S(A_1, \ldots, A_n)| : A_1, \ldots, A_n \in \mathcal{S} \right\} : \mathbb{N} \to \mathbb{N}.$$

We say that $\mathcal{S}$ is **independent** (in $X$) if $\pi^*_{\mathcal{S}}(n) = 2^n$ for every $n$, and **dependent** (in $X$) otherwise.

**Example ($X = \mathbb{R}^2$, $\mathcal{S} =$ half planes in $\mathbb{R}^2$)**

$\pi^*_{\mathcal{S}}(n) = \begin{cases} \text{maximum number of regions into which } n \text{ half} \\ \text{planes partition the plane.} \end{cases}$



Adding one half plane to $n-1$ given half planes divides at most $n$ of the existing regions into $2$ pieces. So $\pi^*_{\mathcal{S}}(n) = O(n^2)$.

The function $\pi^*_{\mathcal{S}}$ is called the **dual shatter function of $\mathcal{S}$.**

# VC duality

Let $X$, $Y$ be infinite sets, $\Phi \subseteq X \times Y$ a binary relation. Put

$$\mathcal{S}_\Phi := \{\Phi_y : y \in Y\} \subseteq 2^X \quad \text{where } \Phi_y := \{x \in X : (x,y) \in \Phi\},$$

and

$$\begin{aligned} \pi_\Phi &:= \pi_{\mathcal{S}_\Phi}, & \pi^*_\Phi &:= \pi^*_{\mathcal{S}_\Phi}, \\ \mathrm{VC}(\Phi) &:= \mathrm{VC}(\mathcal{S}_\Phi), & \mathrm{vc}(\Phi) &:= \mathrm{vc}(\mathcal{S}_\Phi). \end{aligned}$$

We also write

$$\Phi^* \subseteq Y \times X := \big\{(y,x) \in Y \times X : (x,y) \in \Phi\big\}.$$

In this way we obtain two set systems: $(X, \mathcal{S}_\Phi)$ and $(Y, \mathcal{S}_{\Phi^*})$

Given a finite set $A \subseteq X$ we have a bijection

$$A' \mapsto \bigcap_{x \in A'} \Phi^*_x \cap \bigcap_{x \in A \setminus A'} Y \setminus \Phi^*_x : \quad \mathcal{S}_\Phi \cap A \to S(\Phi^*_x : x \in A).$$

Hence $\pi_\Phi = \pi^*_{\Phi^*}$ and $\pi_{\Phi^*} = \pi^*_\Phi$, and thus

$$\mathcal{S}_\Phi \text{ is a VC class} \iff \mathcal{S}_{\Phi^*} \text{ is dependent,}$$
$$\mathcal{S}_{\Phi^*} \text{ is a VC class} \iff \mathcal{S}_\Phi \text{ is dependent.}$$

Moreover (first noticed by Assouad):

$$\mathcal{S}_\Phi \text{ is a VC class} \iff \mathcal{S}_{\Phi^*} \text{ is a VC class.}$$

We fix:

$\mathcal{L}$: a first-order language,
$x = (x_1, \ldots, x_m)$: object variables,
$y = (y_1, \ldots, y_n)$: parameter variables,
$\varphi(x; y)$: a partitioned $\mathcal{L}$-formula,
$\boldsymbol{M}$: an infinite $\mathcal{L}$-structure, and
$T$: a complete $\mathcal{L}$-theory without finite models.

The set system (on $M^m$) associated with $\varphi$ in $\boldsymbol{M}$:

$$\mathcal{S}_\varphi^{\boldsymbol{M}} := \{\varphi^{\boldsymbol{M}}(M^m; b) : b \in M^n\}$$

If $\boldsymbol{M} \equiv \boldsymbol{N}$, then $\pi_{\mathcal{S}_\varphi^{\boldsymbol{M}}} = \pi_{\mathcal{S}_\varphi^{\boldsymbol{N}}}$. So, picking $\boldsymbol{M} \models T$ arbitrary, set

$$\pi_\varphi := \pi_{\mathcal{S}_\varphi^{\boldsymbol{M}}}, \quad \mathrm{VC}(\varphi) := \mathrm{VC}(\mathcal{S}_\varphi^{\boldsymbol{M}}), \quad \mathrm{vc}(\varphi) := \mathrm{vc}(\mathcal{S}_\varphi^{\boldsymbol{M}}).$$

The *dual* of $\varphi(x;y)$ is $\varphi^*(y;x) := \varphi(x;y)$. Put

$$\mathrm{VC}^*(\varphi) := \mathrm{VC}(\varphi^*), \qquad \mathrm{vc}^*(\varphi) := \mathrm{vc}(\varphi^*).$$

We have $\pi^*_\varphi = \pi_{\varphi^*}$, hence $\mathrm{VC}^*(\varphi)$ and $\mathrm{vc}^*(\varphi)$ can be computed using the dual shatter function of $\varphi$.

If $\mathrm{VC}(\varphi) < \infty$ then we say that $\varphi$ is **dependent** in $T$. The theory $T$ does **not have the independence property** (is **NIP**) if every partitioned $\mathcal{L}$-formula is dependent in $T$.

An important theorem of Shelah (given other proofs by Laskowski and others) says that for $T$ to be NIP it is enough for for every $\mathcal{L}$-formula $\varphi(x;y)$ with $|x| = 1$ to be dependent.

Many (but not all) well-behaved theories arising naturally in model theory are NIP.

## Some questions about $\mathrm{vc}$ in model theory

1. Possible values of $\mathrm{vc}(\varphi)$. There exists a formula $\varphi(x; y)$ in $\mathcal{L}_{\mathrm{rings}}$ with $|y| = 4$ such that

$$\mathrm{vc}^{\mathrm{ACF}_0}(\varphi) = \tfrac{4}{3}; \quad \mathrm{vc}^{\mathrm{ACF}_p}(\varphi) = \tfrac{3}{2} \text{ for } p > 0.$$

   We do not know an example of a formula $\varphi$ in a NIP theory with $\mathrm{vc}(\varphi) \notin \mathbb{Q}$.

2. Growth of $\pi_\varphi$. There is an example of an $\omega$-stable $T$ and an $\mathcal{L}$-formula $\varphi(x; y)$ with $|y| = 2$ and

$$\pi_\varphi(n) = \tfrac{1}{2} n \log n \, (1 + o(1)).$$

3. Uniform bounds on $\mathrm{vc}(\varphi)$.

# Uniform bounds on VC density

Some reasons why it should be interesting to obtain bounds on $\mathrm{vc}(\varphi)$ in terms of $|y| =$ number of free parameters:

1. uniform bounds on VC density often "explain" why certain bounds on the complexity of geometric arrangements, used in computational geometry, are polynomial in the number of objects involved;
2. connections to strengthenings of the NIP concept: if $\mathrm{vc}(\varphi) < 2$ for each $\varphi(x; y)$ with $|y| = 1$ then $T$ is *dp-minimal*.

# Uniform bounds on VC density

**Theorem**

*Suppose $T$ expands the theory of linearly ordered sets, and assume that $T$ is* **weakly o-minimal,** *i.e., in every $\boldsymbol{M} \models T$, every definable subset of $M$ is a finite union of convex subsets of $M$. Then for each $\varphi(x; y)$ we have $\pi_\varphi(t) = O(t^{|y|})$, hence $\mathrm{vc}(\varphi) \leqslant |y|$.*

- Generalizes results due to Karpinski-Macintyre and Wilkie;
- Idea of the proof:
  - generalize definition of $\pi^*_\varphi$ to finite sets $\Delta$ of formulas instead of a single $\varphi$;
  - the number of parameters needed in a *uniform definition of $\Delta$-types* over finite parameter sets yields a bound on $\pi^*_\Delta(t)$;
  - reduce to the case where $|x| = 1$ and each instance of $\varphi \in \Delta$ defines an initial segment, in which case each finite $\Delta$-type can be defined by a single parameter.

# Uniform bounds on VC density

Interesting classes of NIP theories are provided by certain valued fields. By a non-trivial elaboration of our methods:

**Theorem**

*Suppose $M = \mathbb{Q}_p$ is the field of $p$-adic numbers, construed as a structure in the language of rings. Then $\mathrm{vc}(\varphi) \leqslant 2|y| - 1$.*

We also have uniform bounds on VC density (obtained by other techniques) for certain stable structures; e.g., we characterize those abelian groups for which we have such uniform bounds.

There are many open questions in this subject.

**Open question**

If there is some $d_1$ such that $\mathrm{vc}(\varphi) \leqslant d_1$ for each $\varphi(x; y)$ with $|y| = 1$, is there is some $d_m$ such that $\mathrm{vc}(\varphi) \leqslant d_m$ for each $\varphi(x; y)$ with $|y| = m$?

# From Lipschitz maps to VC density

## Question

Let $f\colon A \to \mathbb{R}^n$, $A \subseteq \mathbb{R}^m$, be $L$-**Lipschitz** (where $L \in \mathbb{R}^{\geqslant 0}$), i.e.,

$$||f(x) - f(y)|| \leqslant L \cdot ||x - y|| \qquad \text{for all } x, y \in A.$$

Can one extend $f$ to an $L$-Lipschitz map $\mathbb{R}^m \to \mathbb{R}^n$?

## Kirszbraun (1934): *yes* for all $n$

There always exists an $L$-Lipschitz extension $\mathbb{R}^m \to \mathbb{R}^n$ of $f$.

The usual proofs of this theorem all use some sort of transfinite induction.

### Question (Chris Miller)

Let $f\colon A \to \mathbb{R}^n$, $A \subseteq \mathbb{R}^m$, be $L$-Lipschitz and *semialgebraic*. Is there a *semialgebraic* $L$-Lipschitz map $\mathbb{R}^m \to \mathbb{R}^n$ extending $f$?

More generally, one may ask this for Lipschitz maps definable in an o-minimal expansion $\boldsymbol{R} = (R, 0, 1, +, \times, <, \dots)$ of a real closed field $R$, instead of $\overline{\mathbb{R}} = (\mathbb{R}, 0, 1, +, \times, <)$.

Why is this extra generality interesting?

- no "higher-order" Tarski principle for transfer from o-minimal expansions of $\overline{\mathbb{R}}$ to $\boldsymbol{R}$;
- brings out the inherent uniformities in the construction.

In fact, o-minimality even turns out to be an unnecessarily strong assumption.

# From Lipschitz maps to VC density

### Theorem A (A.-Fischer, Proc. LMS 2011)

Let $\boldsymbol{R} = (R, 0, 1, +, \times, <, \dots)$ be a **definably complete** expansion of an ordered field: every non-empty definable subset of $R$ which is bounded from above has a supremum. Then every definable $L$-Lipschitz map $A \to R^n$ ($A \subseteq R^m$, $L \in R^{\geqslant 0}$) has a definable $L$-Lipschitz extension $R^m \to R^n$.

The proof of this theorem used convex analysis and is based on a relationship between Lipschitz maps and monotone set-valued maps (Minty; more recently, Bauschke & Wang).

Another crucial ingredient (in the case where $R \neq \mathbb{R}$) is a definable version of a classical theorem of Helly:

**Theorem B (A.-Fischer, Proc. LMS 2011)**

Let $\boldsymbol{R}$ be a definably complete expansion of an ordered field. Let $\mathcal{C}$ be a definable family of closed bounded *convex* subsets of $R^n$. Suppose $\mathcal{C}$ is $(n+1)$-**consistent**:

$$\bigcap \mathcal{C}' \neq \emptyset \qquad \text{for all } \mathcal{C}' \subseteq \mathcal{C} \text{ with } |\mathcal{C}'| \leqslant n+1.$$

Then $\bigcap \mathcal{C} \neq \emptyset$.

Our proof of this theorem uses an optimization argument.

S. Starchenko pointed out that in the case of an o-minimal $\boldsymbol{R}$, our theorem follows from an analysis of the model-theoretic notion of *forking* in o-minimal structures due to A. Dolich.

# From Lipschitz maps to VC density

A subset $T$ of $X$ is called a **transversal** of a set system $\mathcal{S}$ on $X$ if every member of $\mathcal{S}$ contains an element of $T$.

### Theorem (Dolich '04, made explicit by Peterzil & Pillay '07)

Let $\boldsymbol{R}$ be an o-minimal expansion of a real closed field, and let $\mathcal{C} = \{C_a\}_{a \in A}$ be a definable family of closed and bounded subsets of $R^n$ parameterized by a subset $A$ of $R^m$. If $\mathcal{C}$ is $N(m,n)$-consistent, where

$$N(m,n) = (1+2^m) \cdot (1+2^{2^m}) \cdots \qquad (n \text{ factors}),$$

then $\mathcal{C}$ has a finite transversal.

# From Lipschitz maps to VC density

**Question**

Can one do better than the bound $N(m, n)$?

**Theorem (Matoušek, 2004)**

*Let $(X, \mathcal{S})$ be a set system of finite dual VC density $\mathrm{vc}^*(\mathcal{S})$. Suppose $\mathcal{S}$ is $d$-consistent, where $d > \mathrm{vc}^*(\mathcal{S})$. Assume that $X$ comes equipped with a topology making all sets in $\mathcal{S}$ compact. Then $\mathcal{S}$ has a finite transversal.*

**Corollary**

*Let $\boldsymbol{R}$ be an o-minimal structure on $\mathbb{R}$, and let $\mathcal{C} = \{C_a\}_{a \in A}$ be a definable family of compact subsets of $\mathbb{R}^n$. If $\mathcal{C}$ is $(n+1)$-consistent, then $\mathcal{C}$ has a finite transversal.*

## Proof of Theorem B in the o-minimal case (Starchenko)

Suppose $\boldsymbol{R}$ is o-minimal, and write $\mathcal{C} = \{C_a\}_{a \in A}$.

By Helly's Theorem for finite families, the (definable) family whose members are the intersections of $n+1$ members of $\mathcal{C}$ is finitely consistent.

Apply Dolich's Theorem to this family to obtain a finite set $P \subseteq R^n$ with $P \cap C_{a_1} \cap \cdots \cap C_{a_{n+1}} \neq \emptyset$ for all $a_1, \ldots, a_{n+1} \in A$.

Thus

$$\mathcal{P} = \{\mathrm{conv}(C_a \cap P)\}_{a \in A}$$

is a family of convex subsets of $R^n$ with only finitely many distinct members, and $\mathcal{P}$ is $(n+1)$-consistent.

Hence $\emptyset \neq \bigcap \mathcal{P} \subseteq \bigcap \mathcal{C}$ by Helly's Theorem for finite families. $\square$