VC-DENSITY IN AN ADDITIVE REDUCT OF P-ADIC NUMBERS

ANTON BOBKOV

ABSTRACT. Aschenbrenner et. al. computed a bound $\mathrm{vc}(n) \leq 2n-1$ for the VC density function in the field of p-adic numbers, but it is not known to be optimal. I 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 $\mathrm{vc}(n) = n$ for that structure.

VC density was introduced into model theory in [1] by Aschenbrenner, Dolich, Haskell, MacPherson, and Starchenko as a natural notion of dimension for definable families of sets in NIP theories. In a NIP theory T we can define the vc-function

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

where $\mathrm{vc}(n)$ measures the worst-case complexity of families of definable sets in an n-dimensional space. The simplest possible behavior is $\mathrm{vc}(n) = n$ for all n. For $T = \mathrm{Th}(\mathbb{Q}_p)$, the paper [1] computes an upper bound for this function to be $2n-1$, and it is not known whether it is optimal. This same bound would hold in any reduct of the field of p-adic numbers, so one may expect that the simplified structure of the reduct would allow a better bound. In [2], Leenknegt provides a cell decomposition result for a certain P-minimal additive reduct of the field 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. More in depth exposition of vc-density can be found in [1]. Section 2 defines and states some basic facts about thoery of p-adic numbers. Here we also introduce the reduct we will be working with. Section 3 sets up basic definition and lemmas that will be needed for the proof. We define trees and intervals and show how it helps 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 $\mathbb{Q}_p$-vector space $\mathbb{Q}_p^n$. $|x|$ refers to the arity of the variable. First-order formulas will have parameter variables separated $\phi(x; y)$.

1. VC-DIMENSION AND VC-DENSITY

Definition 1.1. Throughout this section we work with a collection $\mathcal{F}$ of subsets of a set X. We call the pair $(X, \mathcal{F})$ a <u>set system</u>.

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

Definition 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 $\mathrm{VC}(\mathcal{F})$.

Note 1.3. We may drop X from the previous definition, as it VC-dimension doesn't depend on the base set and is determined by $(\bigcup\mathcal{F},\mathcal{F})$.

This allows us to distinguish between well behaved set systems of finite VC-dimension which tend to have good combinatorial properties and poorly behaved set systems with infinite VC dimension.

Another natural combinatorial notion is that of a dual system:

Definition 1.4. For $a\in X$ define $X_a=\{F\in\mathcal{F}\mid a\in F\}$. Let $\mathcal{F}^*=\{X_a\}_{a\in X}$. We define $(\mathcal{F},\mathcal{F}^*)$ as 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 1.5. *A set system has finite VC-dimension if and only if its dual system has finite VC-dimension. More precisely*

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

For a more refined notion we look at the traces of our family on finite sets:

Definition 1.6. Define the shatter function $\pi_{\mathcal{F}}\colon\mathbb{N}\longrightarrow\mathbb{N}$ and the dual shatter function $\pi^*_{\mathcal{F}}\colon\mathbb{N}\longrightarrow\mathbb{N}$ of $\mathcal{F}$ by

$\pi_{\mathcal{F}}(n)=\max\{|A\cap\mathcal{F}|\mid A\subset X \text{ and } |A|=n\}$

$\pi^*_{\mathcal{F}}(n)=\max\{\text{number of atoms in Boolean algebra generated by } B\mid B\subset\mathcal{F},|B|=n\}$

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

Theorem 1.7 (Sauer-Shelah '72)**.** *If the set system $(X,\mathcal{F})$ has finite VC-dimension d then $\pi_{\mathcal{F}}(n)\le\binom{n}{\le d}$ where $\binom{n}{\le d}=\binom{n}{d}+\binom{n}{d-1}+\ldots+\binom{n}{1}$.*

Thus the systems with a finite VC-dimension are precisely the systems where the shatter function grows polynomially. Define vc-density to be the degree of that polynomial:

Definition 1.8. Define vc-density and dual vc-density of $\mathcal{F}$ as

$$\mathrm{vc}(\mathcal{F})=\limsup_{n\to\infty}\frac{\log\pi_{\mathcal{F}}(n)}{\log n}\in\mathbb{R}^{\ge 0}\cup\{+\infty\}$$
$$\mathrm{vc}^*(\mathcal{F})=\limsup_{n\to\infty}\frac{\log\pi^*_{\mathcal{F}}(n)}{\log n}\in\mathbb{R}^{\ge 0}\cup\{+\infty\}$$

Generally speaking a shatter function that is bounded by a polynomial doesn't itself have to be a polynomial. Proposition 4.12 in [1] 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 1.9. Work in a structure M. Fix a finite collection of formulas $\Phi(x,y) = \{\phi_i(x,y)\}$.

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

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 1.10.

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

Lemma 1.11.

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

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.

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

Definition 1.12. Define ϕ to be NIP if it has finite VC-dimension.

[**?**] shows that in a general combinatorial context, vc-density can be any real number in $0 \cup [1, \infty)$. Less is known if we restrict our attention to NIP theories. Proposition 4.6 in [1] 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 context.

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

Definition 1.13. For a given NIP structure M, define the vc-function

$$\begin{aligned} \mathrm{vc}^M(n) &= \sup\{\mathrm{vc}^*(\phi(x,y)) \mid \phi \in \mathcal{L}(M), |x| = n\} \\ &= \sup\{\mathrm{vc}(\phi(x,y)) \mid \phi \in \mathcal{L}(M), |y| = n\} \end{aligned}$$

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 1.14 (Lemma 3.22 in [1])**.**

$$\begin{aligned} \mathrm{vc}(1) &\geq 1 \\ \mathrm{vc}(n) &\geq n\,\mathrm{vc}(1) \end{aligned}$$

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

2. P-adic numbers

The field 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 of fields together with unary predicates P_n interpreted in $\mathbb{Q}_p$ by

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

and a divisibility relation where $a|b$ holds when $\operatorname{val} a \le \operatorname{val} b$.

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

Theorem 2.1 (Macintyre '76)**.** *The $\mathcal{L}_{Mac}$-structure $\mathbb{Q}_p$ has quantifier elimination.*

There is also a cell decomposition result:

Definition 2.2. Define k-cell recursively. 0-cells are points in $\mathbb{Q}_p$. An $(k+1)$-cell is a subset of $\mathbb{Q}_p^{k+1}$ of the following form:

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

where D is an k-cell, $a_1(x), a_2(x), c(x)$ are $\emptyset$-definable, $\square$ is $<, \le$ or no condition, and $\lambda \in \mathbb{Q}_p$.

Theorem 2.3 (Denef '84)**.** *Any subset of $\mathbb{Q}_p$ defined by a $\mathcal{L}_{Mac}$-formula $\phi(x,t)$ with $|x| = n$ and $|t| = 1$ decomposes into a finite union of $(k+1)$-cells.*

In [1], Aschenbrenner, Dolich, Haskell, Macpherson, and Starchenko show that this structure has $\mathrm{vc}(n) \le 2n - 1$, however it is not known whether this bound is optimal.

In [2], Leenknegt analyzes the reduct of p-adic numbers 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 $\bar{c}$ is a scalar multiplication by c, $a|b$ stands for $\operatorname{val} a \le \operatorname{val} b$, and $Q_{m,n}$ is a unary predicate

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

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 the extra relation symbols are definable in the $\mathcal{L}_{Mac}$-structure $\mathbb{Q}_p$. The paper [2] provides a cell decomposition result with the following cells:

Definition 2.4. A 0-cell is a point in $\mathbb{Q}_p$. An $(k+1)$-cell is a subset of $\mathbb{Q}_p^{k+1}$ of the following form:

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

where D is an k-cell called the base of the cell, $a_1(x), a_2(x), c(x)$ are degree ≤ 1 polynomials, $\square$ is $<$ or no condition, and $\lambda \in \mathbb{Q}_p$.

Theorem 2.5 (Leenknegt '12)**.** *Any formula $\phi(x,t)$ in $(\mathbb{Q}_p, \mathcal{L}_{aff})$ with $|x| = n$ and $|t| = 1$ decomposes into a union of $(k+1)$-cells.*

Moreover, [2] shows that $(\mathbb{Q}_p, \mathcal{L}_{aff})$ is a P-minimal reduct, that is the one-dimensional definable sets of $(\mathbb{Q}_p, \mathcal{L}_{aff})$ coincide with the one-dimensional definable sets in the full structure $(\mathbb{Q}_p, \mathcal{L}_{Mac})$.

I am able to compute the vc-function for this structure:

Theorem 2.6. *$(\mathbb{Q}_p, \mathcal{L}_{aff})$ has $\mathrm{vc}(n) = n$.*

3. Key Lemmas and Definitions

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

Lemma 3.1. *Any formula $\phi(x; y)$ in $(\mathbb{Q}_p, \mathcal{L}_{aff})$ can be written as a boolean combination of formulas from the following collection*

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

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, and $\lambda_k \in \mathbb{Q}_p$.

Proof. Let $l = |x| + |y|$. Apply the cell decomposition theorem to $\phi(x; y)$ to obtain $\mathscr{D}^l$, a collection of l-cells. Let $\mathscr{D}^{l-1}$ be a collection $l - 1$ of bases of cells in $\mathscr{D}^l$. Similarly, construct by induction $\mathscr{D}^i$ for each $0 \le j < l$, where $\mathscr{D}_j$ is a collection of j-cells which are the bases of cells in $\mathscr{D}_{j+1}$. Let $\mathscr{D} = \bigcup \mathscr{D}_j$. Choose m, n large enough to cover all n', m' for $Q_{n',m'}$ that show up in the cells of $\mathscr{D}$. Choose λ_k to go over all the cosets of $Q_{m,n}$. Let $q_i(x, y)$ enumerate all of the polynomials $a_1(x), a_2(x), t - c(x)$ that show up in the cells of $\mathscr{D}$. Those are all polynomials of degree ≤ 1 in variables x, y. We can split each of them as $q_i(x, y) = p_i(x) - c_j(y)$ where the constant term goes into c_j. 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. $\square$

Lemma 3.2. *If ϕ can be written as a boolean combination of formulas from Φ then*

$$\mathrm{vc}^*(\Phi) \le n \implies \mathrm{vc}^*(\phi) \le n$$

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 1.11. $\square$

Therefore to show that $\mathrm{vc}^*(\phi) \le |x|$, it suffices to bound $\mathrm{vc}^*(\Phi) \le |x|$. 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 a set $T = T(\Psi, B) = \{c_i(b) \mid b \in B, i \in I\} \subset \mathbb{Q}_p$. 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 3.3.

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

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

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

- Define a collection of balls $\mathscr{B} = \{B(t_1, \mathrm{val}(t_1 - t_2))\}_{t_1,t_2 \in T}$. Those form a (directed) boolean algebra of sets in $\mathbb{Q}_p$. We refer to the atoms in that algebra as intervals.
- Let's introduce some notation for the intervals. For $t \in T$ and $\alpha_L . \alpha_U \in \mathbb{Z} \cup \{-\infty, +\infty\}$ define
$$\mathrm{I}(t, \alpha_L, \alpha_U) = B(t, \alpha_L) \backslash \bigcup \{B(t', \alpha_U) \mid t' \in T, \mathrm{val}(t' - t) \geq \alpha_U\}$$
(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.
- Intervals are a natural construction for trees, however we will require a more refined notion to make Lemma 3.12 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 defintion, we define subinterval as an atom of a 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.

Subintervals are fine enough to make Lemma 3.12 work while coarse enough to be $O(N)$ small.

Lemma 3.4.
- *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 interval to the total count, so by induction there are at most $|\mathscr{B}' \backslash \mathscr{B}|$ more subintervals than there are intervals. □

Definition 3.5. Suppose $a \in \mathbb{Q}_p$ lies in an interval $\mathrm{I}(t, \alpha_L, \alpha_U)$. Define T-valuation of a to be $\text{T-val}(a) = \mathrm{val}(a - t)$.

This a natural notion having the following properties:

Lemma 3.6.
(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 a subinterval* $\mathrm{I}(t, \alpha_L, \alpha_U)$ *(as opposed to an interval), then* $\text{T-val}(a) = \mathrm{val}(a - t)$ *as well (this works for any refinement of intervals).*
(c) *If* $a \in \mathbb{Q}_p$ *lies in a (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 a (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 a subinterval $\mathrm{I}(t, \alpha'_L, \alpha'_U)$. This subinterval lies inside of an 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 intervlas one of the three following cases has to hold.

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

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

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 $\text{T-val}(a) = \alpha_U$.

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

thus $\mathrm{val}(a - t') \geq \alpha_U$. The interval $\mathrm{I}(t, \alpha_L, \alpha_U)$ excludes 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: $\mathrm{val}(t - t') \leq \alpha_L$

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

thus $\mathrm{val}(a - t') = \mathrm{val}(t - t')$ and note that $\mathrm{val}(a - t') \leq \text{T-val}(a)$ as needed. □

Definition 3.7. Suppose $a \in \mathbb{Q}_p$ lies in a subinterval $\mathrm{I}(t, \alpha_L, \alpha_U)$. We say that a is far from boundary if

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

Otherwise we say that it is close to boundary.

Definition 3.8. 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 boundary and $a_1 - t, a_2 - t$ are in the same $Q_{m,n}$ coset.
- Both a_1, a_2 are close to boundary and $\text{T-val}(a_1) = \text{T-val}(a_2) \leq \mathrm{val}(a_1 - a_2) - n$.

Definition 3.9. For $c \in \mathbb{Q}_p$ and $\alpha, \beta \in \mathbb{Z}$ define $c \upharpoonright [\alpha, \beta) \in (\mathbb{Z}/p\mathbb{Z})^{\beta - \alpha}$ to be the record of the coefficients of c for the valuations between $[\alpha, \beta)$. More precisely write c in its power series form

$$c = \sum_{\gamma \in \mathbb{Z}} c_\gamma p^\gamma \text{ with } c_\gamma \in \mathbb{Z}/p\mathbb{Z}$$

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

The following lemma is an adaptation of Lemma 7.4 in [1].

Lemma 3.10. *Fix $m, n \in \mathbb{N}$. For any $x, y, c \in \mathbb{Q}_p$, if*

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

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

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

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

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 $Q_{m,n}$. Conditions of the lemma force $x - c, y - c$ to be similar, thus belonging to the same coset. $\square$

Lemma 3.11. *For each subinterval there are at most $K = K(Q_{m,n})$ many subinterval types (with K not dependent on B 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 boundary. Then they have the same subinterval type if $a - t, a' - t$ are in the same $Q_{m,n}$-coset. Number of such subinterval types is bounded by the number of $Q_{m,n}$-cosets.

Suppose a, a' are close to boundary and

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

Then a, a' have the same subinterval type. Such subinterval type is thus determined by $\text{T-val}(a) - \alpha_L$ and $a \restriction [\text{T-val}(a), \text{T-val}(a) + n)$, 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 those 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 lemma relates tree notions to Φ-types.

Lemma 3.12. *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 3.1). 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 $i \in I, b \in B$ we have $p_i(d) - c_i(b), p_i(d') - c_i(b)$ are in the same $Q_{m,n}$-coset. Fix such i, b. For brievety 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 boundary. Then $\text{T-val}(a) = \text{T-val}(a') \leq \text{val}(a - a') - n$. Using Lemma 3.6d, 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 3.10 shows that $a - c_i(b), a' - c_i(b)$ are in the same Q-coset.

Now, suppose both a, a' are far from boundary. Label their interval as $\mathrm{I}(t, \alpha_L, \alpha_U)$. 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 $\mathrm{val}(t - c_i(b)) \geq \alpha_U$. Then using Lemma 3.6d

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

So by Lemma 3.10 we have $a - c_i(b), a - t$ are in the same Q-coset. By a parallel argument we have $a' - c_i(b), a' - t$ 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 $\mathrm{val}(t - c_i(b)) \leq \alpha_L$. Then using Lemma 3.6d

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

so by Lemma 3.10 we have $a - c_i(b), t - c_i(b)$ are in the same Q-coset. By a parallel argument we have $a' - c_i(b), t - c_i(b)$ 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 3.1). Fix $i, j \in I, b \in B$. We would like to show that

$$\mathrm{val}(p_i(d) - c_i(b)) < \mathrm{val}(p_j(d) - c_j(b)) \iff \mathrm{val}(p_i(d') - c_i(b)) < \mathrm{val}(p_j(d') - c_j(b)) \tag{3.1}$$

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 3.6d yields 4 following cases.

Case 1:

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

Then it is clear that the equivalence (3.1) holds.

Case 2:

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

Then the equivalence (3.1) holds by the third condition of the lemma that order of T-valuations is preserved.

Case 3:

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

If $p_j(d), p_j(d')$ are close to boundary, then $\text{T-val}(p_j(d)) = \text{T-val}(p_j(d'))$ and the equivalence (3.1) clearly holds. Suppose then that $p_j(d), p_j(d')$ are far from boundary.

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

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 (3.1) 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 3.6c, so we only need the far from boundary condition to avoid the edge case of equality to β_j.)

Case 4:

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

Similar to case 3 (switching i, j). □

Note 3.13. This 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, giving a total of $O(N^{|I|})\cdot K^{|I|}\cdot |I|! = O(N^{|I|})$ many types. This implies $\mathrm{vc}^*(\Phi)\le |I|$. 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 an subinterval for $|x|$ many p_i's, the subinterval for the rest should be determined. This would give the required $\mathrm{vc}^*(\Phi)\le |x|$ bound. The next section outlines this idea formally.

4. Main Proof

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

Lemma 4.1. *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 $J\subset I$ and $i\in I$ satisfy*

$$\vec{p}_i\in\mathrm{span}\,\{\vec{p}_j\}_{j\in J},$$

and we have $\vec{c}\in\mathbb{Q}_p^{|x|}, \alpha\in\mathbb{Z}$ with

$$\mathrm{val}(\vec{p}_j\cdot\vec{c})>\alpha \textit{ for all } j\in J$$

Then

$$\mathrm{val}(\vec{p}_i\cdot\vec{c})>\alpha-\gamma$$

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. Fix i, J satisfying the conditions of the lemma. For some $c_j\in\mathbb{Q}_p$ for $j\in J$ we have

$$\vec{p}_i=\sum_{j\in J}c_j\vec{p}_j,$$

hence

$$\vec{p}_i\cdot\vec{c}=\sum_{j\in J}c_j\vec{p}_j\cdot\vec{c}.$$

We have

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

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

$$\begin{aligned}
&\operatorname{val}(c_j \vec{p}_j \cdot \vec{c}) > \alpha - \gamma \text{ for all } j \in J \\
&\operatorname{val}\left(\sum_{j \in J} c_j \vec{p}_j \cdot \vec{c}\right) > \alpha - \gamma \\
&\operatorname{val}(\vec{p}_i \cdot \vec{c}) > \alpha - \gamma
\end{aligned}$$

This 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 Lemma 4.1 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 4.2. Suppose $a \in \mathbb{Q}_p$ lies in a subinterval $(B(t_L, \alpha_L), B(t_U, \alpha_U))$. Define floor of a to be $F(a) = \alpha_L$.

Definition 4.3. 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.

Given a tuple $(a_i)_{i \in I}$ in the segment space, look at the corresponding floors $\{F(a_i)\}_{i \in I}$ and T-valuations $\{\text{T-val}(a_i)\}_{i \in I}$. Partition the segment space by the order types of $\{F(a_i)\}_{i \in I}$ and $\{\text{T-val}(a_i)\}_{i \in I}$ (as subsets of $\mathbb{Z}$).

Work in a fixed partition Sg'. After relabeling we may assume that

$$F(a_1) \geq F(a_2) \geq \dots$$

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 all vectors with indices in J are linearly independent, and all vectors with indices outside of J are a linear combination of preceding vectors. For any index $i \in I$ we call it independent if $i \in J$ and we call it dependent otherwise.

Definition 4.4.

- Denote $\mathbb{Z}/p\mathbb{Z}^\gamma$ as Ct. Note that $|\operatorname{Ct}| = p^\gamma$.
- Let It be the space of all subinterval types. By Lemma 3.11 $|\operatorname{It}| \leq K$.
- Let Sub be the space of all subintervals. By Lemma 3.4 $|\operatorname{Sub}| \leq 3|I|^2 \cdot N = O(N)$.

Definition 4.5. Now, we define the following function

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

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 It^I.

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

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

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

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

Lemma 4.6. *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 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

$$f(c) = (p_i(c))_{i \in I} = (\vec{p}_i \cdot \vec{c})_{i \in I} = (a_i)_{i \in I}$$
$$f(c') = (p_i(c'))_{i \in I} = (\vec{p}_i \cdot \vec{c}')_{i \in I} = (a_i')_{i \in I}$$

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 3.12 ($f(c), f(c')$ are in the same partition ensuring the proper order of T-valuations for the 3rd condition of the lemma). It 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 $F(a_i) \leq F(a_j)$ and $F(a_i') \leq F(a_j')$. Moreover $F(a_j) = F(a_j')$ as a_j, a_j' lie in the same subinterval (using the earlier part of the argument as j is independent).

Claim 4.7. $\mathrm{val}(a_i - a_i') > F(a_j) - \gamma$

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

$$\mathrm{val}(a_k - a_k') > F(a_k) \text{ for all } k \in K$$

as a_k, a_k' lie in the same subinterval (using the earlier part of the argument as k is independent).

$$\mathrm{val}(a_k - a_k') > F(a_j) \text{ for all } k \in K \text{ by monotonicity of } F(a_k)$$
$$\mathrm{val}(\vec{p}_k \cdot \vec{c} - \vec{p}_k \cdot \vec{c}') > F(a_j) \text{ for all } k \in K$$
$$\mathrm{val}(\vec{p}_k \cdot (\vec{c} - \vec{c}')) > F(a_j) \text{ for all } k \in K$$

$K \subset I, i \in I, \vec{c} - \vec{c}' \in \mathbb{Q}_p^{|x|}, F(a_j) \in \mathbb{Z}$ satisfy the requirements of Lemma 4.1, so we apply it to conclude

$$\mathrm{val}(\vec{p}_i \cdot (\vec{c} - \vec{c}')) > F(a_j) - \gamma$$
$$\mathrm{val}(\vec{p}_i \cdot \vec{c} - \vec{p}_i \cdot \vec{c}') > F(a_j) - \gamma$$
$$\mathrm{val}(a_i - a_i') > F(a_j) - \gamma$$

as needed, finishing the proof of the claim. □

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

$$\mathrm{val}(a_i - a_i') > F(a_j)$$

We now would like to show that a_i, a_i' lie in the same subinterval. As $F(a_i) \leq F(a_j)$, $F(a_i') \leq F(a_j')$ and $F(a_j) = F(a_j')$ we have that $\mathrm{val}(a_i - a_i') > F(a_i)$ and

$\mathrm{val}(a_i - a_i') > F(a_i')$ Suppose that a_i lies in the subinterval $\mathrm{I}(t, F(a_i), \alpha_U)$ and that a_i' lies in the subinterval $\mathrm{I}(t', F(a_i'), \alpha_U')$. Without loss of generality assume that $F(a_i) \leq F(a_i')$. As $\mathrm{val}(a_i - a_i') > F(a_i')$, this implies that

$$a_i \in B(a_i', F(a_i'))$$
$$a_i \in B(t', F(a_i'))$$
$$B(t, F(a_i)) \cap B(t', F(a_i')) \neq \emptyset$$
$$B(t, F(a_i)) \subset B(t', F(a_i'))$$

For the subintervals to be disjoint we need $F(a_i') \geq \alpha_U$ and $\mathrm{I}(t, F(a_i), \alpha_U) \cap B(t', F(a_i')) = \emptyset$. But $\mathrm{val}(t' - a_i) > F(a_i')$ implying that $a_i \in \mathrm{I}(t, F(a_i), \alpha_U) \cap B(t', F(a_i'))$ giving a contradiction. Therefore the subintervals coicide finishing the proof. □

Corollary 4.8. *$\Phi(x, y)$ has dual* vc-*density* $\leq |x|$.

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

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

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. □

Corollary 4.9 (Theorem 2.6)**.** *$(\mathbb{Q}_p, \mathcal{L}_{aff})$ has* $\mathrm{vc}(n) = n$.

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

This proof relies heavily on the linearity of functions a_1, a_2, c in the cell deomposition result (see Definition 2.4). Linearity is used to separate x and y variables as well as for Lemma 4.1 to reduce the number of independent factors from $|I|$ to $|x|$. The paper [2] has cell decomposition results for more expressive reducts of $\mathbb{Q}_p$, including, for exapmple, restricted multiplication. While our results don't apply to it directly, it is this author's hope that similar techniques can be used to compute $\mathrm{vc}(n)$ function for those structures.

References

[1] 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

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

E-mail address: `bobkov@math.ucla.edu`