Distality in Combinatorics

Aaron Anderson

UCLA

January 25, 2024

Table of Contents

The Lazy Caterer's Problem

Distal Combinatorics at Purdue

Aaron Anderson

VC Theory

Incidences and Cuttings

Distality

Regularity

Continuous Logic

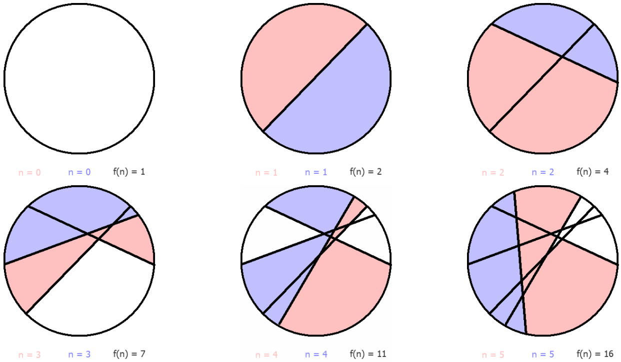


Figure: The maximum number of pieces that n slices cut a circle into

Vapnik-Chervonenkis Theory

Let X be an infinite set, and let $\mathcal{F}$ be a set of subsets of X.

Definition

- Let $\pi^*_{\mathcal{F}}(n)$ be the maximum number of subsets of $A \subset X$ that can be expressed as $S \cap A$ for some $S \in \mathcal{F}$, where $|A| = n$.
- The *dual vc-dimension* of $\mathcal{F}$, $\mathrm{VC}^*(\mathcal{F})$, is the largest d for which $\pi^*_{\mathcal{F}}(d) = 2^d$.
- The *dual vc-density* of $\mathcal{F}$, $\mathrm{vc}^*_{\mathcal{F}}(\mathcal{F})$, is the infimum of all d for which $\pi^*_{\mathcal{F}}(n) = \mathcal{O}(n^d)$.

Vapnik-Chervonenkis Theory: Catering Example

Let X be the circle, and $\mathcal{F}$ be the set of half-planes. Then $\pi^*_{\mathcal{F}}(n)$ is the number of pieces from the Lazy Caterer's Problem.

n	0	1	2	3	4	5
$\pi(n)$	1	2	4	7	11	16

This means $\mathrm{VC}(\mathcal{F}) = 2$.

Vapnik-Chervonenkis Theory: Sauer-Shelah

Theorem (Sauer-Shelah)

For any $\mathcal{F}$ with $\mathrm{VC}^*(\mathcal{F}) = d$,

$$\pi^*_{\mathcal{F}}(n) \leq \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} = \mathcal{O}(n^d)$$

for all $n \in \mathbb{N}$. Thus $\mathrm{vc}^*(\mathcal{F}) \leq \mathrm{VC}(\mathcal{F})$.

Example

In the Lazy Caterer's Problem, we get exactly

$$\pi^*_{\mathcal{F}}(n) = \binom{n}{0} + \binom{n}{1} + \binom{n}{2}.$$

Fix a language $\mathcal{L}$ and an $\mathcal{L}$-structure $\mathcal{M}$.

The dual VC-dimension of a formula $\phi(x; y)$ is the dual VC-dimension of the family

$$\mathcal{F}_\varphi = \{\varphi(M; \bar{b}) : \bar{b} \in M^{|y|}\}.$$

NIP

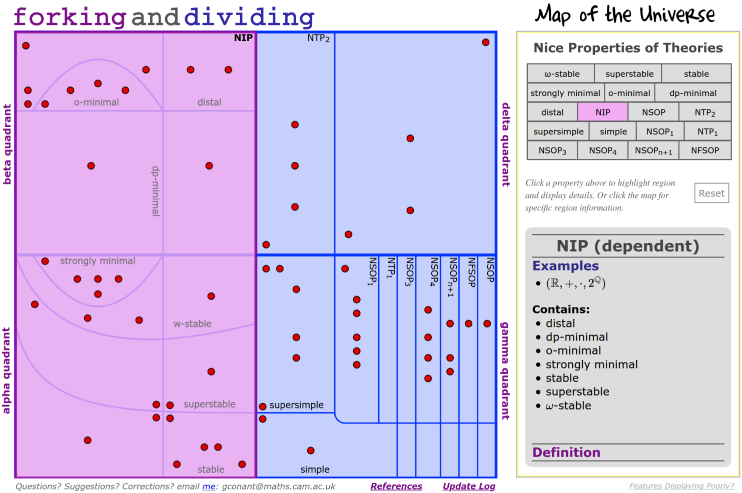
Fact

All formulas $\phi(x; y)$ *with* $|x| = d$ *in the ordered field* $\mathbb{R}$ *have vc-density* d*.*

Definition

If all definable families in a structure have finite vc-density/dimension, we call the structure NIP.

NIP Structures



Definability

We can define each* half-plane as

$$\{(x_1, x_2) \in \mathbb{R}^2 : b_1 x_1 + b_2 x_2 \leq 1\}$$

for some $(b_1, b_2) \in \mathbb{R}^2$, so the half-planes are parametrized by points in the plane.

Fact

The set $\mathcal{F}$ of all half-planes is a definable family in the structure $\langle \mathbb{R}; 0, 1, +, \cdot, < \rangle$ (the real ordered field). It is defined by

$$\varphi(x_1, x_2; y_1, y_2) \iff y_1 x_1 + y_2 x_2 \leq 1.$$

So is the set of circles, defined by

$$\varphi(x_1, x_2; y_1, y_2, y_3) \iff (x_1 - y_1)^2 + (x_2 - y_2)^2 = y_3^2.$$

Definability

Definition

- Let $\mathcal{M}$ be an $\mathcal{L}$-structure.
- Let $\varphi(x; y)$ be a formula made up of symbols from $\mathcal{L}$, where x and y could be multiple variables. Then for $\bar{b} \in M^{|y|}$, the set

$$\varphi(M; \bar{b}) := \{\bar{a} \in M^{|x|} : \varphi(\bar{a}, \bar{b}) \text{ is true}\}$$

  is called *definable*.
- The set

$$\mathcal{F}_\varphi = \{\varphi(M; \bar{b}) : \bar{b} \in M^{|y|}\}$$

  is called a *definable family*.

Table of Contents

Incidences

Many combinatorics problems boil down to something like this: Let P be a set of m points in $\mathbb{R}^d$, let Q be a set of n lines, circles, or curves from some other definable family.
What's the asymptotic growth in m and n of the size of the set of *incidences*:

$$I(P,Q) = \{(p,q) : p \in q\}?$$

$I(P,Q)$ can be thought of as the edges of the (bipartite) *incidence graph*.

Incidence Theorems

Theorem (Kövari-Sós-Túran)

If G is a bipartite graph with parts P and Q, containing no copy of $K_{s,t}$, then the number of edges satisfies

$$|E(P,Q)| = \mathcal{O}(mn^{1-1/t} + n)$$

Theorem (Szemerédi-Trotter)

If P is a set of m points and Q a set of n lines in $\mathbb{R}^2$, then

$$|I(P,Q)| = \mathcal{O}(m^{2/3}n^{2/3} + m + n)$$

The Cutting Lemma

Definition

Say that a set A *is crossed by* another set B if $A \cap B$ and $A \setminus B$ are both nonempty.

Theorem

For every set Q of n lines in $\mathbb{R}^2$, and for every $1 < r < n$, there exists a partition of the plane into $\mathcal{O}(r^2)$ "triangular" pieces with each piece only crossed by n/r of the half-planes.

Szemerédi-Trotter Proof Sketch

Proof.

- If $m^2 \leq n$, it works. Otherwise, let $r = \frac{n^{1/3}}{m^{2/3}}$.
- Apply the cutting lemma to get $\mathcal{O}(r^2)$ barely-intersecting pieces that are each crossed by only n/r lines.
- Use Kövari-Sós-Túran to count incidences on each piece, using the reduced number of points and reduced number of lines.
- Add up the number of incidences in each piece.

□

Table of Contents

Distal Cell Decompositions

- What if we want to count incidences on curves from some other definable family?
- One technique is a better cutting lemma. To get it, use a *distal cell decomposition*.

Distal Cell Decompositions

Definition

A distal cell decomposition $\mathcal{T}$ for a formula $\varphi(x; y)$ is a nice-and-definable method for taking a finite set $S \subset M^{|y|}$ of parameters, and outputting a set of cells, $\mathcal{T}(S)$.
They satisfy the following axioms:

- The union of the cells is $M^{|x|}$.
- Each cell is not crossed by $\varphi(M; \bar{b})$ for any $\bar{b} \in S$.
- The cells are uniformly definable. Roughly, there are some formulas that take in S and give you the cells, and each cell can be defined from only k many parameters from S for some k.

Distal Cell Decompositions in Continuous Logic

Theorem (A., generalizing Chernikov-Simon)

A metric structure M is distal iff every formula $\phi(x; y)$ has a strong honest definition:

A definable predicate $\theta(x; z)$ such that for any $a \in M^x$ and finite $B \subseteq M^y$, there is a tuple d in B with

- $\theta(a; d) = 0$
- *for all $a' \in M^x$, $b \in B$, $|\phi(a', b) - \phi(a, b)| \leq \theta(a; d)$.*

Intuitively, $\theta(x; d)$ bounds how much $\phi(x; b)$ can vary from $\phi(a; b)$, so $\{\theta(x; d) \leq \varepsilon\}$ is a good cell for a cell decomposition.

Distal Exponent/Density

Definition

- If a distal cell decomposition $\mathcal{T}$ satisfies $|\mathcal{T}(S)| = \mathcal{O}(|S|^d)$, then we say $\mathcal{T}$ *has exponent d*.
- The *distal density* of a formula $\varphi(x; y)$ is the infimum of all d such that φ has a distal cell decomposition of exponent d.

Fact

The exponent of $\mathcal{T}$ is at most the number of parameters needed to define its cells. As this has to be finite, so is the exponent.

Distal Structures

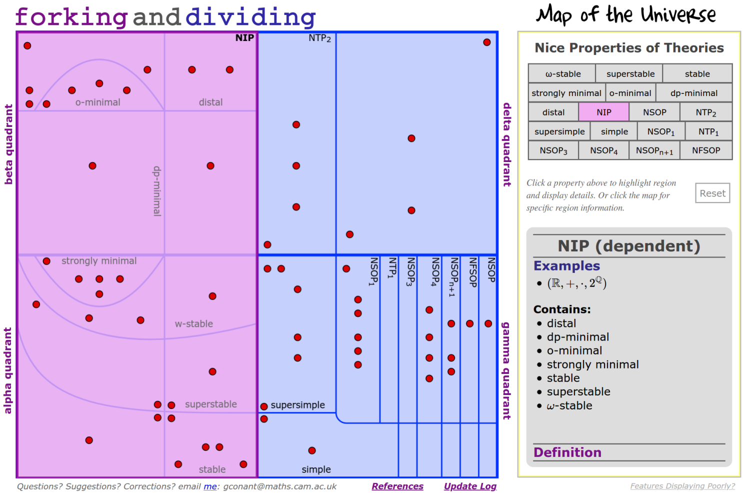
Definition

A *distal* structure is one where every definable family admits a distal cell decomposition.

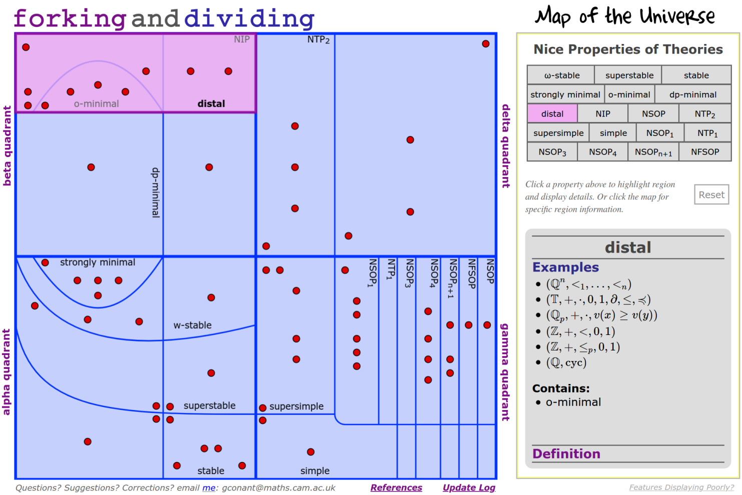
Fact

As the distal density of $\varphi(x; y)$ *is at least* $\mathrm{vc}^*(\mathcal{F}_\varphi)$, *and is always finite if* φ *has a distal cell decomposition, a distal structure is NIP.*

NIP Structures



Distal Structures



Distal Cutting Lemma

Theorem (Chernikov, Galvin, Starchenko)

If $\varphi(x; y)$ admits a distal cell decomposition with exponent d, then for any finite $S \subset M^{|y|}$ of size n and $1 < r < n$, there is a uniformly definable partition of $M^{|x|}$ into $\mathcal{O}(r^d)$ pieces, each of which is crossed by at most n/r of the formulas $\varphi(M; \bar{b})$ for $\bar{b} \in S$.

Distal Incidence Bound

Theorem (Chernikov, Galvin, Starchenko)

Let $\mathcal{M}$ be a structure and $t, d \in \mathbb{R}_{\geq 2}$. Assume that $E(x, y) \subseteq M^{|x|} \times M^{|y|}$ is a graph defined by a formula $\varphi(x; y)$ which has a distal cell decomposition with exponent t, and vc-density d.
Then for any finite $P \subseteq M^{|x|}, Q \subseteq M^{|y|}$, $|P| = m, |Q| = n$, such that the subgraph $E(P, Q)$ does not contain a complete bipartite subgraph $K_{s,s}$:

$$|E(P, Q)| = \mathcal{O}\left(m^{\frac{(t-1)d}{td-1}} n^{\frac{t(d-1)}{td-1}} + m + n\right).$$

Dimension Induction

It suffices to calculate distal density for one dimension.

Theorem (A.)

Let $\mathcal{M}$ be a structure in which all finite sets of formulas with $|x| = 1$ admit a distal cell decomposition $\mathcal{T}_1$ where

- *Every formula ψ of $\mathcal{T}_1$ refers to at most k parameters*
- *For some $d_0 \in \mathbb{N}$, all finite sets of formulas with $|x| = d_0$ have distal density $\leq r$*

All finite sets Φ of formulas with $|x| = d \geq d_0$ have distal density at most $k(d - d_0) + r$.

Dimension Induction Proof

The resulting cell decomposition generalizes the "vertical decomposition" of Chazelle, Edelsbrunner, Guibas, Sharir.

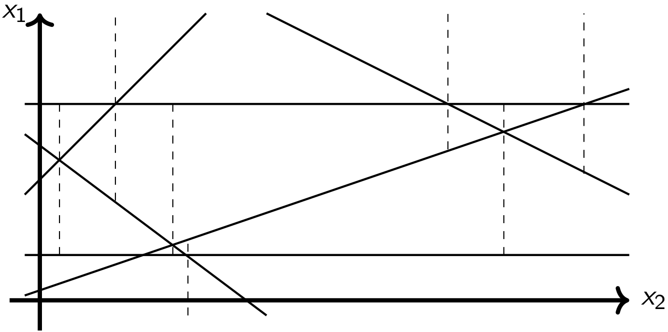


Figure: A "vertical decomposition" for lines in the plane

Bounds on Distal Density

$\mathcal{M}$	Dual VC-Density	Distal Density
$\mathbb{R}$ with field structure (and more)	$\lvert x\rvert$	$2\lvert x\rvert - 2$ (1 if $\lvert x\rvert = 1$)
weakly o-minimal structures	$\lvert x\rvert$	$2\lvert x\rvert - 1$
ordered vector spaces over ordered division rings	$\lvert x\rvert$	$\lvert x\rvert$
$\mathbb{Q}_p$ the valued field	$2\lvert x\rvert - 1$	$3\lvert x\rvert - 2$
$\mathbb{Q}_p$ in the linear reduct	$\lvert x\rvert$	$\lvert x\rvert$

All distal density results other than o-minimal $\lvert x\rvert = 2$ calculated using the dimension induction bound.

Table of Contents

Keisler Measures

A *Keisler measure* is a regular Borel measure on a type space $S_x(A)$.

A *generically stable* Keisler measure resembles a counting measure in an NIP structure, but these are more versatile for model-theoretic proofs.

Generically stable measures are particularly well-behaved (smooth) in distal structures - this characterizes distality.

Definable Strong Erdős-Hajnal

Definition

Let $\phi(x; y)$ be a formula, $A \subseteq M^{|x|}$, $B \subseteq M^{|y|}$. The pair (A, B) is ϕ-homogeneous when $\phi(A; B) = A \times B$ or $\phi(A; B) = \emptyset$.

Theorem

Let M be a distal structure, $\phi(x; y)$ a formula. There is a constant $\delta > 0$ and formulas $\psi^1(x; z_1), \psi^2(y; z_2)$ such that for any generically stable measures $\mu_1 \in \mathfrak{M}_x(M)$ and $\mu_2 \in \mathfrak{M}_y(M)$, there are $c_1 \in M^{z_1}$ and $c_2 \in M^{z_2}$ such that $\mu_1(\psi^1(M^{|x|}; c_1)) \geq \delta$ and $\mu_2(\psi^2(M^{|x|}; c_2)) \geq \delta$, and the pair $(\psi^1(M^{|x|}; c_1), \psi^2(M^{|x|}; c_2))$ is ϕ-homogeneous.

- $\mathcal{M} = \mathbb{R}$: Alon et al.
- *o*-minimal: Basu
- distal: Chernikov, Starchenko
- distal metric structure: A.

Regularity

Definable strong Erdős-Hajnal is just one form of the *distal regularity lemma*:

For any $\varepsilon > 0$, by iteratively applying SEH, we can decompose both $M^{|x|}$ and $M^{|y|}$ into a bounded number of pieces, such that the total measure of the non-homogeneous "rectangles" is at most ε.

Distal vs. SEH

Definable strong Erdős-Hajnal is equivalent to distality, but only a distal expansion is needed for strong Erdős-Hajnal on counting measures.

Does SEH for counting measures - or some other combinatorial property - imply a distal expansion?

Every known example of a non-distality comes from a failure of SEH, except. . .

Table of Contents

Continuous Logic

In continuous logic, we care about

- Metric Structures: bounded metric space, equipped with Lipschitz functions and $[0, 1]$-valued "relations"
- Formulas: Continuous $[0, 1]$-valued functions built out of function and relation symbols
- Definable predicates: Uniform limits of formulas

In this context, we can still talk about all of the definitions of distality.

Oscillation

Instead of asking for cells which are not crossed, look at oscillation:

Definition

Let X be a set, Y a metric space, $\Delta \subseteq X$, $f : X \to Y$. Define

$$\operatorname{osc}(f(x); \Delta) = \sup_{x,y \in \Delta} |f(x) - f(y)|.$$

We can also state a version of strong Erdős-Hajnal, with (ϕ, ε)-homogeneous pairs.

Examples of Distal Metric Structures

- Any discrete distal structure (o-minimal, $\mathbb{Q}_p$, etc.)
- Dual linear continua
- Metric (weakly) o-minimal structures, such as
- Real closed metric valued fields (RCMVF)

Non-Examples:

- The randomization of *any* structure
- Expansions of Infinite-Dimensional Hilbert Spaces, Atomless Probability Algebras

Thank you, Purdue!