Contents

Chapter 1

Inroduction

1.1 VC-dimension and vc-density

Throughout this section we work with a collection $\mathcal{F}$ of subsets of a set X. We call the pair $(X, \mathcal{F})$ a set system.

Definition 1.1.1.

- Given a subset A of X, we define the set system $(A, A \cap \mathcal{F})$ where $A \cap \mathcal{F} = \{A \cap F \mid F \in \mathcal{F}\}$.
- For $A \subset X$ we say that $\mathcal{F}$ shatters A if $A \cap \mathcal{F} = \mathcal{P}(A)$ (the power set of A).

Definition 1.1.2. We say $(X, \mathcal{F})$ has VC-dimension n if the largest subset of X shattered by $\mathcal{F}$ is of size n. If $\mathcal{F}$ shatters arbitrarily large subsets of X, we say that $(X, \mathcal{F})$ has infinite VC-dimension. We denote the VC-dimension of $(X, \mathcal{F})$ by $\mathrm{VC}(X, \mathcal{F})$.

Note 1.1.3. We may drop X from the notation $\mathrm{VC}(X, \mathcal{F})$, as the VC-dimension doesn't depend on the base set and is determined by $(\bigcup \mathcal{F}, \mathcal{F})$.

Set systems of finite VC-dimension tend to have good combinatorial properties, and we consider set systems with infinite VC-dimension to be poorly behaved.

Another natural combinatorial notion is that of a dual system:

Definition 1.1.4. For $a \in X$ define $X_a = \{F \in \mathcal{F} \mid a \in F\}$. Let $\mathcal{F}^* = \{X_a \mid a \in X\}$. We call $(\mathcal{F}, \mathcal{F}^*)$ the dual system of $(X, \mathcal{F})$. The VC-dimension of the dual system of $(X, \mathcal{F})$ is referred to as the dual VC-dimension of $(X, \mathcal{F})$ and denoted by $\mathrm{VC}^*(\mathcal{F})$. (As before, this notion doesn't depend on X.)

Lemma 1.1.5 (see 2.13b in [4]). *A set system $(X, \mathcal{F})$ has finite VC-dimension if and only if its dual system has finite VC-dimension. More precisely*

$$\mathrm{VC}^*(\mathcal{F}) \leq 2^{1+\mathrm{VC}(\mathcal{F})}.$$

For a more refined notion of complexity of $(X, \mathcal{F})$ we look at the traces of our family on finite sets:

Definition 1.1.6. Define the shatter function $\pi_{\mathcal{F}} \colon \mathbb{N} \longrightarrow \mathbb{N}$ of $\mathcal{F}$ and the dual shatter function $\pi^*_{\mathcal{F}} \colon \mathbb{N} \longrightarrow \mathbb{N}$ of $\mathcal{F}$ by

$$\pi_{\mathcal{F}}(n) = \max\{|A \cap \mathcal{F}| \mid A \subset X \text{ and } |A| = n\}$$
$$\pi^*_{\mathcal{F}}(n) = \max\{\mathrm{atoms}(B) \mid B \subset \mathcal{F}, |B| = n\}$$

where $\mathrm{atoms}(B)$ = number of atoms in the boolean algebra of sets generated by B. Note that the dual shatter function is precisely the shatter function of the dual system: $\pi^*_{\mathcal{F}} = \pi_{\mathcal{F}^*}$.

A simple upper bound is $\pi_{\mathcal{F}}(n) \leq 2^n$ (same for the dual). If the VC-dimension of $\mathcal{F}$ is infinite then clearly $\pi_{\mathcal{F}}(n) = 2^n$ for all n. Conversely we have the following remarkable fact:

Theorem 1.1.7 (Sauer-Shelah '72, see [6], [7]). *If the set system $(X, \mathcal{F})$ has finite VC-dimension d then $\pi_{\mathcal{F}}(n) \leq \binom{n}{\leq d}$ for all n, where $\binom{n}{\leq d} = \binom{n}{d} + \binom{n}{d-1} + \ldots + \binom{n}{1}$.*

Thus the systems with a finite VC-dimension are precisely the systems where the shatter function grows polynomially. Define the vc-density of $\mathcal{F}$ to quantify the growth of the shatter function of $\mathcal{F}$:

Definition 1.1.8. Define the vc-density and dual vc-density of $\mathcal{F}$ as

$$\begin{aligned} \mathrm{vc}(\mathcal{F}) &= \limsup_{n\to\infty} \frac{\log \pi_{\mathcal{F}}(n)}{\log n} \in \mathbb{R}^{\geq 0} \cup \{+\infty\}, \\ \mathrm{vc}^*(\mathcal{F}) &= \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}}(n)}{\log n} \in \mathbb{R}^{\geq 0} \cup \{+\infty\}. \end{aligned}$$

Generally speaking a shatter function that is bounded by a polynomial doesn't itself have to be a polynomial. Proposition 4.12 in [2] gives an example of a shatter function that grows like $n \log n$ (so it has vc-density 1).

So far the notions that we have defined are purely combinatorial. We now adapt VC-dimension and vc-density to the model theoretic context.

Definition 1.1.9. Work in a first-order structure M. Fix a finite collection of formulas $\Phi(x, y)$.

- For $\phi(x, y) \in \mathcal{L}(M)$ and $b \in M^{|y|}$ let

$$\phi(M^{|x|}, b) = \{a \in M^{|x|} \mid \phi(a, b)\} \subseteq M^{|x|}.$$

- Let $\Phi(M^{|x|}, M^{|y|}) = \{\phi(M^{|x|}, b) \mid \phi_i \in \Phi, b \in M^{|y|}\} \subseteq \mathcal{P}(M^{|x|})$.
- Let $\mathcal{F}_\Phi = \Phi(M^{|x|}, M^{|y|})$, giving rise to a set system $(M^{|x|}, \mathcal{F}_\Phi)$.
- Define the VC-dimension $\mathrm{VC}(\Phi)$ of Φ, to be the VC-dimension of $(M^{|x|}, \mathcal{F}_\Phi)$, similarly for the dual.

- Define the vc-density $\mathrm{vc}(\Phi)$ of Φ, to be the vc-density of $(M^{|x|}, \mathcal{F}_\Phi)$, similarly for the dual.

We will also refer to the vc-density and VC-dimension of a single formula ϕ viewing it as a one element collection $\Phi = \{\phi\}$.

Counting atoms of a boolean algebra in a model theoretic setting corresponds to counting types, so it is instructive to rewrite the shatter function in terms of types.

Definition 1.1.10.

$$\pi^*_\Phi(n) = \max\{\text{number of } \Phi\text{-types over } B \mid B \subset M, |B| = n\}$$

Here a Φ-type over B is a maximal consistent collection of formulas of the form $\phi(x, b)$ or $\neg\phi(x, b)$ where $\phi \in \Phi$ and $b \in B$.

Functions π^*_Φ and $\pi^*_{\mathcal{F}_\Phi}$ are not equal, as one fixes the size of boolean algebra and another fixes the size of the parameter set. However, as the following lemma demonstrates, they both give the same asymptotic definition of dual vc-density.

Lemma 1.1.11.

$$\mathrm{vc}^*(\Phi) = \textit{degree of polynomial growth of } \pi^*_\Phi(n) = \limsup_{n\to\infty} \frac{\log \pi^*_\Phi(n)}{\log n}$$

Proof. With parameter set of size n, we get $|\Phi|n$ elements in the boolean algebra. We check that asymptotically it doesn't matter whether we look at growth of boolean algebra of size n or size $|\Phi|n$.

$$\begin{aligned}
&\pi^*_{\mathcal{F}_\Phi}(n) \le \pi^*_\Phi(n) \le \pi^*_{\mathcal{F}_\Phi}(|\Phi|n) \\
&\mathrm{vc}^*(\Phi) \le \limsup_{n\to\infty} \frac{\log \pi^*_\Phi(n)}{\log n} \le \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}_\Phi}(|\Phi|n)}{\log n} = \\
&= \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}_\Phi}(|\Phi|n)}{\log |\Phi|n} \frac{\log |\Phi|n}{\log n} = \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}_\Phi}(|\Phi|n)}{\log |\Phi|n} \le \\
&\le \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}_\Phi}(n)}{\log n} = \mathrm{vc}^*(\Phi)
\end{aligned}$$

□

One can check that the shatter function and hence VC-dimension and vc-density of a formula are elementary notions, so they only depend on the first-order theory of the structure M.

NIP theories are a natural context for studying vc-density. In fact we can take the following as the definition of NIP:

Definition 1.1.12. Define ϕ to be NIP if it has finite VC-dimension in a theory T. A theory T is NIP if all the formulas in T are NIP.

In a general combinatorial context for arbitrary set systems, vc-density can be any real number in $0 \cup [1, \infty)$ (see [5]). Less is known if we restrict our attention to NIP theories. Proposition 4.6 in [2] gives examples of formulas that have non-integer rational vc-density in an NIP theory, however it is open whether one can get an irrational vc-density in this model-theoretic setting.

Instead of working with a theory formula by formula, we can look for a uniform bound for all formulas:

Definition 1.1.13. For a given NIP structure M, define the <u>vc-function</u>

$$\begin{aligned}
\mathrm{vc}^M(n) &= \sup\{\mathrm{vc}^*(\phi(x,y)) \mid \phi \in \mathcal{L}(M), |x| = n\} \\
&= \sup\{\mathrm{vc}(\phi(x,y)) \mid \phi \in \mathcal{L}(M), |y| = n\} \in \mathbb{R}^{\ge 0} \cup \{+\infty\}
\end{aligned}$$

As before this definition is elementary, so it only depends on the theory of M. We omit the superscript M if it is understood from the context. One can easily check the following bounds:

Lemma 1.1.14 (Lemma 3.22 in [2])**.** *We have* $\mathrm{vc}(1) \geq 1$ *and* $\mathrm{vc}(n) \geq n\,\mathrm{vc}(1)$*.*

However, it is not known whether the second inequality can be strict or even whether $\mathrm{vc}(1) < \infty$ implies $\mathrm{vc}(n) < \infty$.

Chapter 2

vc-density for trees

We show that for the theory of infinite trees we have $\mathrm{vc}(n) = n$ for all n. This generalizes a result of Simon in [8] showing that the trees are dp-minimal.

VC-density was studied in [2] by Aschenbrenner, Dolich, Haskell, MacPherson, and Starchenko as a natural notion of dimension for NIP theories. In an NIP theory we can define a vc-function

$$\mathrm{vc} : \mathbb{N} \longrightarrow \mathbb{R} \cup \{\infty\},$$

where $\mathrm{vc}(n)$ measures the worst-case complexity of families of definable sets in an n-dimensional space. Simplest possible behavior is $\mathrm{vc}(n) = n$ for all n. Theories with the property that $\mathrm{vc}(1) = 1$ are known to be dp-minimal, i.e., having the smallest possible dp-rank. In general, it is not known whether there can be a dp-minimal theory which doesn't satisfy $\mathrm{vc}(n) = n$.

In this paper we work with trees viewed as posets. Parigot in [9] showed that such structures have NIP. This result was strengthened by Simon in [8] showing that trees are dp-minimal. The paper [2] poses the following problem:

Problem 2.0.15. ([2] p. 47) Determine the VC density function of each (infinite) tree.

Here we settle this question by showing that any model of the theory of trees has $\mathrm{vc}(n) = n$.

Section 1 of the paper consists of a basic introduction to the concepts of VC-dimension and vc-density. In Section 2 we introduce proper subdivisions – the main tool that we use to analyze trees. We also prove some basic properties of proper subdivisions. Section 3 introduces the key constructions of proper subdivisions in tree which will be used in the proof. Section 4 presents the proof of $\mathrm{vc}(n) = n$ via the subdivisions.

We use notation $a \in T^n$ for the tuples of size n. For a variable x or a tuple a we denote their arity by $|x|$ and $|a|$ respectively.

The language for the trees consists of a single binary predicate $\{\leq\}$. The theory of trees states that $\leq$ defines a partial order and for every element a the set $\{x \mid x < a\}$ is linearly ordered by $<$. For visualization purposes we assume that trees grow upwards, with the smaller elements on the bottom and the larger elements on the top. If $a \leq b$ we will say that a is below b and b is above a.

Definition 2.0.16. Work in a tree $\mathcal{T} = (T, \leq)$. For $x \in T$ let $I(x) = \{t \in T \mid t \leq x\}$ denote all the elements below x. The *meet* of two tree elements a, b is the greatest element of $I(a) \cap I(b)$ (if one exists) and is denoted by $a \wedge b$.

The theory of <u>meet trees</u> requires that any two elements in the same connected component have a meet. <u>Colored trees</u> are trees with a finite number of colors added via unary predicates.

From now on assume that all trees are colored. We allow our trees to be disconnected (so really, we work with forests) or finite unless otherwise stated.

2.1 Proper Subdivisions: Definition and Properties

We work with finite relational languages. Given a formula we define its complexity as the depth of quantifiers used to build up the formula. More precisely:

Definition 2.1.1. Define *complexity* of a formula by induction:

$$\text{Complexity}(\text{q.f. formula}) = 0$$

$$\text{Complexity}(\exists x \phi(x)) = \text{Complexity}(\phi(x)) + 1$$

$$\text{Complexity}(\phi \wedge \psi) = \max(\text{Complexity}(\phi), \text{Complexity}(\psi))$$

$$\text{Complexity}(\neg \phi) = \text{Complexity}(\phi)$$

A simple inductive argument verifies that there are (up to equivalence) only finitely many formulas when the complexity and the number of free variables are fixed. We will use the following notation for types:

Definition 2.1.2. Let $\mathcal{B}$ be a structure, $A \subset B$ be a finite parameter set, and a, b be tuples in $\mathcal{B}$, and m, n be natural numbers.

- $\text{tp}^n_{\mathcal{B}}(a/A)$ will stand for the set of all A-formulas of complexity $\leq n$ that are true of a in $\mathcal{B}$. If $A = \emptyset$ we may also write this as $\text{tp}^n_{\mathcal{B}}(a)$. The subscript $\mathcal{B}$ will be omitted as well if it is clear from context. Note that if A is finite, there are finitely many such formulas (up to equivalence). The conjunction of those formulas still has complexity $\leq n$ and so we can just associate a single formula to every type $\text{tp}^n_{\mathcal{B}}(a/A)$.
- $\mathcal{B} \models a \equiv^n_A b$ means that a, b have the same type with complexity $\leq n$ over A in $\mathcal{B}$, i.e., $\text{tp}^n_{\mathcal{B}}(a/A) = \text{tp}^n_{\mathcal{B}}(b/A)$.

- $S^n_{\mathcal{B},m}(A)$ will stand for the set of all m-types of complexity $\leq n$ over A:

$$S^n_{\mathcal{B},m}(A) = \{\mathrm{tp}^n_{\mathcal{B}}(a/A) \mid a \in B^m\}.$$

Definition 2.1.3.

- Let $\mathcal{A}$, $\mathcal{B}$, $\mathcal{T}$ be structures in some (possibly different) finite relational languages. If the underlying sets A, B of $\mathcal{A}, \mathcal{B}$ partition the underlying set T of $\mathcal{T}$ (i.e. $T = A \sqcup B$), then we say that $(\mathcal{A}, \mathcal{B})$ is a *subdivision* of $\mathcal{T}$.

- A subdivision $(\mathcal{A}, \mathcal{B})$ of $\mathcal{T}$ is called *n-proper* if given $p, q \in \mathbb{N}$, $a_1, a_2 \in A^p$ and $b_1, b_2 \in B^q$ with

$$\mathcal{A} \models a_1 \equiv_n a_2$$

$$\mathcal{B} \models b_1 \equiv_n b_2$$

  we have

$$\mathcal{T} \models a_1 b_1 \equiv_n a_2 b_2.$$

- A subdivision $(\mathcal{A}, \mathcal{B})$ of $\mathcal{T}$ is called *proper* if it is n-proper for all $n \in \mathbb{N}$.

Lemma 2.1.4. *Consider a subdivision $(\mathcal{A}, \mathcal{B})$ of $\mathcal{T}$. If $(\mathcal{A}, \mathcal{B})$ is 0-proper then it is proper.*

Proof. We prove that the subdivision is n-proper for all k by induction. The case $n = 0$ is given by the assumption. Suppose we have $\mathcal{T} \models \exists x\, \phi^n(x, a_1, b_1)$ where ϕ^n is some formula of complexity n. Let $a \in T$ witness the existential claim, i.e., $\mathcal{T} \models \phi^n(a, a_1, b_1)$. We can have $a \in A$ or $a \in B$. Without loss of generality assume $a \in A$. Let $\boldsymbol{p} = \mathrm{tp}^n_{\mathcal{A}}(a, a_1)$. Then we have

$$\mathcal{A} \models \exists x \ \mathrm{tp}^n_{\mathcal{A}}(x, a_1) = \boldsymbol{p}$$

(with $\mathrm{tp}^n_{\mathcal{A}}(x, a_1) = \boldsymbol{p}$ a shorthand for $\phi_{\boldsymbol{p}}(x)$ where $\phi_{\boldsymbol{p}}$ is a formula that determines the type $\boldsymbol{p}$). The formula $\mathrm{tp}^n_{\mathcal{A}}(x, a_1) = \boldsymbol{p}$ is of complexity $\leq k$ so $\exists x \ \mathrm{tp}^n_{\mathcal{A}}(x, a_1) = \boldsymbol{p}$ is of complexity $\leq k+1$. By the inductive hypothesis we have

$$\mathcal{A} \models \exists x \ \mathrm{tp}^n_{\mathcal{A}}(x, a_2) = \boldsymbol{p}\,.$$

Let a' witness this existential claim, so that $\mathrm{tp}^n_{\mathcal{A}}(a', a_2) = \boldsymbol{p}$, hence $\mathrm{tp}^n_{\mathcal{A}}(a', a_2) = \mathrm{tp}^n_{\mathcal{A}}(a, a_1)$, that is, $\mathcal{A} \models a'a_2 \equiv_n aa_1$. By the inductive hypothesis we therefore have $\mathcal{T} \models aa_1b_1 \equiv_n a'a_2b_2$; in particular $\mathcal{T} \models \phi^n(a', a_2, b_2)$as $\mathcal{T} \models \phi^n(a, a_1, b_1)$, and $\mathcal{T} \models \exists x \phi^n(x, a_2, b_2)$. $\square$

This lemma is general, but we will use it specifically applied to (colored) trees. Suppose $\mathcal{T}$ is a (colored) tree in some language $\mathcal{L} = \{\leq, \dots\}$. Suppose $\mathcal{A}, \mathcal{B}$ are some structures in languages $\mathcal{L}_A, \mathcal{L}_B$ which expand $\mathcal{L}$, with the $\mathcal{L}$-reducts of $\mathcal{A}, \mathcal{B}$ substructures of $\mathcal{T}$. Furthermore suppose that $(\mathcal{A}, \mathcal{B})$ is 0-proper. Then by the previous lemma $(\mathcal{A}, \mathcal{B})$ is a proper subdivision of $\mathcal{T}$. From now on all the subdivisions we work with will be of this form.

Example 2.1.5. Suppose a tree consists of two connected components C_1, C_2. Then those components $(C_1, \leq)$, $(C_2, \leq)$ viewed as substructures form a proper subdivision. To see that we only need to show that this subdivision is 0-proper. But that is immediate as any $c_1 \in C_1$ and $c_2 \in C_2$ are incomparable.

Example 2.1.6. Fix a tree $\mathcal{T}$ in the language $\{\leq\}$ and $a \in T$. Let $B = \{t \in T \mid a < t\}$, $S = \{t \in T \mid t \leq a\}$, $A = T - B$. Then $(A, \leq, S)$ and $(B, \leq)$ form a proper subdivision, where $\mathcal{L}_A$ has a unary predicate interpreted by S. To see this, again, we show that the subdivision is 0-proper. The only time $a \in A$ and $b \in B$ are comparable is when $a \in S$, and this is captured by the language. (See proof of Lemma 2.2.7 for more details.)

Definition 2.1.7. For $\phi(x, y)$, $A \subseteq T^{|x|}$ and $B \subseteq T^{|y|}$

- let $\phi(A, b) = \{a \in A \mid \phi(a, b)\} \subseteq A$, and
- let $\phi(A, B) = \{\phi(A, b) \mid b \in B\} \subseteq \mathcal{P}(A)$.

Thus $\phi(A, B)$ is a collection of subsets of A definable by ϕ with parameters from B. We notice the following bound when A, B are parts of a proper subdivision.

Corollary 2.1.8. *Let $\mathcal{A}, \mathcal{B}$ be a proper subdivision of $\mathcal{T}$ and $\phi(x, y)$ be a formula of complexity n. Then $|\phi(A^{|x|}, B^{|y|})|$ is bounded by $|S^n_{\mathcal{B},|y|}|$.*

Proof. Take some $a \in A^{|x|}$ and $b_1, b_2 \in B^{|y|}$ with $\mathrm{tp}^n_{\mathcal{B}}(b_1) = \mathrm{tp}^n_{\mathcal{B}}(b_2)$. We have $\mathcal{B} \models b_1 \equiv_n b_2$ and (trivially) $\mathcal{A} \models a \equiv_n a$. Thus we have $\mathcal{T} \models ab_1 \equiv_n ab_2$, so $T \models \phi(a, b_1) \leftrightarrow \phi(a, b_2)$. Since a was arbitrary we have $\phi(A^{|x|}, b_1) = \phi(A^{|x|}, b_2)$ as different traces can only come from parameters of different types. Thus $|\phi(A^{|x|}, B^{|y|}) \leq |S^n_{\mathcal{B},|y|}|$. ∎

We note that the size of the type space $|S^n_{\mathcal{B},|y|}|$ can be bounded uniformly:

Definition 2.1.9. Fix a (finite relational) language $\mathcal{L}_B$. Let $N = N(n, m, \mathcal{L}_B)$ be smallest integer such that for any structure $\mathcal{B}$ in $\mathcal{L}_B$ we have $|S^n_{\mathcal{B},m}| \leq N$. This integer exists as there is a finite number (up to equivalence) of possible formulas of complexity $\leq n$ with m free variables. Note that $N(n, m, \mathcal{L}_B)$ is increasing in all variables:

$$n \leq n', m \leq m', \mathcal{L}_B \subseteq \mathcal{L}'_B \Rightarrow N(n, m, \mathcal{L}_B) \leq N(n', m', \mathcal{L}'_B)$$

2.2 Proper Subdivisions: Constructions

Throughout this section, $\mathcal{T}$ denotes a colored meet tree. First, we describe several constructions of proper subdivisions that are needed for the proof.

Definition 2.2.1. We use $E(b, c)$ to express that b and c are in the same connected component:

$$E(b, c) \Leftrightarrow \exists x \, (b \geq x) \wedge (c \geq x).$$

Definition 2.2.2. Given an element a of the tree we call the sets fo all the elements above a, i.e. the set $\{x \mid x \geq a\}$, the *closed cone* above a. Connected components of that cone can be thought of as *open cones* above a. With that interpretation in mind, the notation $E_a(b, c)$ means that b and c are in the same open cone above a. More formally:

$$E_a(b, c) \Leftrightarrow E(b, c) \text{ and } (b \wedge c) > a.$$

Fix a language $\mathcal{L}$ for a colored tree $\mathcal{L} = \{\leq, C_1, \ldots C_n\} = \{\leq, \vec{C}\}$. In the following four definitions structures denoted by $\mathcal{B}$ are going to be in the same language $\mathcal{L}_B = \mathcal{L} \cup \{U\}$ with U a unary predicate. It is not always necessary to have this predicate but we keep it for the sake of uniformity. Structures denoted by $\mathcal{A}$ will have different languages $\mathcal{L}_A$ (those are not as important in later applications).

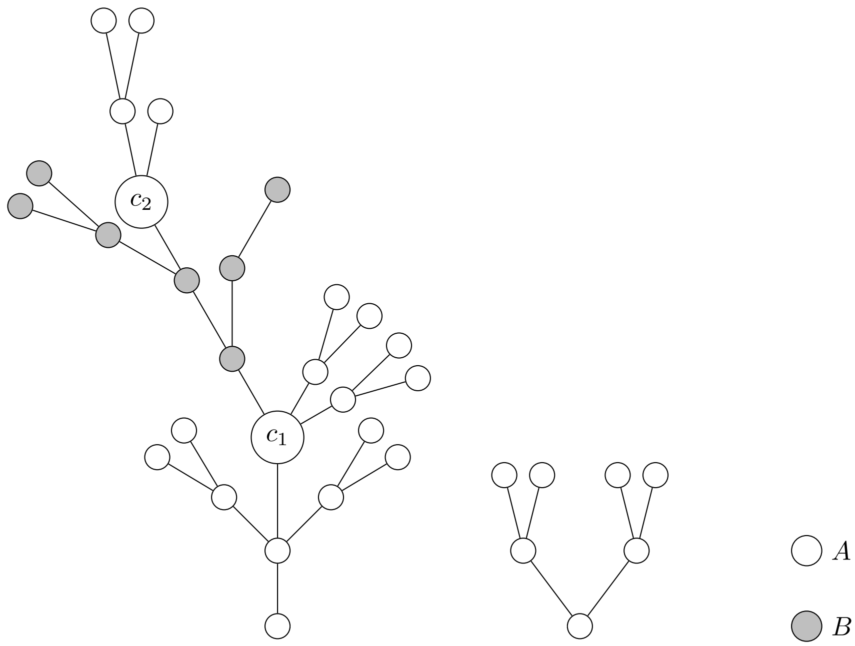

Figure 2.1: Proper subdivision for $(A, B) = (A_{c_2}^{c_1}, B_{c_2}^{c_1})$

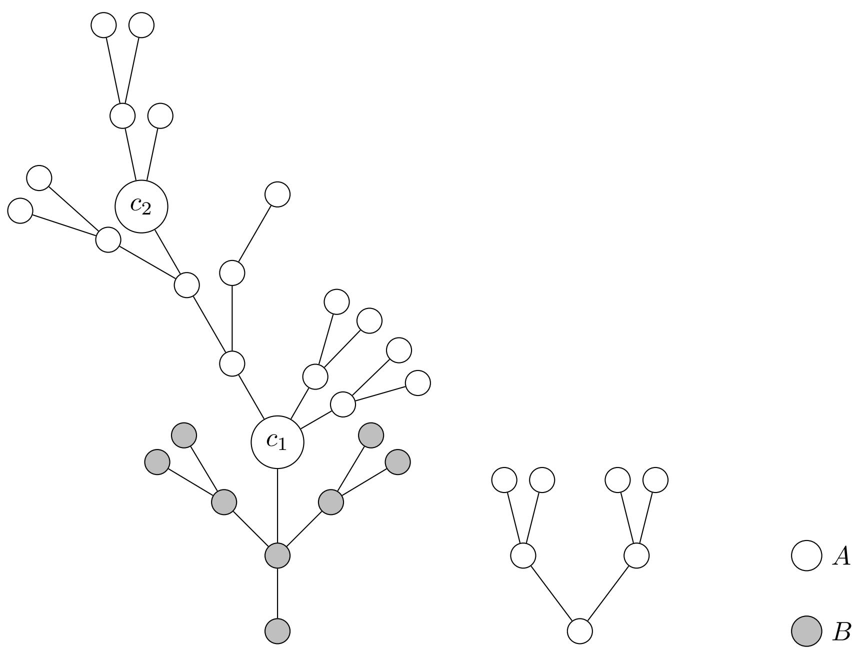

Figure 2.2: Proper subdivision for $(A, B) = (A_{c_1}, B_{c_1})$

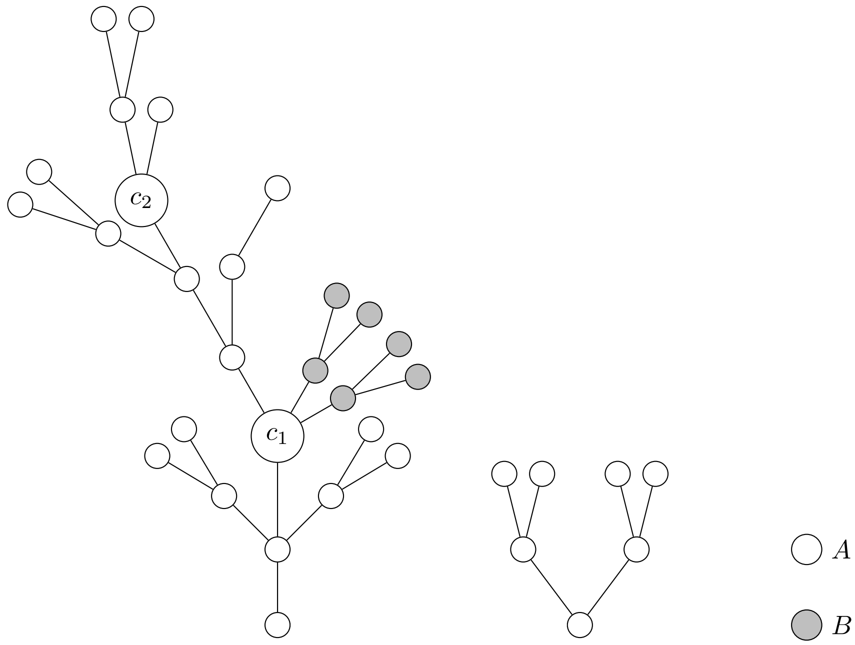

Figure 2.3: Proper subdivision for $(A, B) = (A_G^{c_1}, B_G^{c_1})$ for $S = \{c_2\}$

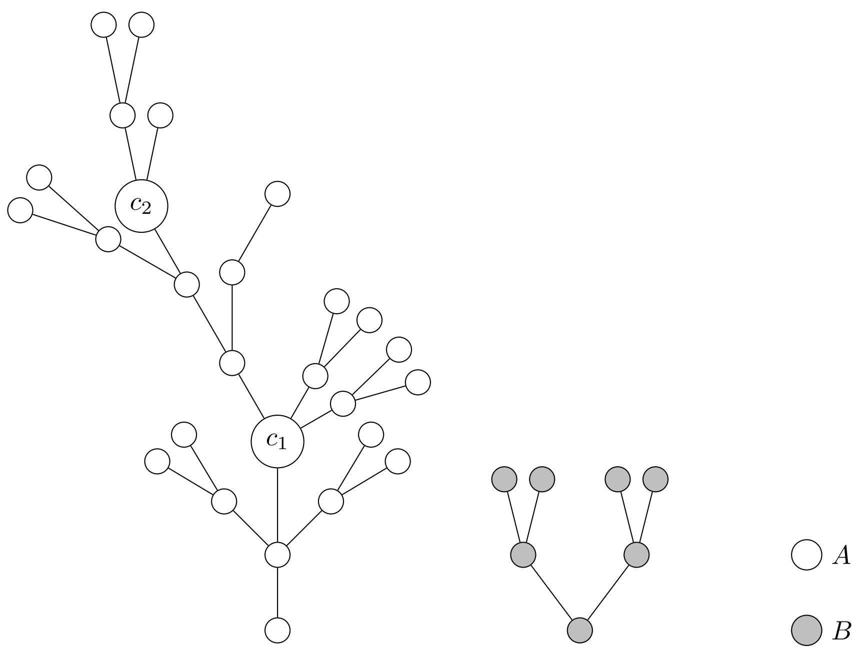

Figure 2.4: Proper subdivision for $(A, B) = (A_G, B_G)$ for $S = \{c_1, c_2\}$

Definition 2.2.3. Fix $c_1 < c_2$ in T. Let

$$B = \{b \in T \mid E_{c_1}(c_2, b) \wedge \neg(b \geq c_2)\},$$

$$A = T - B,$$

$$S_1 = \{t \in T \mid t < c_1\},$$

$$S_2 = \{t \in T \mid t < c_2\},$$

$$S_B = S_2 - S_1,$$

$$T_A = \{t \in T \mid c_2 \leq t\}.$$

Define structures $\mathcal{A}^{c_1}_{c_2} = (A, \leq, \vec{C} \cap A, S_1, T_A)$ where $\vec{C} \cap A = \{C_1 \cap A, \ldots, C_n \cap A\}$ and $\mathcal{B}^{c_1}_{c_2} = (B, \leq, \vec{C} \cap B, S_B)$ where $\mathcal{L}_A$ is an expansion of $\mathcal{L}$ by two unary predicate symbols (and $\mathcal{L}_B$ as defined above). Note that $c_1, c_2 \notin B$.

Definition 2.2.4. Fix c in T. Let

$$B = \{b \in T \mid \neg(b \geq c) \wedge E(b, c)\},$$

$$A = T - B,$$

$$S_1 = \{t \in T \mid t < c\}.$$

Define structures $\mathcal{A}_c = (A, \leq, \vec{C} \cap A)$ and $\mathcal{B}_c = (B, \leq, \vec{C} \cap B, S_1)$ where $\mathcal{L}_A = \mathcal{L}$ (and $\mathcal{L}_B$ as defined above). Note that $c \notin B$. (cf example 2.1.6).

Definition 2.2.5. Fix c in T and a finite subset $S \subseteq T$. Let

$$B = \{b \in T \mid (b > c) \text{ and for all } s \in S \text{ we have } \neg E_c(s, b)\},$$

$$A = T - B,$$

$$S_1 = \{t \in T \mid t \leq c\}.$$

Define structures $\mathcal{A}^c_S = (A, \leq, \vec{C} \cap A, S_1)$ and $\mathcal{B}^c_S = (B, \leq, \vec{C} \cap B, B)$ where

$\mathcal{L}_A$ is an expansion of $\mathcal{L}$ by a single unary predicate (and $U \in \mathcal{L}_B$ vacuously interpreted by B). Note that $c \notin B$ and $S \cap B = \emptyset$.

Definition 2.2.6. Fix a finite subset $S \subseteq T$. Let

$$B = \{b \in T \mid \text{ for all } s \in S \text{ we have } \neg E(s, b)\},$$
$$A = T - B.$$

Define structures $\mathcal{A}_S = (A, \leq)$ and $\mathcal{B}_S = (B, \leq, \vec{C} \cap B, B)$ where $\mathcal{L}_A = \mathcal{L}$ (and $U \in \mathcal{L}_B$ vacuously interpreted by B). Note that $S \cap B = \emptyset$. (cf. example 2.1.5)

Lemma 2.2.7. *The pairs of structures defined above are all proper subdivisions of $\mathcal{T}$.*

Proof. We only show this holds for the pair $(\mathcal{A}, \mathcal{B}) = (\mathcal{A}^{c_1}_{c_2}, \mathcal{B}^{c_1}_{c_2})$. The other cases follow by a similar argument. The sets A, B partition T by definition, so (A, B) is a subdivision of $\mathcal{T}$. To show that it is proper, by Lemma 2.1.4 we only need to check that it is 0-proper. Suppose we have

$$a = (a_1, a_2, \ldots, a_p) \in A^p,$$
$$a' = (a'_1, a'_2, \ldots, a'_p) \in A^p,$$
$$b = (b_1, b_2, \ldots, b_q) \in B^q,$$
$$b' = (b'_1, b'_2, \ldots, b'_q) \in B^q.$$

with $\mathcal{A} \models a \equiv_0 a'$ and $\mathcal{B} \models b \equiv_0 b'$. We need to show that ab has the same quantifier-free type in $\mathcal{T}$ as $a'b'$. Any two elements in T can be related in the four following ways:

$$
\begin{aligned}
&x = y, \\
&x < y, \\
&x > y, \text{ or} \\
&x, y \text{ are incomparable.}
\end{aligned}
$$

We need to check that for all i, j the same relations hold for (a_i, b_j) as do for (a'_i, b'_j).

- It is impossible that $a_i = b_j$ as they come from disjoint sets.
- Suppose $a_i < b_j$. This forces $a_i \in S_1$ thus $a'_i \in S_1$ and $a'_i < b'_j$.
- Suppose $a_i > b_j$. This forces $b_j \in S_B$ and $a \in T_A$, thus $b'_j \in S_B$ and $a'_i \in T_A$, so $a'_i > b'_j$.
- Suppose a_i and b_j are incomparable. Two cases are possible:
  - $b_j \notin S_B$ and $a_i \in T_A$. Then $b'_j \notin S_B$ and $a'_i \in T_A$ making a'_i, b'_j incomparable.
  - $b_j \in S_B$, $a_i \notin T_A$, $a_i \notin S_1$. Similarly this forces a'_i, b'_j to be incomparable.

Also we need to check that ab has the same colors as $a'b'$. But that is immediate as having the same color in a substructure means having the same color in the tree. $\square$

2.3 Main proof

The basic idea for the proof is as follows. Suppose we have a formula with q parameters over a parameter set of size n. We are able to split our parameter space into $O(n)$ many partitions. Each of q parameters can come from any of those $O(n)$ partitions giving us $O(n)^q$ many choices for parameter configuration. When every parameter is coming from a fixed partition the number of definable sets is constant and in fact is uniformly bounded above by some N. This gives us at most $N \cdot O(n)^q$ possibilities for different definable sets.

First, we generalize Corollary 2.1.8. (This is required for computing vc-density for formulas $\phi(x, y)$ with $|y| > 1$).

Lemma 2.3.1. *Consider a finite collection $(\mathcal{A}_i, \mathcal{B}_i)_{i \le n}$ satisfying the following properties:*

- *$(\mathcal{A}_i, \mathcal{B}_i)$ is either a proper subdivision of $\mathcal{T}$ or $A_i = T$ and $B_i = \{b_i\}$,*
- *all $\mathcal{B}_i$ have the same language $\mathcal{L}_B$,*
- *sets $\{B_i\}_{i \le n}$ are pairwise disjoint.*

Let $A = \bigcap_{i \in I} A_i$. Fix a formula $\phi(x, y)$ of complexity m . Let $N = N(m, |y|, \mathcal{L}_B)$ be as in Definition 2.1.9. Consider any $B \subseteq T^{|y|}$ of the form

$$B = B_1^{i_1} \times B_2^{i_2} \times \ldots \times B_n^{i_n} \text{ with } i_1 + i_2 + \ldots + i_n = |y|.$$

(some of the indeces can be zero). Then we have the following bound:

$$\phi(A^{|x|}, B) \le N^{|y|}.$$

Proof. We show this result by counting types.

Claim 2.3.2. *Suppose we have*

$$b_1, b_1' \in B_1^{i_1} \text{ with } b_1 \equiv_m b_1' \text{ in } \mathcal{B}_1,$$
$$b_2, b_2' \in B_2^{i_2} \text{ with } b_2 \equiv_m b_2' \text{ in } \mathcal{B}_2,$$
$$\dots$$
$$b_n, b_n' \in B_n^{i_n} \text{ with } b_n \equiv_m b_n' \text{ in } \mathcal{B}_n.$$

Then

$$\phi(A^{|x|}, b_1, b_2, \dots b_n) \iff \phi(A^{|x|}, b_1', b_2', \dots b_n').$$

Proof. Define $\bar{b}_i = (b_1, \dots, b_i, b_{i+1}', \dots, b_n') \in B$ for $i \in [0..n]$. (That is, a tuple where first i elements are without prime, and elements after that are with a prime.) We have $\phi(A^{|x|}, \bar{b}_i) \iff \phi(A^{|x|}, \bar{b}_{i+1})$ as either $(\mathcal{A}_{i+1}, \mathcal{B}_{i+1})$ is m-proper or $\mathcal{B}_{i+1}$ is a singleton, and the implication is trivial. (Notice that $b_i \in \mathcal{A}_j$ for $j \neq i$ by disjointness assumption.) Thus, by induction we get $\phi(A^{|x|}, \bar{b}_0) \iff \phi(A^{|x|}, \bar{b}_n)$ as needed. □

Thus $\phi(A^{|x|}, B)$ only depends on the choice of the types for the tuples:

$$|\phi(A^{|x|}, B)| \leq |S^m_{\mathcal{B}_1, i_1}| \cdot |S^m_{\mathcal{B}_2, i_2}| \cdot \ldots \cdot |S^m_{\mathcal{B}_n, i_n}|$$

Now for each type space we have an inequality

$$|S^m_{\mathcal{B}_j, i_j}| \leq N(m, i_j, \mathcal{L}_B) \leq N(m, |y|, \mathcal{L}_B) \leq N$$

(For singletons $|S^m_{\mathcal{B}_j, i_j}| = 1 \leq N$). Only non-zero indices contribute to the product and there are at most $|y|$ of those (by the equality $i_1 + i_2 + \ldots + i_n = |y|$).

Thus we have

$$|\phi(A^{|x|}, B)| \leq N^{|y|}$$

as needed. $\square$

For subdivisions to work out properly, we will need to work with subsets closed under meets. We observe that the closure under meets doesn't add too many new elements.

Lemma 2.3.3. *Suppose $S \subseteq T$ is a finite subset of size $n \geq 1$ in a meet tree and S' is its closure under meets. Then $|S'| \leq 2n - 1$.*

Proof. We can partition S into connected components and prove the result separately for each component. Thus we may assume all elements of S lie in the same connected component. We prove the claim by induction on n. The base case $n = 1$ is clear. Suppose we have S of size k with closure of size at most $2k - 1$. Take a new point s, and look at its meets with all the elements of S. Pick the smallest one, s'. Then $S \cup \{s, s'\}$ is closed under meets. $\square$

Putting all of those results together we are able to compute the vc-density of formulas in meet trees.

Theorem 2.3.4. *Let $\mathcal{T}$ be an infinite (colored) meet tree and $\phi(x, y)$ a formula with $|x| = p$ and $|y| = q$. Then $\mathrm{vc}(\phi) \leq q$.*

Proof. Pick a finite subset of $S_0 \subset T^p$ of size n. Let $S_1 \subset T$ consist of the components of the elements of S_0. Let $S \subset T$ be the closure of S_1 under meets. Using Lemma 2.3.3 we have $|S| \leq 2|S_1| \leq 2p|S_0| = 2pn = O(n)$. We have $S_0 \subseteq S^p$, so $|\phi(S_0, T^q)| \leq |\phi(S^p, T^q)|$. Thus it is enough to show $|\phi(S^p, T^q)| = O(n^q)$.

Label $S = \{c_i\}_{i \in I}$ with $|I| \leq 2pn$. For every c_i we construct two partitions in the following way. We have that c_i is either minimal in S or it has a predecessor in S (greatest element less than c). If it is minimal, construct $(\mathcal{A}_{c_i}, \mathcal{B}_{c_i})$. If there is a predecessor p, construct $(\mathcal{A}^p_{c_i}, \mathcal{B}^p_{c_i})$. For the second subdivision let G be all the elements in S greater than c_i and construct $(\mathcal{A}^c_G, \mathcal{B}^c_G)$. So far we have constructed two subdivisions for every $i \in I$. Additionally construct $(\mathcal{A}_S, \mathcal{B}_S)$. We end up with a finite collection of proper subdivisons $(\mathcal{A}_j, \mathcal{B}_j)_{j \in J}$ with $|J| = 2|I| + 1$. Before we proceed, we note the following two lemmas describing our partitions.

Lemma 2.3.5. *For all $j \in J$ we have $S \subseteq A_j$. Thus $S \subseteq \bigcap_{j \in J} A_j$ and $S^p \subseteq \bigcap_{j \in J} (A_j)^p$.*

Proof. Check this for each possible choice of partition. Cases for partitions of the type $\mathcal{A}_S, \mathcal{A}^c_G, \mathcal{A}_c$ are easy. Suppose we have a partition $(\mathcal{A}, \mathcal{B}) = (\mathcal{A}^{c_1}_{c_2}, \mathcal{B}^{c_1}_{c_2})$. We need to show that $B \cap S = \emptyset$. By construction we have $c_1, c_2 \notin B$. Suppose we have some other $c \in S$ with $c \in B$. We have $E_{c_1}(c_2, c)$ i.e. there is some b such that $(b > c_1)$, $(b \leq c_2)$ and $(b \leq c)$. Consider the meet $(c \wedge c_2)$. We have $(c \wedge c_2) \geq b > c_1$. Also as $\neg(c \geq c_2)$ we have $(c \wedge c_2) < c_2$. To summarize: $c_2 > (c \wedge c_2) > c_1$. But this contradicts our construction as S is closed under meets, so $(c \wedge c_2) \in S$ and c_1 is supposed to be a predecessor of c_2 in S. $\square$

Lemma 2.3.6. *$\{B_j\}_{j \in J}$ is a disjoint partition of $T - S$ i.e. $T = \bigsqcup_{j \in J} B_j \sqcup S$*

Proof. This more or less follows from the choice of partitions. Pick any $b \in S - T$. Take all the elements in S greater than b and take the minimal one a. Take all the elements in S less than b and take the maximal one c (possible as S is closed under meets). Also take all the elements in S incomparable to b and denote them G. If both a and c exist we have $b \in \mathcal{B}^a_c$. If only the upper bound exists we have $b \in \mathcal{B}_G^a$. If only the lower bound exists we have $b \in \mathcal{B}_c$. If neither exists we have $b \in \mathcal{B}_G$. □

Note 2.3.7. Those two lemmas imply $S = \bigcap_{j \in J} A_j$.

Note 2.3.8. For one-dimensional case $q = 1$ we don't need to do any more work. We have partitioned the parameter space into $|J| = O(n)$ many pieces and over each piece the number of definable sets is uniformly bounded. By Corollary 2.1.8 we have that $|\phi((A_j)^p, B_j)| \le N$ for any $j \in J$ (letting $N = N(n_\phi, q, \mathcal{L} \cup \{S\})$ where n_ϕ is the complexity of ϕ and S is a unary predicate). Compute

$$\begin{aligned}
|\phi(S^p, T)| &= \left| \bigcup_{j \in J} \phi(S^p, B_j) \cup \phi(S^p, S) \right| \le \\
&\le \sum_{j \in J} |\phi(S^p, B_j)| + |\phi(S^p, S)| \le \\
&\le \sum_{j \in J} |\phi((A_j)^p, B_j)| + |S| \le \\
&\le \sum_{j \in J} N + |I| \le \\
&\le (4pn + 1)N + 2pn = (4pN + 2p)n + N = O(n)
\end{aligned}$$

Basic idea for the general case $q \ge 1$ is that we have q parameters and $|J| = O(n)$ many partitions to pick each parameter from giving us $|J|^q = O(n^q)$ choices for the parameter configuration, each giving a uniformly constant number of definable subsets of S. (If every parameter is picked from a fixed partition, Lemma 2.3.1 provides a uniform bound). This yields $\text{vc}(\phi) \le q$ as needed. The rest of the proof is stating this idea formally.

First, we extend our collection of subdivisions $(\mathcal{A}_j, \mathcal{B}_j)_{j \in J}$ by the following singleton sets. For each $c_i \in S$ let $B_i = \{c_i\}$ and $A_i = T$ and add $(\mathcal{A}_i, \mathcal{B}_i)$ to our collection with $\mathcal{L}_B$ the language of B_i interpreted arbitrarily. We end up with a new collection $(\mathcal{A}_k, \mathcal{B}_k)_{k\in K}$ indexed by some K with $|K| = |J| + |I|$ (we added $|S|$ new pairs). Now $\{B_k\}_{k\in K}$ partitions T, so $T = \bigsqcup_{k\in K} B_k$ and $S = \bigcap_{j\in J} A_j = \bigcap_{k\in K} A_k$. For $(k_1, k_2, \ldots k_q) = \vec{k} \in K^q$ denote

$$B_{\vec{k}} = B_{k_1} \times B_{k_2} \times \ldots \times B_{k_q}$$

Then we have the following identity

$$T^q = (\bigsqcup_{k\in K} B_k)^q = \bigsqcup_{\vec{k}\in K^q} B_{\vec{k}}$$

Thus we have that $\{B_{\vec{k}}\}_{\vec{k}\in K^q}$ partition T^q. Compute

$$\begin{aligned} |\phi(S^p, T^q)| = \left| \bigcup_{\vec{k}\in K^q} \phi(S^p, B_{\vec{k}}) \right| \leq \\ \leq \sum_{\vec{k}\in K^q} |\phi(S^p, B_{\vec{k}})| \end{aligned}$$

We can bound $|\phi(S^p, B_{\vec{k}})|$ uniformly using Lemma 2.3.1. $(\mathcal{A}_k, \mathcal{B}_k)_{k\in K}$ satisfies the requirements of the lemma and $B_{\vec{k}}$ looks like B in the lemma after possibly permuting some variables in ϕ. Applying the lemma we get

$$|\phi(S^p, B_{\vec{k}})| \leq N^q$$

with N only depending on q and complexity of ϕ. We complete our computation

$$\begin{aligned}
|\phi(S^p, T^q)| &\leq \sum_{\vec{k}\in K^q} |\phi(S^p, B_{\vec{k}})| \leq \\
&\leq \sum_{\vec{k}\in K^q} N^q \leq \\
&\leq |K^q| N^q \leq \\
&\leq (|J| + |I|)^q N^q \leq \\
&\leq (4pn + 1 + 2pn)^q N^q = N^q(6p + 1/n)^q n^q = O(n^q)
\end{aligned}$$

□

Corollary 2.3.9. *In the theory of infinite (colored) meet trees we have $vc(n) = n$ for all n.*

We get the general result for the trees that aren't necessarily meet trees via an easy application of interpretability.

Corollary 2.3.10. *In the theory of infinite (colored) trees we have $vc(n) = n$ for all n.*

Proof. Let $\mathcal{T}'$ be a tree. We can embed it in a larger tree $\mathcal{T}$ that is closed under meets. Expand $\mathcal{T}$ by an extra color and interpret it by coloring the subset $\mathcal{T}'$. Thus we can interpret $\mathcal{T}'$ in T. By Corollary 3.17 in [2] we get that $\mathrm{vc}^{\mathcal{T}'}(n) \leq \mathrm{vc}^{T}(1 \cdot n) = n$ thus $\mathrm{vc}^{\mathcal{T}'}(n) = n$ as well. □

This settles the question of *vc*-function for trees. Lacking a quantifier elimination result, a lot is still not known. One can try to adapt these techniques to compute the vc-density of a fixed formula, and see if it can take non-integer values. It is also not known whether trees have VC 1 property (see [2] 5.2 for the definition). Our techniques can be used to show that VC 2 property holds but this doesn't give the optimal vc-function.

One can also try to apply similar techniques to more general classes of partially ordered sets. For example, vc-density values are not known for lattices. Similarly, dropping the order, one can look at nicely behaved families of graphs, such as planar graphs or flat graphs. Those are known to be dp-minimal, so one would expect a simple vc-function. It is this author's hope that the techniques developed in this paper can be adapted to yield fruitful results for a more general class of structures.

Chapter 3

vc-density in an additive reduct of the P-adic numbers

Aschenbrenner et. al. computed a linear bound for the vc-density function in the field of p-adic numbers, but it is not known to be optimal. In this paper we investigate a certain P-minimal additive reduct of the field of p-adic numbers and use a cell decomposition result of Leenknegt to compute an optimal bound for that structure.

VC-density was studied in model theory in [2] by Aschenbrenner, Dolich, Haskell, MacPherson, and Starchenko as a natural notion of dimension for definable families of sets in NIP theories. In a complete NIP theory T we can define the vc-function

$$\mathrm{vc}^T = \mathrm{vc} : \mathbb{N} \longrightarrow \mathbb{R} \cup \{\infty\}$$

where $\mathrm{vc}(n)$ measures the worst-case complexity of families of definable sets in an n-fold Cartesian power of the underlying set of a model of T (see 1.1.13 below for a precise definition of vc^T). The simplest possible behavior is $\mathrm{vc}(n) = n$ for all n, satisfied, for example, if T is o-minimal. For $T = \mathrm{Th}(\mathbb{Q}_p)$, the paper [2] computes an upper bound for this function to be $2n-1$, and it is not known whether this is optimal. This same bound holds in any reduct of the field of p-adic numbers, but one may expect that the simplified structure of the reduct would allow a better bound. In [3], Leenknegt provides a cell decomposition result for a certain P-minimal additive reduct of the field of p-adic numbers. Using this result, in this paper we improve the bound for the vc-function, showing that in Leenknegt's structure $\mathrm{vc}(n) = n$.

Section 1 defines vc-density and states some basic lemmas about it. A more in depth exposition of vc-density can be found in [2]. Section 2 defines and states some basic facts about the theory of p-adic numbers. Here we also introduce the reduct which we will be working with. Section 3 sets up basic definitions and lemmas that will be needed for the proof. We define trees and intervals and show how they help with vc-density calculations. Section 4 concludes the proof.

Throughout the paper, variables and tuples of elements will be simply denoted as $x, y, a, b, \ldots$. We will occasionally write $\vec{a}$ instead of a for a tuple in $\mathbb{Q}_p^n$ to emphasize it as an element of the $\mathbb{Q}_p$-vector space $\mathbb{Q}_p^n$. We denote the arity of a tuple x of variables by $|x|$. Natural numbers are $\mathbb{N} = \{0, 1, \ldots\}$.

3.1 P-adic numbers

The field $\mathbb{Q}_p$ of p-adic numbers is often studied in the language of Macintyre

$$\mathcal{L}_{Mac} = \{0, 1, +, -, \cdot, |, \{P_n\}_{n\in\mathbb{N}}\}$$

which is a language $\{0, 1, +, -, \cdot\}$ of rings together with unary predicates P_n interpreted in $\mathbb{Q}_p$ so as to satisfy

$$P_n x \leftrightarrow \exists y \; y^n = x$$

and a divisibility relation where $a|b$ holds in $\mathbb{Q}_p$ when $\operatorname{val} a \leq \operatorname{val} b$.

Note that $P_n \setminus \{0\}$ is a multiplicative subgroup of $\mathbb{Q}_p$ with finitely many cosets.

Theorem 3.1.1 (Macintyre '76)**.** *The $\mathcal{L}_{Mac}$-structure $\mathbb{Q}_p$ has quantifier elimination.*

There is also a cell decomposition result:

Definition 3.1.2. Define <u>k-cells</u> recursively as follows. A <u>0-cell</u> is a singleton subset of $\mathbb{Q}_p$. A <u>$(k+1)$-cell</u> is a subset of $\mathbb{Q}_p^{k+1}$ of the following form:

$$\{(x,t) \in D \times \mathbb{Q}_p \mid \operatorname{val} a_1(x) \;\square_1\; \operatorname{val}(t - c(x)) \;\square_2\; \operatorname{val} a_2(x), t - c(x) \in \lambda P_n\}$$

where D is a k-cell, $a_1(x), a_2(x), c(x)$ are definable functions $D \longrightarrow \mathbb{Q}_p$, each of $\square_i$ is $<, \leq$ or no condition, $n \in \mathbb{N}$, and $\lambda \in \mathbb{Q}_p$.

Theorem 3.1.3 (Denef '84)**.** *Any definable subset of $\mathbb{Q}_p^n$ defined by an $\mathcal{L}_{Mac}$-formula decomposes into a finite disjoint union of n-cells.*

In [2], Aschenbrenner, Dolich, Haskell, Macpherson, and Starchenko show that $\mathbb{Q}_p$ as $\mathcal{L}_{Mac}$-structure satisfies $\mathrm{vc}(n) \leq 2n-1$, however it is not known whether this bound is optimal.

In [3], Leenknegt analyzes the reduct of $\mathbb{Q}_p$ to the language

$$\mathcal{L}_{aff} = \left\{0, 1, +, -, \{\bar{c}\}_{c\in\mathbb{Q}_p}, |, \{Q_{m,n}\}_{m,n\in\mathbb{N}}\right\}$$

where $\bar{c}$ denotes a scalar multiplication by c, $a|b$ as above stands for $\mathrm{val}\, a \leq \mathrm{val}\, b$, and $Q_{m,n}$ is a unary predicate interpreted as

$$Q_{m,n} = \bigcup_{k\in\mathbb{Z}} p^{km}(1 + p^n\mathbb{Z}_p).$$

Note that $Q_{m,n} \setminus \{0\}$ is a subgroup of the multiplicative group of $\mathbb{Q}_p$ with finitely many cosets. One can check that these extra relation symbols are definable in the $\mathcal{L}_{Mac}$-structure $\mathbb{Q}_p$. The paper [3] provides a cell decomposition result with the following cells:

Definition 3.1.4. A 0-cell is a singleton subset of $\mathbb{Q}_p$. A $(k+1)$-cell is a subset of $\mathbb{Q}_p^{k+1}$ of the following form:

$$\{(x,t) \in D \times \mathbb{Q}_p \mid \mathrm{val}\, a_1(x) \,\square_1\, \mathrm{val}(t - c(x)) \,\square_2\, \mathrm{val}\, a_2(x), t - c(x) \in \lambda Q_{m,n}\}$$

where D is a k-cell, called the base of the cell, $a_1(x), a_2(x), c(x)$ are polynomials of degree ≤ 1, called the defining polynomials each of $\square_1, \square_2$ is $<$ or no condition, $m, n \in \mathbb{N}$, and $\lambda \in \mathbb{Q}_p$. We call $\mathbb{Q}_{m,n}$ the defining predicate.

Theorem 3.1.5 (Leenknegt '12)**.** *Any definable subset of $\mathbb{Q}_p^n$ defined by an $\mathcal{L}_{aff}$-formula decomposes into a finite disjoint union of n-cells.*

Moreover, [3] shows that $\mathcal{L}_{aff}$-structure $\mathbb{Q}_p$ is a P-minimal reduct, that is, the one-dimensional definable sets of $\mathcal{L}_{aff}$-structure $\mathbb{Q}_p$ coincide with the one-dimensional definable sets in the full structure $\mathcal{L}_{Mac}$-structure $\mathbb{Q}_p$.

The main result of this paper is the computation of the vc-function for this structure:

Theorem 3.1.6. *$\mathcal{L}_{aff}$-structure $\mathbb{Q}_p$ has* $\mathrm{vc}(n) = n$.

3.2 Key Lemmas and Definitions

To show that $\mathrm{vc}(n) = n$ it suffices to bound $\mathrm{vc}^*(\phi) \leq |x|$ for every $\mathcal{L}_{aff}$-formula $\phi(x; y)$. Fix such a formula $\phi(x; y)$. Instead of working with it directly, we first simplify it using quantifier elimination. The required quantifier elimination result can be easily obtained from cell decomposition:

Lemma 3.2.1. *Any formula $\phi(x; y)$ in $\mathcal{L}_{aff}$-structure $\mathbb{Q}_p$. can be written as a boolean combination of formulas from a collection*

$$\begin{aligned}\Phi(x; y) = &\{\mathrm{val}(p_i(x) - c_i(y)) < \mathrm{val}(p_j(x) - c_j(y))\}_{i,j \in I} \cup \\ &\{p_i(x) - c_i(y) \in \lambda_k Q_{m,n}\}_{i \in I, k \in K}\end{aligned}$$

of $\mathcal{L}_{aff}$-formulas where I, K are finite index sets, each p_i is a degree ≤ 1 polynomial in x without a constant term, each c_i is a degree ≤ 1 polynomial in y, $m, n \in \mathbb{N}$, and $\lambda_k \in \mathbb{Q}_p$.

Proof. Let $l = |x| + |y|$. Partition the subset of $\mathbb{Q}_p^l$ defined by ϕ to obtain $\mathscr{D}^l$, a collection of l-cells. Let $\mathscr{D}^{l-1}$ be the collection of the bases of the cells in $\mathscr{D}^l$. Similarly, construct by induction $\mathscr{D}^i$ for each $0 \leq j < l$, where $\mathscr{D}^j$ is the collection of j-cells which are the bases of cells in $\mathscr{D}^{j+1}$. Set

$$m = \prod \left\{m' \mid Q_{m',n'} \text{ is the defining predicate of a cell in } \mathscr{D}^j \text{ for } 0 \leq j \leq l\right\}$$
$$n = \max \left\{n' \mid Q_{m',n'} \text{ is the defining predicate of a cell in } \mathscr{D}^j \text{ for } 0 \leq j \leq l\right\}$$

This way if a, a' are in the same coset of $Q_{m',n'}$ then they are in the same coset of $Q_{m,n}$. Choose $\{\lambda_k\}_{k\in K}$ to range over all the cosets of $Q_{m,n}$. Let $q_i(x, y)$ enumerate all of the defining polynomials $a_1(x), a_2(x), t - c(x)$ that show up in the cells of $\mathscr{D}^j$ for any j. All if those are all polynomials of degree ≤ 1 in variables x, y. We can split each of them as $q_i(x, y) = p_i(x) - c_i(y)$ where the constant term of q_i goes into c_i. This gives us the appropriate finite collection of formulas Φ. From the cell decomposition it is easy to see that when a, a' have the same Φ-type, then they have the same ϕ-type. Thus ϕ can be written as a boolean combination of formulas from Φ. □

Lemma 3.2.2. *Let $\Phi(x; y)$ be a finite collection of formulas. If ϕ can be written as a boolean combination of formulas from Φ then*

$$\mathrm{vc}^*(\Phi) \leq r \implies \mathrm{vc}^*(\phi) \leq r \quad \text{for all } r \in \mathbb{R}.$$

Proof. If a, a' have the same Φ-type over B, then they have the same ϕ-type over B, where B is some parameter set. Therefore the number of ϕ-types is bounded by the number of Φ-types. The bound follows from Lemma 1.1.11. □

For the remainder of the paper fix $\Phi(x; y)$ to be the collection of formulas defined by Lemma 3.2.1. By the previous lemma, to show that $\mathrm{vc}^*(\phi) \leq |x|$, it suffices to bound $\mathrm{vc}^*(\Phi) \leq |x|$. More precisely, it is sufficient to show that if there is a parameter set B of size N then the number of Φ-types over B is $O(N^{|x|})$. Fix such a parameter set B and work with it from now on. We will compute a bound for the number of Φ-types over B.

Consider the set $T = T(\Phi, B) = \{c_i(b) \mid b \in B, i \in I\} \subset \mathbb{Q}_p$. In this definition B is the parameter set that we have fixed and $c_i(b)$ come from the collection of formulas Φ from the quantifier elimination above. View T as a tree as follows:

Definition 3.2.3.

- For $c \in \mathbb{Q}_p, \alpha \in \mathbb{Z}$ define a ball

$$B(c,\alpha) = \{c' \in \mathbb{Q}_p \mid \mathrm{val}\,(c'-c) > \alpha\}\,.$$

  We also let $B(c,-\infty) = \mathbb{Q}_p$ and $B(c,+\infty) = \emptyset$.

- Define a collection of balls $\mathscr{B} = \{B(t_1, \mathrm{val}(t_1 - t_2))\}_{t_1,t_2 \in T}$. Note that $\mathscr{B}$ is a (directed) boolean algebra of sets in $\mathbb{Q}_p$. We refer to the atoms in that algebra as intervals. Note that the intervals partition $\mathbb{Q}_p$ so any element $a \in \mathbb{Q}_p$ belongs to a unique interval.

- Let's introduce some notation for the intervals. For $t \in T$ and $\alpha_L, \alpha_U \in \mathbb{Z} \cup \{-\infty, +\infty\}$ define

$$\mathrm{I}(t,\alpha_L,\alpha_U) = B(t,\alpha_L) \setminus \bigcup \{B(t',\alpha_U) \mid t' \in T, \mathrm{val}(t'-t) \geq \alpha_U\}$$

  (this is sometimes referred to as the swiss cheese construction). One can check that every interval is of the form $\mathrm{I}(t,\alpha_L,\alpha_U)$ for some values of t, α_L, α_U. The quantities α_L, α_U are uniquely determined by the interval $\mathrm{I}(t,\alpha_L,\alpha_U)$, while t might not be.

- Intervals are a natural construction for trees, however we will require a more refined notion to make Lemma 3.2.12 below work. Define a larger collection of balls

$$\mathscr{B}' = \mathscr{B} \cup \{B(c_i(b), \mathrm{val}(c_j(b) - c_k(b)))\}_{i,j,k \in I, b \in B}\,.$$

  Similar to the previous definition, we define a subinterval to be an atom of the boolean algebra generated by $\mathscr{B}'$. Subintervals refine intervals. Moreover, as before, each subinterval can be written as $\mathrm{I}(t,\alpha_L,\alpha_U)$ for some values of t, α_L, α_U. As before, α_L, α_U are uniquely determined by the subinterval $\mathrm{I}(t, \alpha_L, \alpha_U)$, while t might not be.

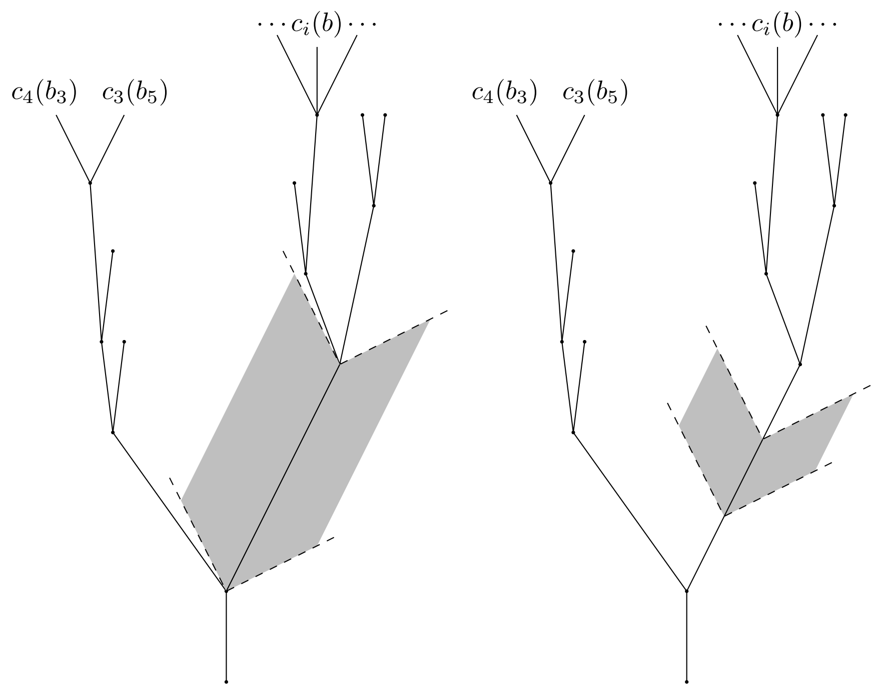

Figure 3.1: A typical interval (left) and subinterval (right) on a tree $\{c_i(b) \mid i \in I, b \in B\}$.

Subintervals are fine enough to make Lemma 3.2.12 below work while coarse enough to be $O(N)$ small:

Lemma 3.2.4.

- *There are at most* $2|T| = 2N|I| = O(N)$ *different intervals.*
- *There are at most* $2|T| + |B| \cdot |I|^3 = O(N)$ *different subintervals.*

Proof. Each new element in the tree T adds at most two intervals to the total count, so by induction there can be at most $2|T|$ many intervals. Each new ball in $\mathscr{B}' \backslash \mathscr{B}$ adds at most one subinterval to the total count, so by induction there are at most $|\mathscr{B}' \backslash \mathscr{B}|$ more subintervals than there are intervals. $\square$

Definition 3.2.5. Suppose $a \in \mathbb{Q}_p$ lies in the interval $\mathrm{I}(t, \alpha_L, \alpha_U)$. Define the T-valuation of a to be $\text{T-val}(a) = \mathrm{val}(a - t)$.

This is a natural notion having the following properties:

Lemma 3.2.6.

(a) $\text{T-val}(a)$ *is well-defined, independent of choice of* t *to represent the interval.*

(b) If $a \in \mathbb{Q}_p$ *lies in the subinterval* $\mathrm{I}(t, \alpha_L, \alpha_U)$, *then* $\text{T-val}(a) = \mathrm{val}(a - t)$.

(c) If $a \in \mathbb{Q}_p$ *lies in the (sub)interval* $\mathrm{I}(t, \alpha_L, \alpha_U)$ *then* $\alpha_L < \text{T-val}(a) \leq \alpha_U$.

(d) For any $a \in \mathbb{Q}_p$ *lying in the (sub)interval* $\mathrm{I}(t, \alpha_L, \alpha_U)$ *and* $t' \in T$*:*

- *If* $\mathrm{val}(t - t') \geq \alpha_U$, *then* $\mathrm{val}(a - t') = \text{T-val}(a)$.
- *If* $\mathrm{val}(t - t') \leq \alpha_L$, *then* $\mathrm{val}(a - t') = \mathrm{val}(t - t')\,(\leq \alpha_L < \text{T-val}(a))$.

Proof. (a)-(c) are clear. For (d) fix $t' \in T$ and suppose $a \in \mathbb{Q}_p$ lies in the subinterval $\mathrm{I}(t, \alpha'_L, \alpha'_U)$. This subinterval lies inside of a unique interval $\mathrm{I}(t, \alpha_L, \alpha_U)$ for some choice of α_L, α_U and by the definition of intervals (or more specifically $\mathscr{B}$):

$$\mathrm{val}(t - t') \geq \alpha_U \iff \mathrm{val}(t - t') \geq \alpha'_U,$$
$$\mathrm{val}(t - t') \geq \alpha_L \iff \mathrm{val}(t - t') \geq \alpha'_L.$$

Therefore without loss of generality we may assume that $a \in \mathbb{Q}_p$ lies in an interval $\mathrm{I}(t, \alpha_L, \alpha_U)$. By (c) and the definition of intervals one of the three following cases has to hold.

Case 1: $\mathrm{val}(t - t') \geq \alpha_U$ and $\text{T-val}(a) < \alpha_U$. Then

$$\mathrm{val}(t - t') \geq \alpha_U > \text{T-val}(a) = \mathrm{val}(a - t),$$

thus $\mathrm{val}(a-t') = \mathrm{val}(a-t) = \text{T-val}(a)$ as needed.

Case 2: $\mathrm{val}(t-t') \geq \alpha_U$ and $\text{T-val}(a) = \alpha_U$. Then

$$\text{T-val}(a) = \mathrm{val}(a-t) = \mathrm{val}(t-t') \geq \alpha_U,$$

thus $\mathrm{val}(a-t') \geq \alpha_U$. The interval $\mathrm{I}(t, \alpha_L, \alpha_U)$ is disjoint from the ball $B(t', \alpha_U)$, so $a \notin B(t', \alpha_U)$, that is, $\mathrm{val}(a-t') \leq \alpha_U$. Combining this with the previous inequality we get that $\mathrm{val}(a-t') = \alpha_U = \text{T-val}(a)$ as needed.

Case 3: $\mathrm{val}(t-t') \leq \alpha_L$. Then

$$\mathrm{val}(t-t') \leq \alpha_L < \text{T-val}(a) = \mathrm{val}(a-t),$$

thus $\mathrm{val}(a-t') = \mathrm{val}(t-t')$ as needed. □

Definition 3.2.7. Suppose $a \in \mathbb{Q}_p$ lies in the subinterval $\mathrm{I}(t, \alpha_L, \alpha_U)$. We say that a is far from the boundary tacitly (of $\mathrm{I}(t, \alpha_L, \alpha_U)$) if

$$\alpha_L + n \leq \text{T-val}(a) \leq \alpha_U - n.$$

Here n is as in Lemma 3.2.1. Otherwise we say that it is close to the boundary (of $\mathrm{I}(t, \alpha_L, \alpha_U)$).

Definition 3.2.8. Suppose $a_1, a_2 \in \mathbb{Q}_p$ lie in the same subinterval $\mathrm{I}(t, \alpha_L, \alpha_U)$. We say a_1, a_2 have the same subinterval type if one of the following holds:

- Both a_1, a_2 are far from the boundary and $a_1 - t, a_2 - t$ are in the same $Q_{m,n}$-coset. (Here $Q_{m,n}$ is an in Lemma 3.2.1.)
- Both a_1, a_2 are close to the boundary and

$$\text{T-val}(a_1) = \text{T-val}(a_2) \leq \mathrm{val}(a_1 - a_2) - n.$$

Definition 3.2.9. For $c \in \mathbb{Q}_p$ and $\alpha, \beta \in \mathbb{Z}, \alpha < \beta$ define $c \upharpoonright [\alpha, \beta)$ to be the record of the coefficients of c for the valuations between $[\alpha, \beta)$. More precisely write c in its power series form

$$c = \sum_{\gamma \in \mathbb{Z}} c_\gamma p^\gamma \text{ with } c_\gamma \in \{0, 1, \ldots, p-1\}$$

Then $c \upharpoonright [\alpha, \beta)$ is just $(c_\alpha, c_{\alpha+1}, \ldots c_{\beta-1}) \in \{0, 1, \ldots, p-1\}^{\beta-\alpha}$.

The following lemma is an adaptation of Lemma 7.4 in [2].

Lemma 3.2.10. *Fix $m, n \in \mathbb{N}$. For any $x, y, c \in \mathbb{Q}_p$, if*

$$\operatorname{val}(x-c) = \operatorname{val}(y-c) \leq \operatorname{val}(x-y) - n,$$

then $x-c, y-c$ are in the same coset of $Q_{m,n}$.

Proof. Call $a, b \in \mathbb{Q}_p$ similar if $\operatorname{val} a = \operatorname{val} b$ and

$$a \upharpoonright [\operatorname{val} a, \operatorname{val} a + n) = b \upharpoonright [\operatorname{val} b, \operatorname{val} b + n).$$

If a, b are similar then

$$a \in Q_{m,n} \iff b \in Q_{m,n}.$$

Moreover for any $\lambda \in \mathbb{Q}_p^\times$, if a, b are similar then so are $\lambda a, \lambda b$. Thus if a, b are similar, then they belong to the same coset of $Q_{m,n}$. The hypothesis of the lemma force $x-c, y-c$ to be similar, thus belonging to the same coset. $\square$

Lemma 3.2.11. *For each subinterval there are at most $K = K(Q_{m,n})$ many subinterval types (with K not depending on B or on the subinterval).*

Proof. Let $a, a' \in \mathbb{Q}_p$ lie in the same subinterval $\mathrm{I}(t, \alpha_L, \alpha_U)$.

Suppose a, a' are far from the boundary. Then they have the same subinterval type if $a - t, a' - t$ are in the same $Q_{m,n}$-coset. So the number of such subinterval types is bounded by the number of $Q_{m,n}$-cosets.

Suppose a, a' are close to the boundary and

$$\text{T-val}(a) - \alpha_L = \text{T-val}(a') - \alpha_L < n \text{ and}$$
$$a \restriction [\text{T-val}(a), \text{T-val}(a) + n) = a' \restriction [\text{T-val}(a'), \text{T-val}(a') + n).$$

Then a, a' have the same subinterval type. Such a subinterval type is thus determined by $\text{T-val}(a) - \alpha_L$ and the tuple $a \restriction [\text{T-val}(a), \text{T-val}(a) + n)$, therefore there are at most np^n many such types.

A similar argument works for a with $\alpha_U - \text{T-val}(a) \leq n$.

Adding all this up we get that there are at most

$$K = (\text{number of } Q_{m,n} \text{ cosets}) + 2np^n$$

many subinterval types. $\square$

The following critical lemma relates tree notions to Φ-types.

Lemma 3.2.12. *Suppose $d, d' \in \mathbb{Q}_p^{|x|}$ satisfy the follwing three conditions:*

- *For all $i \in I$ $p_i(d)$ and $p_i(d')$ are in the same subinterval.*
- *For all $i \in I$ $p_i(d)$ and $p_i(d')$ have the same subinterval type.*
- *For all $i, j \in I$, $\text{T-val}(p_i(d)) > \text{T-val}(p_j(d))$ iff $\text{T-val}(p_i(d')) > \text{T-val}(p_j(d'))$.*

Then d, d' have the same Φ-type over B.

Proof. There are two kinds of formulas in Φ (see Lemma 3.2.1). First we show that d, d' agree on formulas of the form $p_i(x) - c_i(y) \in \lambda_k Q_{m,n}$. It is enough to show that for every $i \in I, b \in B$, $p_i(d) - c_i(b), p_i(d') - c_i(b)$ are in the same

$Q_{m,n}$-coset. Fix such i, b. For brevity let $a = p_i(d), a' = p_i(d')$ and $Q = Q_{m,n}$. We want to show that $a - c_i(b), a' - c_i(b)$ are in the same Q-coset.

Suppose a, a' are close to the boundary. Then $\text{T-val}(a) = \text{T-val}(a') \leq \text{val}(a - a') - n$. Using Lemma 3.2.6d, we have

$$\text{val}(a - c_i(b)) = \text{val}(a' - c_i(b)) \leq \text{T-val}(a) \leq \text{val}(a - a') - n.$$

Lemma 3.2.10 shows that $a - c_i(b), a' - c_i(b)$ are in the same Q-coset.

Now, suppose both a, a' are far from the boundary. Let $\mathrm{I}(t, \alpha_L, \alpha_U)$ be the interval containing a, a'. Then we have

$$\alpha_L + n \leq \text{val}(a - t) \leq \alpha_U - n,$$
$$\alpha_L + n \leq \text{val}(a' - t) \leq \alpha_U - n$$

(as being far from the subinterval's boundary also makes a, a' far from interval's boundary). We have either $\text{val}(t - c_i(b)) \geq \alpha_U$ or $\text{val}(t - c_i(b)) \leq \alpha_L$ (as otherwise it would contradict the definition of intervals, or more specifically $\mathscr{B}$).

Suppose it is the first case $\text{val}(t - c_i(b)) \geq \alpha_U$. Then using Lemma 3.2.6d

$$\text{val}(a - c_i(b)) = \text{val}(a - t) \leq \alpha_U - n \leq \text{val}(t - c_i(b)) - n.$$

So by Lemma 3.2.10 $a - c_i(b), a - t$ are in the same Q-coset. By an analogous argument, $a' - c_i(b), a' - t$ are in the same Q-coset. As a, a' have the same subinterval type, $a - t, a' - t$ are in the same Q-coset. Thus by transitivity we get that $a - c_i(b), a' - c_i(b)$ are in the same Q-coset.

For the second case, suppose $\text{val}(t - c_i(b)) \leq \alpha_L$. Then using Lemma 3.2.6d

$$\text{val}(a - c_i(b)) = \text{val}(t - c_i(b)) \leq \alpha_L \leq \text{val}(a - t) - n,$$

so by Lemma 3.2.10, $a - c_i(b), t - c_i(b)$ are in the same Q-coset. Similarly $a' - c_i(b), t - c_i(b)$ are in the same Q-coset. Thus by transitivity we get that $a - c_i(b), a' - c_i(b)$ are in the same Q-coset.

Next, we need to show that d, d' agree on formulas of the form $\mathrm{val}(p_i(x) - c_i(y)) < \mathrm{val}(p_j(x) - c_j(y))$ (again, referring to the presentation in Lemma 3.2.1). Fix $i, j \in I, b \in B$. We would like to show the following equivalence:

$$\mathrm{val}(p_i(d) - c_i(b)) < \mathrm{val}(p_j(d) - c_j(b)) \iff \\ \iff \mathrm{val}(p_i(d') - c_i(b)) < \mathrm{val}(p_j(d') - c_j(b)) \quad (3.2.1)$$

Suppose $p_i(d), p_i(d')$ are in the subinterval $\mathrm{I}(t_i, \alpha_i, \beta_i)$ and $p_j(d), p_j(d')$ are in the subinterval $\mathrm{I}(t_j, \alpha_j, \beta_j)$. Lemma 3.2.6d yields the following four cases.

Case 1:

$$\mathrm{val}(p_i(d) - c_i(b)) = \mathrm{val}(p_i(d') - c_i(b)) = \mathrm{val}(t_i - c_i(b))$$
$$\mathrm{val}(p_j(d) - c_j(b)) = \mathrm{val}(p_j(d') - c_j(b)) = \mathrm{val}(t_j - c_j(b))$$

Then it is clear that the equivalence (3.2.1) holds.

Case 2:

$$\mathrm{val}(p_i(d) - c_i(b)) = \text{T-val}(p_i(d)) \text{ and } \mathrm{val}(p_i(d') - c_i(b)) = \text{T-val}(p_i(d'))$$
$$\mathrm{val}(p_j(d) - c_j(b)) = \text{T-val}(p_j(d)) \text{ and } \mathrm{val}(p_j(d') - c_j(b)) = \text{T-val}(p_j(d'))$$

Then the equivalence (3.2.1) holds by the third hypothesis of the lemma (that order of T-valuations is preserved).

Case 3:

$$\mathrm{val}(p_i(d) - c_i(b)) = \mathrm{val}(p_i(d') - c_i(b)) = \mathrm{val}(t_i - c_i(b))$$
$$\mathrm{val}(p_j(d) - c_j(b)) = \text{T-val}(p_j(d)) \text{ and } \mathrm{val}(p_j(d') - c_j(b)) = \text{T-val}(p_j(d'))$$

If $p_j(d), p_j(d')$ are close to the boundary, then $\text{T-val}(p_j(d)) = \text{T-val}(p_j(d'))$ and the equivalence (3.2.1) clearly holds. Suppose then that $p_j(d), p_j(d')$ are far from the boundary.

$$\alpha_j + n \leq \text{T-val}(p_j(d)), \text{T-val}(p_j(d')) \leq \beta_j - n$$
$$\alpha_j < \text{T-val}(p_j(d)), \text{T-val}(p_j(d')) < \beta_j$$

and $\mathrm{val}(t_i - c_i(b))$ lies outside of the (α_j, β_j) by the definition of subinterval (more specifically definition of $\mathscr{B}'$). Therefore (3.2.1) has to hold. (Note that we always have $\text{T-val}(p_j(d)), \text{T-val}(p_j(d')) \in (\alpha_j, \beta_j]$ by Lemma 3.2.6c, so we only need the condition on being far from the boundary to avoid the edge case of equality to β_j.)

Case 4:

$$\mathrm{val}(p_i(d) - c_i(b)) = \text{T-val}(p_i(d)) \text{ and } \mathrm{val}(p_i(d') - c_i(b)) = \text{T-val}(p_i(d'))$$
$$\mathrm{val}(p_j(d) - c_j(b)) = \mathrm{val}(p_j(d') - c_j(b)) = \mathrm{val}(t_j - c_j(b))$$

Similar to case 3 (switching i, j). $\square$

The previous lemma gives us an upper bound on the number of types - there are at most $|2I|!$ many choices for the order of T-val, $O(N)$ many choices for the subinterval for each p_i, and K many choices for the subinterval type for each p_i (where K is as in Lemma 3.2.11), giving a total of $O(N^{|I|}) \cdot K^{|I|} \cdot |I|! = O(N^{|I|})$ many types. This implies $\mathrm{vc}^*(\Phi) \leq |I|$. The biggest contribution to this bound are the choices among the $O(N)$ many subintervals for each p_i with $i \in I$. Are all of those choices realized? Intuitively there are $|x|$ many variables and $|I|$ many equations, so once we choose a subinterval for $|x|$ many p_i's, the subintervals for the rest should be determined. This would give the required bound $\mathrm{vc}^*(\Phi) \leq |x|$. The next section outlines this idea formally.

3.3 Main Proof

An alternative way to write $p_i(c)$ is as a scalar product $\vec{p}_i \cdot \vec{c}$, where $\vec{p}_i$ and $\vec{c}$ are vectors in $\mathbb{Q}_p^{|x|}$ (as $p_i(x)$ is homogeneous linear).

Lemma 3.3.1. *Suppose we have a finite collection of vectors $\{\vec{p}_j\}_{j \in J}$ with each $\vec{p}_j \in \mathbb{Q}_p^{|x|}$. Suppose $\vec{p} \in \mathbb{Q}_p^{|x|}$ satisfies $\vec{p} \in \operatorname{span}\{\vec{p}_j\}_{j\in J}$, and we have $\vec{c} \in \mathbb{Q}_p^{|x|}, \alpha \in \mathbb{Z}$ with $\operatorname{val}(\vec{p}_j \cdot \vec{c}) > \alpha$ for all $j \in J$. Then $\operatorname{val}(\vec{p} \cdot \vec{c}) > \alpha - \gamma$ for some $\gamma \in \mathbb{N}$. Moreover γ can be chosen independently from $\vec{c}, \alpha$ depending only on $\{\vec{p}_j\}_{j\in J}$.*

Proof. For some $c_j \in \mathbb{Q}_p$ for $j \in J$ we have $\vec{p} = \sum_{j\in J} c_j \vec{p}_j$, hence $\vec{p} \cdot \vec{c} = \sum_{j \in J} c_j \vec{p}_j \cdot \vec{c}$. Thus

$$\operatorname{val}(c_j \vec{p}_j \cdot \vec{c}) = \operatorname{val}(c_j) + \operatorname{val}(\vec{p}_j \cdot \vec{c}) > \operatorname{val}(c_j) + \alpha.$$

Let $\gamma = \max(0, -\max_{j \in J} \operatorname{val}(c_j))$. Then we have

$$\operatorname{val}(\vec{p} \cdot \vec{c}) = \operatorname{val}\left(\sum_{j\in J} c_j \vec{p}_j \cdot \vec{c}\right) \geq$$
$$\geq \min_{j \in J} \operatorname{val}(\sum_{j \in J} c_j \vec{p}_j \cdot \vec{c}) > \min_{j\in J} \operatorname{val}(c_j) + \alpha \geq \alpha - \gamma$$

as required. ∎

Corollary 3.3.2. *Suppose we have a finite collection of vectors $\{\vec{p}_i\}_{i\in I}$ with each $\vec{p}_i \in \mathbb{Q}_p^{|x|}$. Suppose $J \subseteq I$ and $i \in I$ satisfy $\vec{p}_i \in \operatorname{span}\{\vec{p}_j\}_{j\in J}$, and we have $\vec{c} \in \mathbb{Q}_p^{|x|}, \alpha \in \mathbb{Z}$ with $\operatorname{val}(\vec{p}_j \cdot \vec{c}) > \alpha$ for all $j \in J$. Then $\operatorname{val}(\vec{p}_i \cdot \vec{c}) > \alpha - \gamma$ for some $\gamma \in \mathbb{N}$. Moreover γ can be chosen independently from $J, j, \vec{c}, \alpha$ depending only on $\{\vec{p}_i\}_{i\in I}$.*

Proof. The previous lemma shows that we can pick such γ for a given choice of i, J, but independent from $\alpha, \vec{c}$. To get a choice independent from i, J, go over all such eligible choices (i ranges over I and J ranges over subsets of I), pick γ for each, and then take the maximum of those values. □

Fix γ according to Corollary 3.3.2 corresponding to $\{\vec{p}_i\}_{i\in I}$ given by our collection of formulas Φ. (The lemma above is a general result, but we only use it applied to the vectors given by Φ.)

Definition 3.3.3. Suppose $a \in \mathbb{Q}_p$ lies in the subinterval $\mathrm{I}(t, \alpha_L, \alpha_U)$. Define the $\underline{T\text{-floor}}$ of a to be $\text{T-fl}(a) = \alpha_L$.

Definition 3.3.4. Let $f : \mathbb{Q}_p^{|x|} \longrightarrow \mathbb{Q}_p^I$ with $f(c) = (p_i(c))_{i\in I}$. Define the segment space Sg to be the image of f. Equvalently:

$$\mathrm{Sg} = \left\{(p_i(c))_{i\in I} \mid c \in \mathbb{Q}_p^{|x|}\right\} \subseteq \mathbb{Q}_p^I$$

Without loss of generality, we may assume that $I = \{1, 2, \ldots, k\}$ (that is the formulas are labeled by consecutive natural numbers). Given a tuple $(a_i)_{i\in I}$ in the segment space, look at the corresponding T-floors $\{\text{T-fl}(a_i)\}_{i\in I}$ and T-valuations $\{\text{T-val}(a_i)\}_{i\in I}$. Partition the segment space by the order types of $\{\text{T-fl}(a_i)\}_{i\in I}$ and $\{\text{T-val}(a_i)\}_{i\in I}$ (as subsets of $\mathbb{Z}$).

Work in a fixed set Sg' of the partition. After relabeling the p_i we may assume that

$$\text{T-fl}(a_1) \geq \text{T-fl}(a_2) \geq \dots \text{ for all } a_i \in \mathrm{Sg}'$$

Consider the (relabeled) sequence of vectors $\vec{p}_1, \vec{p}_2, \dots, \vec{p}_I$. There is a unique subset $J \subset I$ such that the set of all vectors with indices in J is linearly independent, and all vectors with indices outside of J are a linear combination of preceding vectors. (We can pick those using a greedy algorithm for finding a linearly independent subset of vectors.) We call indices in I independent and we call the indices in $I \setminus J$ dependent.

Definition 3.3.5.

- Denote $\{0, 1, \dots, p-1\}$ as Ct.
- Let Tp be the space of all subinterval types. By Lemma 3.2.11 we have $|\,\mathrm{Tp}\,| \leq K$.
- Let Sub be the space of all subintervals. By Lemma 3.2.4 we have $|\,\mathrm{Sub}\,| \leq 3|I|^2 \cdot N = O(N)$.

Definition 3.3.6. Now, we define a function

$$g_{\mathrm{Sg}'} : \mathrm{Sg}' \longrightarrow \mathrm{Tp}^I \times \mathrm{Sub}^J \times \mathrm{Ct}^{I \setminus J}$$

as follows:

Let $a = (a_i)_{i \in I} \in \mathrm{Sg}'$. To define $g_{\mathrm{Sg}'}(a)$ we need to specify where it maps a in each individual component of the product.

For each a_i record its subinterval type, giving the first component in Tp^I.

For a_j with $j \in J$, record the subinterval of a_j, giving the second component in Sub^J.

For the third component (an element of $\mathrm{Ct}^{I\setminus J}$) do the following computation. Pick a_i with i dependent. Let j be the largest independent index with $j < i$. Record $a_i \upharpoonright [\text{T-fl}(a_j) - \gamma, \text{T-fl}(a_j))$.

Combine $g_{\mathrm{Sg}'}$ for all the partitions to get a function

$$g : \mathrm{Sg} \longrightarrow \mathrm{Tp}^I \times \mathrm{Sub}^J \times \mathrm{Ct}^{I\setminus J} .$$

Lemma 3.3.7. *Suppose we have $c, c' \in \mathbb{Q}_p^{|x|}$ such that $f(c), f(c')$ are in the same set Sg' of the partition of Sg and $g(f(c)) = g(f(c'))$. Then c, c' have the same Φ-type over B.*

Proof. Let $a_i = \vec{p}_i \cdot \vec{c}$ and $a'_i = \vec{p}_i \cdot \vec{c}\,'$ so that

$$\begin{aligned} f(c) &= (p_i(c))_{i\in I} = (\vec{p}_i \cdot \vec{c})_{i\in I} = (a_i)_{i\in I} \\ f(c') &= (p_i(c'))_{i\in I} = (\vec{p}_i \cdot \vec{c}\,')_{i\in I} = (a'_i)_{i\in I} \end{aligned}$$

For each i we show that a_i, a'_i are in the same subinterval and have the same subinterval type, so the conclusion follows by Lemma 3.2.12 ($f(c), f(c')$ are in the same partition ensuring the proper order of T-valuations for the 3rd condition of the lemma). Tp records the subinterval type of each element, so if $g(\bar{a}) = g(\bar{a}')$ then a_i, a'_i have the same subinterval type for all $i \in I$. Thus it remains to show that a_i, a'_i lie in the same subinterval for all $i \in I$. Suppose i is an independent index. Then by construction, Sub records the subinterval for a_i, a'_i, so those have to belong to the same subinterval. Now suppose i is dependent. Pick the largest $j < i$ such that j is independent. We have $\text{T-fl}(a_i) \leq \text{T-fl}(a_j)$ and $\text{T-fl}(a'_i) \leq \text{T-fl}(a'_j)$. Moreover $\text{T-fl}(a_j) = \text{T-fl}(a'_j)$ as a_j, a'_j lie in the same subinterval (using the earlier part of the argument as j is independent).

Claim 3.3.8. $\mathrm{val}(a_i - a_i') > \text{T-fl}(a_j) - \gamma$

Proof. Let K be the set of the independent indices less than i. Note that by the definition for dependent indices we have $\vec{p}_i \in \operatorname{span}\{\vec{p}_k\}_{k\in K}$. We also have

$$\mathrm{val}(a_k - a_k') > \text{T-fl}(a_k) \text{ for all } k \in K$$

as a_k, a_k' lie in the same subinterval (using the earlier part of the argument as k is independent). Now $\mathrm{val}(a_k - a_k') > \text{T-fl}(a_j)$ for all $k \in K$ by monotonicity of $\text{T-fl}(a_k)$. Moreover $a_k - a_k' = \vec{p}_k \cdot \vec{c} - \vec{p}_k \cdot \vec{c}' = \vec{p}_k \cdot (\vec{c} - \vec{c}')$. Combining the two, we get that $\mathrm{val}(\vec{p}_k \cdot (\vec{c} - \vec{c}')) > \text{T-fl}(a_j)$ for all $k \in K$. Now observe that $K \subset I, i \in I, \vec{c} - \vec{c}' \in \mathbb{Q}_p^{|x|}, \text{T-fl}(a_j) \in \mathbb{Z}$ satisfy the requirements of Lemma 3.3.2, so we apply it to obtain $\mathrm{val}(\vec{p}_i \cdot (\vec{c} - \vec{c}')) > \text{T-fl}(a_j) - \gamma$. Similarly to before, we have $\vec{p}_i \cdot (\vec{c} - \vec{c}') = \vec{p}_i \cdot \vec{c} - \vec{p}_i \cdot \vec{c}' = a_i - a_i'$. Therefore we can conclude that $\mathrm{val}(a_i - a_i') > \text{T-fl}(a_j) - \gamma$ as needed, finishing the proof of the claim. $\square$

Additionally a_i, a_i' have the same image in the Ct component, so we have

$$\mathrm{val}(a_i - a_i') > \text{T-fl}(a_j).$$

We now would like to show that a_i, a_i' lie in the same subinterval. As $\text{T-fl}(a_i) \leq \text{T-fl}(a_j)$, $\text{T-fl}(a_i') \leq \text{T-fl}(a_j')$ and $\text{T-fl}(a_j) = \text{T-fl}(a_j')$ we have that $\mathrm{val}(a_i - a_i') > \text{T-fl}(a_i)$ and $\mathrm{val}(a_i - a_i') > \text{T-fl}(a_i')$. Suppose that a_i lies in the subinterval $\mathrm{I}(t, \text{T-fl}(a_i), \alpha_U)$ and that a_i' lies in the subinterval $\mathrm{I}(t', \text{T-fl}(a_i'), \alpha_U')$. Without loss of generality assume that $\text{T-fl}(a_i) \leq \text{T-fl}(a_i')$. As $\mathrm{val}(a_i - a_i') > \text{T-fl}(a_i')$, this implies that $a_i \in B(a_i', \text{T-fl}(a_i'))$. Then $a_i \in B(t', \text{T-fl}(a_i'))$ as $\mathrm{val}(a_i - t') > \text{T-fl}(a_i')$. This implies that $B(t, \text{T-fl}(a_i)) \cap B(t', \text{T-fl}(a_i')) \neq \emptyset$ as they both contain a_i. As balls are directed, the non-zero intersection means that one ball has to be contained in another. Given our assumption that $\text{T-fl}(a_i) \leq \text{T-fl}(a_i')$, we have $B(t, \text{T-fl}(a_i)) \subset B(t', \text{T-fl}(a_i'))$. For the subintervals to be disjoint we need $\text{I}(t, \text{T-fl}(a_i), \alpha_U) \cap B(t', \text{T-fl}(a_i')) = \emptyset$. But $\text{val}(t' - a_i) > \text{T-fl}(a_i')$ implying that $a_i \in \text{I}(t, \text{T-fl}(a_i), \alpha_U) \cap B(t', \text{T-fl}(a_i'))$ giving a contradiction. Therefore the subintervals coincide finishing the proof. $\square$

Corollary 3.3.9. *The dual* vc*-density of* $\Phi(x, y)$ *is* $\leq |x|$.

Proof. Suppose we have $c, c' \in \mathbb{Q}_p^{|x|}$ such that $f(c), f(c')$ are in the same partition and $g(f(c)) = g(f(c'))$. Then by the previous lemma c, c' have the same Φ-type. Thus the number of possible Φ-types is bounded by the size of the range of g times the number of possible partitions

$$(\text{number of partitions}) \cdot |\operatorname{Tp}|^{|I|} \cdot |\operatorname{Sub}|^{|J|} \cdot |\operatorname{Ct}|^{|I-J|}.$$

There are at most $(|2I|!)^2$ many partitions of Sg, so in the product above, the only component dependent on B is

$$|\operatorname{Sub}|^{|J|} \leq (N \cdot 3|I|^2)^{|J|} = O(N^{|J|}).$$

Every p_i is an element of a $|x|$-dimensional vector space, so there can be at most $|x|$ many independent vectors. Thus we have $|J| \leq |x|$ and the bound follows. $\square$

Corollary 3.3.10 (Theorem 3.1.6)**.** *The* $\mathcal{L}_{aff}$*-structure* $\mathbb{Q}_p$ *satisfies* $\text{vc}(n) = n$.

Proof. The previous lemma implies that $\text{vc}^*(\phi) \leq \text{vc}^*(\Phi) \leq |x|$. As choice of ϕ was arbitrary, this implies that the vc-density of any formula is bounded by the arity of x. $\square$

This proof relies heavily on the linearity of the defining polynomials a_1, a_2, c in the cell decomposition result (see Definition 3.1.4). Linearity is used to separate the x and y variables as well as for Corollary 3.3.2 to reduce the number of independent factors from $|I|$ to $|x|$. The paper [3] has cell decomposition results for more expressive reducts of $\mathbb{Q}_p$, including, for example, restricted multiplication. While our results don't apply to it directly, it is this author's hope that similar techniques can be used to also compute the vc-function for those structures.

Another interesting question whether the reduct studied in this paper has VC 1 property (see [2] 5.2 for the definition). If so, this would imply the linear vc-density bound directly. The paper [2] implies that the reduct has VC 2 property. While there are techniques for showing that a structure has a given VC property, less is known about showing that a structure doesn't have a given VC property. Perhaps the simple structure of the $\mathcal{L}_{aff}$-reduct can help understand this property better.

Chapter 4

shelah-spencer graphs

We investigate vc-density in Shelah-Spencer graphs. We provide an upper bound on formula-by-formula basis and show that there isn't a uniform lower bound forcing $\mathrm{vc}(n) = \infty$.

VC-density was studied in [2] by Aschenbrenner, Dolich, Haskell, MacPherson, and Starchenko as a natural notion of dimension for NIP theories. In an NIP theory we can define a vc-function

$$\mathrm{vc} : \mathbb{N} \longrightarrow \mathbb{N}$$

Where $\mathrm{vc}(n)$ measures the worst-case complexity of definable sets in an n-dimensional space. Simplest possible behavior is $\mathrm{vc}(n) = n$ for all n. Theories with the property that $\mathrm{vc}(1) = 1$ are known to be dp-minimal, i.e. having the smallest possible dp-rank. In general, it is not known whether there can be a dp-minimal theory which doesn't satisfy $\mathrm{vc}(n) = n$.

In this paper, we investigate vc-density of definable sets in Shelah-Spencer graphs. In our description of Shelah-Spencer graphs we follow closely the treatment in [1]. A Shelah-Spencer graph is a limit of random structures $G(n, n^{-\alpha})$ for an irrational $\alpha \in (0,1)$. $G(n, n^{-\alpha})$ is a random graph on n vertices with edge probability $n^{-\alpha}$.

Our first result is that in Shelah-Spencer graphs

$$\mathrm{vc}(n) = \infty$$

which implies that they are not dp-minimal. Our second result is providing an upper bound on a vc-density for a formula ϕ

$$\mathrm{vc}(\phi) \leq K(\phi)\frac{Y(\phi)}{\epsilon(\phi)}$$

where $K(\phi), Y(\phi), \epsilon(\phi)$ are paramters easily computable from the quantifier free form of ϕ.

Chapter 1 introduces basic facts about VC-dimension and vc-density. More can be found in [2]. Chapter 2 summarizes notation and basic facts concerning Shelah-Spencer graphs. We direct the reader to [1] for a more in-depth treatment. In chapter 3 we introduce some measure of dimension for quantifier free formulas as well as proving some elementary facts about it. Chapter 4 computes a lower bound for vc-density to demonstrate that $\mathrm{vc}(n) = \infty$. Chapter 5 computes an upper bound for vc-density on a formula-by-formula basis.

4.1 Graph Combinatorics

We denote graph by $\mathcal{A}$, set of its vertices by A.

Definition 4.1.1. Fix $\alpha \in (0,1)$, irrational.

- For a finite graph $\mathcal{A}$ let

$$\delta(\mathcal{A}) = |A| - \alpha e(\mathcal{A})$$

where $e(\mathcal{A})$ is the number of edges in $\mathcal{A}$.

- For finite $\mathcal{A}, \mathcal{B}$ with $\mathcal{A} \subseteq \mathcal{B}$ define $\delta(\mathcal{B}/\mathcal{A}) = \delta(\mathcal{B}) - \delta(\mathcal{A})$.

- We say that $\mathcal{A} \leq \mathcal{B}$ if $\mathcal{A} \subseteq \mathcal{B}$ and $\delta(\mathcal{A}'/\mathcal{B}) > 0$ for all $\mathcal{A} \subseteq \mathcal{A}' \subsetneq \mathcal{B}$.

- We say that finite $\mathcal{A}$ is <u>positive</u> if for all $\mathcal{A}' \subseteq \mathcal{A}$ we have $\delta(\mathcal{A}') \geq 0$.

- We work in theory S_α axiomatized by
  - Every finite substructure is positive.
  - For a model $\mathscr{M}$ given $\mathcal{A} \leq \mathcal{B}$ every embedding $f : \mathcal{A} \longrightarrow \mathscr{M}$ extends to $g : \mathcal{B} \longrightarrow \mathscr{M}$.

- For $\mathcal{A}, \mathcal{B}$ positive, $(\mathcal{A}, \mathcal{B})$ is called a <u>minimal pair</u> if $\mathcal{A} \subseteq \mathcal{B}$, $\delta(\mathcal{B}/\mathcal{A}) < 0$ but $\delta(\mathcal{A}'/\mathcal{A}) \geq 0$ for all proper $\mathcal{A} \subseteq \mathcal{A}' \subsetneq \mathcal{B}$.

- $\langle \mathcal{A}_i \rangle_{i \leq m}$ is called a <u>minimal chain</u> if $(\mathcal{A}_i, \mathcal{A}_i + 1)$ is a minimal pair (for all $i < m$).

- For a positive $\mathcal{A}$ let $\delta_{\mathcal{A}}(\bar{x})$ be the atomic diagram of $\mathcal{A}$. For positive $\mathcal{A} \subset \mathcal{B}$ let

$$\Psi_{\mathcal{A},\mathcal{B}}(\bar{x}) = \delta_{\mathcal{A}}(\bar{x}) \wedge \exists \bar{y}\ \delta_{\mathcal{B}}(\bar{x}, \bar{y}).$$

Such formula is called a <u>chain-minimal extension formula</u> if in addition we have that there is a minimal chain starting at $\mathcal{A}$ and ending in $\mathcal{B}$. Denote such formulas as $\Psi_{\langle \mathcal{M}_i \rangle}$.

Theorem 4.1.2 (5.6 in [1]). *S_α admits quantifier elimination down to boolean combination of chain-minimal extension formulas.*

4.2 Basic Definitions and Lemmas

Fix tuples $x = (x_1, \dots x_n), y = (y_1, \dots, y_m)$. We refer to chain-minimal extension formulas as basic formulas. Let $\phi_{\langle \mathcal{M}_i \rangle}(x, y)$ be a basic formula.

Definition 4.2.1. Define $\mathcal{X}$ to be the graph on vertices $\{x_i\}$ with edges as defined by $\phi_{\langle \mathcal{M}_i \rangle}$. Similarly define $\mathcal{Y}$. We define those abstractly, i.e. on a new set of vertices disjoint from $\mathscr{M}$.

Note that $\mathcal{X}$, $\mathcal{Y}$ are positive as they are subgraphs of $\mathcal{M}_0$. As usual X, Y will refer to vertices of those graphs.

We restrict our attention to formulas that define no edges between X and Y.

Note 4.2.2. We can handle edges between x and y as separate elements of the minimal chain extension.

Definition 4.2.3. For a basic formula $\phi = \phi_{\langle \mathcal{M}_i \rangle_{i \le k}}(x, y)$ let

- $\epsilon_i(\phi) = -\dim(M_i/M_{i-1})$.
- $\epsilon_L(\phi) = \sum_{[1..k]} \epsilon_i(\phi)$.
- $\epsilon_U(\phi) = \min_{[1..k]} \epsilon_i(\phi)$.
- Let $\mathcal{Y}'$ be a subgraph of $\mathcal{Y}$ induced by vertices of $\mathcal{Y}$ that are connected to $M_k - (X \cup Y)$.
- Let $Y(\phi) = \dim(\mathcal{Y}')$. In particular if $\mathcal{Y} = \mathcal{Y}'$ and $\mathcal{Y}$ is disconnected then $Y(\phi)$ is just the arity of the tuple y.

We conclude this section by stating a couple of technical lemmas that will be useful in our proofs later.

Lemma 4.2.4. *Suppose we have a set B and a minimal pair (M, A) with $A \subset B$ and $\dim(M/A) = -\epsilon$. Then either $M \subseteq B$ or $\dim((M \cup B)/B) < -\epsilon$.*

Proof. By diamond construction

$$\dim((M \cup B)/B) \leq \dim(M/(M \cap B))$$

and

$$\dim(M/(M \cap B)) = \dim(M/A) - \dim(M/(M \cap B))$$

$$\dim(M/A) = -\epsilon$$

$$\dim(M/(M \cap B)) > 0$$

□

Lemma 4.2.5. *Suppose we have a set B and a minimal chain M_n with $M_0 \subset B$ and dimensions $-\epsilon_i$. Let ϵ be the minimal of ϵ_i. Then either $M_n \subseteq B$ or $\dim((M_n \cup B)/B) < -\epsilon$.*

Proof. Let $\bar{M}_i = M_i \cup B$

$$\dim(\bar{M}_n/B) = \dim(\bar{M}_n/\bar{M}_{n-1}) + \ldots + \dim(\bar{M}_2/\bar{M}_1) + \dim(\bar{M}_1/B)$$

Either $M_n \subseteq B$ or one of the summands above is nonzero. Apply previous lemma. □

Lemma 4.2.6. *Suppose we have a minimal chain M_n with dimensions $-\epsilon_i$. Let ϵ be the sum of all ϵ_i. Suppose we have some B with $B \subseteq M_n$. Then*

$\dim B/(M_0 \cap B) \geq -\epsilon$.

Proof. Let $B_i = B \cap M_i$. We have $\dim B_{i+1}/B_i \geq \dim M_{i+1}/M_i$ by minimality. $\dim B/(M_0 \cap B) = \dim B_n/B_0 = \sum \dim B_{i+1}/B_i \geq -\epsilon$. □

4.3 Lower bound

As a simplification for our lower bound computation we assume that all the basic formulas involved we have $\mathcal{Y}' = \mathcal{Y}$ (see Definition 4.2.3).

We work with formulas that are boolean combinations of basic formulas written in disjunctive-conjunctive form. First, we extend our definition of ϵ.

Definition 4.3.1 (Negation)**.** If ϕ is a basic formula, then define

$$\epsilon_L(\neg\phi) = \epsilon_L(\phi)$$

Definition 4.3.2 (Conjunction)**.** Take a collection of formulas $\phi_i(x, y)$ where each ϕ_i is positive or negative basic formula. If both positive and negative formulas are present then $\epsilon_L(\phi) = \infty$. We don't have a lower bound for that case. If different formulas define $\mathcal{X}$ or $\mathcal{Y}$ differently then $\epsilon_L(\phi) = \infty$. In that case of the conflicting definitions would make the formula have no realizations. Otherwise

$$\epsilon_L(\bigwedge \phi_i) = \sum \epsilon_L(\phi_i)$$

Definition 4.3.3 (Disjunction)**.** Take a collection of formulas ψ_i where each instance is a conjunction of positive and negative instances of basic formulas that agree on $\mathcal{X}$ and $\mathcal{Y}$.

$$\epsilon_L(\bigvee \psi_i) = \min \epsilon_L(\psi_i).$$

Theorem 4.3.4. *For a formula ϕ as above*

$$\operatorname{vc}\phi \geq \left\lfloor \frac{Y(\phi)}{\epsilon_L(\phi)} \right\rfloor$$

where $Y(\phi)$ is $Y(\psi)$ for ψ one the basic components of ϕ (all basic componenets agree on $\mathcal{Y}$).

Proof. First work with a formula that is a conjunction of positive basic formulas

$$\psi = \bigwedge_{j \leq J} \phi_j.$$

Then as we defined above

$$\epsilon_L(\psi) = \sum \epsilon_L(\phi_j)$$

Let ϕ be one of the basic formulas in ψ with a chain $\langle M_i \rangle_{i \leq k}$. Let $K_\phi = |M_k|$ i.e. the size of the extension. Let K be the largest such size among all ϕ_i.

Let n be the integer such that $n\epsilon_L(\psi) < Y$ and $(n+1)\epsilon_L(\psi) > Y$.

Label $\mathcal{Y}$ by an tuple b.

Pick parameter set $A \subset \mathscr{M}$ such that

$$A = \bigcup_{i<N} b_i$$

a disjoint union where each b_i is an ordered tuple of size $|x|$ connected according to ψ. We also require A to be $N \cdot I \cdot K$-strong.

Fix n arbitrary elements out of A, label them a_j.

For each ϕ_i, a_j pick an abstract realization M_{ij} of ϕ_i over (a_j, b) (abstract meaning disjoint from $\mathscr{M}$).

Let $\bar{M}$ be an abstract disjoint union of all those realizations.

Claim 4.3.5. *$(A \cap \bar{M}) \leq \bar{M}$.*

Proof. Consider some $(A \cap \bar{M}) \subseteq B \subseteq \bar{M}$. Let $B_{ij} = B \cap M_{ij} \subseteq M_{ij}$. Then B_{ij}'s are disjoint over $\bar{A} = A \cup b$. In particular $\dim B/(\bar{A} \cap B) = \sum \dim B_{ij}/(\bar{A} \cap B_{ij})$. $\dim B_{ij}/\bar{A} \geq -\epsilon_L(\phi_i)$ by Lemma 4.2.6. Thus $\dim B/(\bar{A} \cap B) \geq -n\epsilon(\psi)$. Thus $\dim B/(A \cap B) \geq \dim(B) - n\epsilon(\psi)$. By construction we have $Y - n\epsilon_L(\psi) > 0$ as needed. □

$|\bar{M}| \leq N \cdot I \cdot K$ and A is $\leq N \cdot I \cdot K$-strong. Thus a copy of $\bar{M}$ can be embedded over A into our ambient model $\mathscr{M}$. Our choice of b_i's was arbitrary, so we get $\binom{N}{n}$ choices out of $N|x|$ many elements. Thus we have $O(|A|^n)$ many traces.

Lemma 4.3.6. *There are arbitrarily large sets with properties of A.*

Proof. A is positive, as each of its disjoint components is positive. Thus $0 \leq A$. We apply proposition 4.4 in Laskoswki paper, embedding A into our structure $\mathscr{M}$ while avoiding all nonpositive extensions of size at most $N \cdot I \cdot K$. □

This shows

$$\mathrm{vc}\,\psi \geq n = \left\lfloor \frac{Y}{\epsilon_L} \right\rfloor$$

Now consider the formula which is a conjunction consists of negative basic formulas

$$\psi = \bigwedge \neg\phi_i$$

Let

$$\bar{\psi} = \bigwedge \phi_i$$

Do the construction above for $\bar{\psi}$ and suppose its trace is $X \subset A$ for some b. Then over b the same construction gives trace $(A - X)$ for ψ. Thus we get as many traces.

Finally consider a formula which is a disjunction of formulas considered above. Choose the one with the smallest ϵ_L, this yields the lower bound for the entire formula. $\square$

Claim 4.3.7 (4.1 in [1])**.** *We can find a minimal extension M with arbitrarily small dimension.*

Corollary 4.3.8. *This shows that the vc-function is infinite in Shelah-Spencer random graphs.*

$$\mathrm{vc}(n) = \infty$$

In particular, this implies that Shelah-Spencer graphs are not dp-minimal.

4.4 Upper bound

We bound the number of types of some finite collection of formulas $\Psi(\vec{x}, \vec{y}) = \{\phi_i(\vec{x}, \vec{y})\}_{i \in I}$ over a parameter set B of size N, where ϕ_i is a basic formula.

Fix a formula ϕ from our collection. Suppose it defines a minimal chain extension over $\{x, y\}$. Record the size of that extension as $K(\phi)$ and its total dimension $\epsilon(\phi) = \epsilon_U(\phi)$. Define dimension of that formula $D(\phi) = |\vec{y}| \frac{K(\phi)}{\epsilon(\phi)}$ Define dimension of the entire collection as $D(\Psi) = \max_{i \in I} D(\phi_i)$

In general we have parameter set $B \subset \mathscr{M}^{|y|}$, however without loss of generality we may work with a parameter set $B^{|y|}$, with $B \subset \mathscr{M}$.

Let $S = \lfloor D(\Psi) \rfloor$.

For our proof to work we also need B to be S-strong. We can achieve this by taking (the unique) S-strong closure of B. If size of B is N then the size of its closure is $O(N)$. So without loss of generality we can assume that B is S-strong.

Definition 4.4.1. A witness of a is a union of realizations of the existential formulas $\phi_i(a, b)$ for all i, b so that the formula holds.

Definition 4.4.2. For sets C, B define the boundary of C over B

$$\partial(C, B) = \{b \in B \mid \text{there is an edge between } b \text{ and element of } C - B\}$$

Definition 4.4.3. For each a pick some $\bar{M}_a$ to be its witness. Define two quantities

- ∂_a is the boundary $\partial(\bar{M}_a, B \cup a)$
- Suppose G_1, G_2 are some subgraphs of our model and $a_1 \subset G_1, a_2 \subset G_2$ finite tuples of vertices. Call $f\colon (G_1, a_1) \longrightarrow (G_2, a_2)$ a ∂-isomorphism if it is a graph isomorphism, f and f^{-1} are constant on B, and $f(a_1) = a_2$.
- Define $\mathscr{I}_a$ as the ∂-isomorphism class of $(\bar{M}_a, a)$.

Lemma 4.4.4. *If $\mathscr{I}_{a_1} = \mathscr{I}_{a_2}$ then a_1, a_2 have the same Ψ-type over B.*

Proof. Fix a ∂-isomorphism $f\colon (\bar{M}_{a_1}, a_1) \longrightarrow (\bar{M}_{a_1}, a_2)$. Suppose we have $\phi(a_1, b)$ for some $b \in B$. Pick witness of this existential formula $M_1 \subset \bar{M}_{a_1}$. Then $f(M_1)$ is a witness for $\phi(a_2, b)$. $\square$

Thus to bound the number of traces it is sufficient to bound the number of possibilities for $\mathscr{I}_a$.

Theorem 4.4.5.

$$|\partial_a| \leq D(\Psi)$$
$$|\bar{M}_b - \bar{A}| \leq D(\Psi)$$

Corollary 4.4.6.

$$\mathrm{vc}(\phi) \leq K(\phi)\frac{Y(\phi)}{\epsilon(\phi)}$$

Proof. We count possible ∂-isomorphism classes $\mathscr{I}_b$. Let $W = K(\phi)\frac{Y(\phi)}{\epsilon(\phi)}$. If the parameter set A is of size N then there are $\binom{N}{W}$ choices for boundary ∂_b. On top of the boundary there are at most W extra vertices and $(2W)^2$ extra edges. Thus there are at most

$$W \cdot 2^{(2W)^2}$$

configurations up to a graph isomorphism. In total this gives us

$$\binom{N}{W} \cdot W \cdot 2^{(2W)^2} = O(N^W)$$

options for ∂-isomorphism classes. By Lemma 4.4.4 there are at most $O(N^W)$ many traces, giving the required bound. $\square$

Proof. (of Theorem 4.4.5) Fix some b-trace A_b. Enumerate $A_b = \{a_1, \ldots, a_I\}$.

Let $M_i/\{a_i, b\}$ be a witness of $\phi(a_i, b)$ for each $i \leq I$. Let $\bar{M}_i = \bigcup_{j<i} M_j$. Let $\bar{M} = \bigcup M_i$, a witness of A_b

Claim 4.4.7.

$$\left|\partial(M_iM,\bar{A})-\partial(M,\bar{A})\right|\leq |M_i|=K(\phi)$$
$$\dim(M_iM\bar{A}/M\bar{A})>-\epsilon(\phi)$$

Definition 4.4.8. $(j-1,j)$ is called a <u>jump</u> if some of the following conditions happen

- New vertices are added outside of $\bar{A}$ i.e.

$$\bar{M}_j-\bar{A}\neq \bar{M}_{j-1}-\bar{A}$$

- New vertices are added to the boundary, i.e.

$$\partial(\bar{M}_j,\bar{A})\neq\partial(\bar{M}_{j-1},\bar{A})$$

Definition 4.4.9. We now let m_i count all jumps below i

$$m_i=|\{j<i \mid (j-1,j) \text{ is a jump}\}|$$

Lemma 4.4.10.

$$\dim(\bar{M}_i/\bar{A})\leq -m_i\cdot\epsilon(\phi)$$
$$|\partial(\bar{M}_i,\bar{A})|\leq m_i\cdot K(\phi)$$
$$|\bar{M}_j-\bar{A}|\leq m_i\cdot K(\phi)$$

Proof. (of Lemma 4.4.10) Proceed by induction. Second and third propositions are clear. For the first proposition base case is clear.

Induction step. Suppose $\bar{M}_j\cap(A\cup b)=\bar{M}_{j+1}$ and $\partial(\bar{M}_j,A)=\partial(\bar{M}_{j+1},A)$.

Then $m_i = m_{i+1}$ and the quantities don't change. Thus assume at least one of these equalities fails.

Apply Lemma 4.2.5 to $\bar{M}_j \cup (A \cup b)$ and $(M_{j+1}, a_{j+1}b)$. There are two options

- $\dim(\bar{M}_{j+1} \cup (A \cup b)/\bar{M}_i \cup (A \cup b)) \leq -\epsilon_U$. This implies the proposition.
- $M_{j+1} \subset \bar{M}_j \cup (A \cup b)$. Then by our assumption it has to be $\partial(\bar{M}_j, A) \neq \partial(\bar{M}_{j+1}, A)$. There are edges between $M_{j+1} \cap (\partial(\bar{M}_{j+1}, A) - \partial(\bar{M}_j, A))$ so they contribute some negative dimension $\leq \epsilon_U$.

This ends the proof for Lemma 4.4.10. $\square$

(Proof of Theorem 4.4.5 continued) First part of lemma 4.4.10 implies that we have $\dim(\bar{M}/\bar{A}) \leq -m_I \cdot \epsilon(\phi)$. The requirement of A to be S-strong forces

$$\begin{aligned} m_I \cdot \epsilon(\phi) &< Y(\phi) \\ m_I &< \frac{Y(\phi)}{\epsilon(\phi)} \end{aligned}$$

Applying the rest of 4.4.10 gives us

$$\begin{aligned} |\partial(\bar{M}, A)| &\leq m_I \cdot K(\phi) \leq \frac{K(\phi)Y(\phi)}{\epsilon(\phi)} \\ |\bar{M} \cap A| &\leq m_I \cdot K(\phi) \leq \frac{K(\phi)Y(\phi)}{\epsilon(\phi)} \end{aligned}$$

as needed. This ends the proof for Theorem 4.4.5. $\square$

So far we have computed an upper bound for a single basic formula ϕ.

To bound an arbitrary formula, write it as a boolean combination of basic formulas ϕ_i (via quantifier elimination) It suffices to bound vc-density for collection of formulas $\{\phi_i\}$ to obtain a bound for the original formula.

In general work with a collection of basic formulas $\{\phi_i\}_{i\in I}$. The proof generalizes in a straightforward manner. Instead of $A^{|x|}$ we now work with $A^{|x|}\times I$ separating traces of different formulas. Formula with the largest quantity $Y(\phi)\frac{K(\phi)}{\epsilon(\phi)}$ contributes the most to the vc-density. Thus we have

$$\Phi = \{\phi_i\}_{i\in I}$$
$$\mathrm{vc}(\Phi) \leq \max_{i\in I} Y(\phi_i)\frac{K(\phi_i)}{\epsilon_{\phi_i}}$$

Chapter 5

Flat graphs

We show that the theory of superflat graphs is dp-minimal.

5.1 Preliminaries

Superflat graphs were introduced in [10] as a natural class of stable graphs. Here we present a direct proof showing dp-minimality.

First, we introduce some basic graph-theoretic definitions.

Definition 5.1.1. Work in an infinite graph G. Let $A, V \subset V(G)$ (where $V(G)$ denotes vertices of G)

1. $G' = G[V]$ is called *induced* subgraph of G *spanned* by V if it is obtained as a subgraph of G by taking all edges between vertices in V.

2. For $a, b \in V(G)$ define the *distance* $d(a, b)$ to be the length of the shortest path between a and b in G.

3. For $a, b \in V(G)$ define $d_A(a, b)$ to be the distance between a and b in $G[V(G) - A]$. Equivalently it is the shortest path between a and b that avoids vertices in A.

4. We say that A *separates* V if for all $a, b \in V$, $d_A(a, b) = \infty$.

5. We say that V has *connectivity* n if there are no sets of size $n-1$ in $V(G)$ that separates V.

6. Suppose V has finite connectivity n. *Connectivity hull* of V is defined to be the union of all sets separating V of size $n-1$.

In [11] we find a generalization of Megner's Theorem for infinite graphs

Theorem 5.1.2 (Megner, Erdos, Aharoni, Berger)**.** *Let A and B be two sets of vertices in a possibly infinite digraph. Then there exist a set P of disjoint AB paths, and a set S of vertices separating A from B, such that S consists of a choice of precisely one vertex from each path in P.*

We use the following easy consequence

Corollary 5.1.3. *Let V be a subset of a graph G with connectivity n. Then there exists a set of n disjoint paths from V into itself.*

Corollary 5.1.4. *With assumptions as above, connectivity hull of V is finite.*

Proof. All the separating sets have to have exactly one vertex in each of those paths. □

Definition 5.1.5. Denote by K_n^m an m-subdivision of the complete graph on n vertices. Graph is called superflat if for every $m \in \mathbb{N}$ there is $n \in \mathbb{N}$ such that the graph avoids K_n^m as a subgraph.

Theorem 2 in [10] gives a useful characterization of superflat graphs.

Theorem 5.1.6. *The following are equivalent*

1. *G is superflat*

2. *For every $n \in \mathbb{N}$ and an infinite set $A \subset V(G)$, there exists a finite $B \subset V(G)$ and infinite $A' \subseteq A$ such that for all $x, y \in A'$ we have $d_B(x, y) > n$.*

Roughly, in superflat graphs every infinite set contains a sparse infinite subset (possibly after throwing away finitely many nodes).

5.2 Indiscernible sequences

In this section we work in a superflat graph G. Stability implies that all the indiscernible sequences are totally indiscernible.

Definition 5.2.1. Let $V \subset V(G)$. $P_n(V)$, a subgraph of G denotes a union of all paths of length $\leq n$ between points of V.

Lemma 5.2.2. *Let $I = (a_i : i \in \mathcal{I})$ be a countable indiscernible sequence over A. Fix $n \in \mathbb{N}$. There exists a finite set B such that*

$$\forall i \neq j \ d_B(a_i, a_j) > n$$

Proof. By a 5.1.6 we can find an infinite $\mathcal{J} \subset \mathcal{I}$ and a finite set B' such that each pair from $J = (a_j : j \in \mathcal{J})$ have distance $> n$ over B'. Using total indiscernibility we have an automorphism sending J to I fixing A. Image of B' under this automorphism is the required set B. □

In other words, B separates I when viewed inside subgraph $P_n(I)$. This shows that I has finite connectivity in $P_n(I)$. Applying Corollary 5.1.4 we obtain that connectivity hull of I in $P_n(I)$ is finite.

Definition 5.2.3. We call a set $H \subseteq V(G)$ uniformly definable from an indiscernible sequence I if there is a formula $\phi(x, y)$ such that for every $J \subset I$ of size $|y|$ we have

$$H = \{g \in G \mid \phi(g, J)\}$$

where J is considered a tuple.

Lemma 5.2.4. *Fix a countable indiscernible sequence $I = (a_i : i \in \mathcal{I})$. Let H be its connectivity hull inside of graph $P_n(I)$. Then H is uniformly definable from I in G.*

Definition 5.2.5. Given a graph G and $V \subset V(G)$ define $H(G, V)$ to be connectivity hull of V in G.

Note 5.2.6. Given a finite V we have $H(P_n(V), V)$ is V-definable.

Proof. Corollary 5.1.3 tells us that there finitely many paths between elements of V such that $H(P_n(I), I)$ is inside those paths. Take $J \subset I$ be the endpoints of those paths. $H(P_n(J), J)$ is J-definable as noted above. Moreover we have $H(P_n(I), I) \subseteq H(P_n(J), J)$, both finite sets. In particular we have $H \subset \operatorname{acl}(J)$. $I - J$ is indiscernible over $A \cup J$.

□

Proof. (of 5.2.4) Consider finite parts of the sequence $I_i = \{a_1, a_2, \ldots, a_i\}$. Define $H_i = H(P_n(I_i), I_i)$. It is I_i-definable. Corollary 5.1.3 tells us that there finitely many paths between elements of V such that $H(P_n(I), I)$ is inside those paths. But for large enough i, say $i \geq N$, $P_n(I_i)$ will contain all of those paths. Thus for $i \geq N$ we have $H(P_n(I), I) \subseteq P_n(I_i)$. If a set separates I then would be inside $P_n(I_i)$ and would separate I_i as well. Thus for $i \geq N$ we have $H(P_n(I), I) \subseteq H(P_n(I_i), I_i)$. If the two sets are not equal, it is due to some elements in $H(P_n(I_i), I_i)$ failing to separate entire I. There are finitely many of those, so for large enough i, say $i \geq M$ we have $H(P_n(I_i), I_i) = H(P_n(I), I)$ stabilizing. This shows that for $i \geq M$ we have $H_i = H_{i+m} = H(P_n(I), I)$. By symmetry of indiscernible sequence we have that any subset of size N defines the connectivity hull. □

Lemma 5.2.7. *I is indiscernible over the $A \cup H(P_n(I), I)$.*

Proof. Denote the hull by H. Fix an A-formula $\phi(x, y)$. Consider a collection of traces $\phi(a, H^{\{}|y|\})$ for $a \in I^{|x|}$. If two of them are distinct, then by indiscernibility all of them are. But that is impossible as H has finitely many subsets. Thus all the traces are identical. This shows that for any $a, b \in I^{|x|}$ and $h \in H^{\{}|y|\}$ we have $\phi(a, h) \iff \phi(b, h)$. As choice of ϕ was arbitrary, this shows that I is indiscernible over $A \cup H(P_n(I), I)$. $\square$

Corollary 5.2.8. *Let $(a_i)_{i \in I}$ be a countable indiscernible sequence over A. Then there is a countable B such that (a_i) is indiscernible over $A \cup B$ and*

$$\forall i \neq j \ d_B(a_i, a_j) = \infty$$

Proof. Let $B_n = H(P_n(I), I)$. This is well defined by Lemma 5.2.2 and has property

$$\forall i \neq j \ d_{B_n}(a_i, a_j) > n$$

and I is indiscernible over $A \cup B_i$ by Corollary 5.2.7. Let $B = \bigcup_{n \in \mathbb{N}} B_n$. $\square$

That is every indiscernible sequence can be upgraded to have infinite distance over its parameter set.

5.3 Superflat graphs are dp-minimal

Lemma 5.3.1. *Suppose $x \equiv_A y$ and $d_A(x, c) = d_A(y, c) = \infty$. Then $x \equiv_{Ac} y$*

Proof. Define an equivalence relation $G - A$. Two points p, q are equivalent if $d_A(p, q)$ is finite. There is an automorphism f of G fixing A sending x to y. Denote by X and Y equivalence classes of x and y respectively. It's easy to see that $f(X) = Y$. Define the following function

$$g = f \text{ on } X$$

$$g = f^{-1} \text{ on } Y$$

$$\text{identity otherwise}$$

It is easy to see that g is an automorphism fixing Ac that sends x to y. $\square$

Theorem 5.3.2. *Let G be a flat graph with $(a_i)_{i \in \mathbb{Q}}$ indiscernible over A and $b \in G$. There exists $c \in \mathbb{Q}$ such that all $(a_i)_{i \in \{\mathbb{Q} - c\}}$ have the same type over Ab.*

Proof. Find $B \supseteq A$ such that (a_i) is indiscernible over B and has infinite distance over B. All the elements of the indiscernible sequence fall into distinct equivalence classes. b can be in at most one of them. Exclude that element from the sequence. Remaining sequence elements are all infinitely far away from b. By previous lemma we have that elements of indiscernible sequence all have the same type over Bb. $\square$

But this is exactly what it means to be dp-minimal, as given, say, in [8] Lemma 1.4.4

Corollary 5.3.3. *Flat graphs are dp-minimal.*

Bibliography

[1] Michael C. Laskowski, *A simpler axiomatization of the Shelah-Spencer almost sure theories*, Israel J. Math. **161** (2007), 157186. MR MR2350161

[2] M. Aschenbrenner, A. Dolich, D. Haskell, D. Macpherson, S. Starchenko, *Vapnik-Chervonenkis density in some theories without the independence property*, I, Trans. Amer. Math. Soc. 368 (2016), 5889-5949

[3] E. Leenknegt. *Reducts of p-adically closed fields*, Archive for Mathematical Logic, 53(3):285-306, 2014

[4] P. Assouad, *Densite et dimension*, Ann. Inst. Fourier (Grenoble) 33 (1983), no. 3, 233282.

[5] P. Assouad, *Observations sur les classes de Vapnik-Cervonenkis et la dimension combinatoire de Blei*, in: Seminaire dAnalyse Harmonique, 19831984, pp. 92112, Publications Mathematiques dOrsay, vol. 85-2, Universite de Paris-Sud, Departement de Mathematiques, Orsay, 1985.

[6] N. Sauer, *On the density of families of sets*, J. Combinatorial Theory Ser. A 13 (1972), 145147.

[7] S. Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific J. Math. 41 (1972), 247261.

[8] P. Simon, *On dp-minimal ordered structures*, J. Symbolic Logic 76 (2011), no. 2, 448-460

[9] Michel Parigot. Théories d'arbres. *Journal of Symbolic Logic*, 47, 1982, 841-853

[10] Klaus-Peter Podewski and Martin Ziegler. Stable graphs. *Fund. Math.*, 100:101-107, 1978.

[11] Aharoni, Ron and Berger, Eli (2009). "Menger's Theorem for infinite graphs". *Inventiones Mathematicae* 176: 162