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

ANTON BOBKOV

ABSTRACT. Aschenbrenner et. al. computed a bound $\mathrm{vc}(n) = 2n - 1$ 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 NIP theories. In an NIP theory we can define the vc-function

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

where $\mathrm{vc}(n)$ measures complexity of the definable sets in an n-dimensional space. The simplest possible behavior is $\mathrm{vc}(n) = n$ for all n. [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 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 p-adic numbers. Using that I'm able to improve the bound for the VC function, showing that $\mathrm{vc}(n) = n$.

1. VC-dimension and VC-density

Definition 1.1. Throughout this section we work with a collection $\mathcal{F}$ of subsets of X.

- Call it a <u>set system</u> $(X, \mathcal{F})$.

- Define intersection $A \cap \mathcal{F} = \{A \cap F\}_{F \in \mathcal{F}}$.
- For $A \subset X$ we say that $\mathcal{F}$ shatters A if $A \cap \mathcal{F} = \mathcal{P}(A)$.

Definition 1.2. We say $(X, \mathcal{F})$ has VC-dimension n if the largest set it shatters is of size n. If it can shatter arbitrarily large sets, we say that it has infinite VC-dimension. Denote it by $\mathrm{VC}(\mathcal{F})$.

This distinguishes 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 of a dual system:

Definition 1.3. For $a \in X$ define $X_a = \{F \in \mathcal{F} \mid a \in F\}$. Let $X^* = \{X_a\}_{a \in X}$. We define $(\mathcal{F}, X^*)$ as the dual system of $(X, \mathcal{F})$. VC-dimension of a dual system is referred to as the dual VC-dimension and denoted by $\mathrm{VC}^*(\mathcal{F})$.

Lemma 1.4 (Lemma 2.5 in [1])**.** *A set system has finite VC-dimension if and only if its dual has finite VC-dimension. More precisely*

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

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

Definition 1.5. Define a shatter function $\pi_{\mathcal{F}}(n)$ and a dual shatter function $\pi^*_{\mathcal{F}}(n)$

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

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

Note that the dual shatter function is precisely the shatter function of the dual system.

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

Theorem 1.6 (Sauer-Shelah '72)**.** *If the set system $(X, \mathcal{F})$ has finite VC-dimension d then $\pi_{\mathcal{F}}(n) \le \binom{n}{\le d}$.*

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

$$\mathrm{vc}(\mathcal{F}) = \limsup_{n\to\infty} \frac{\log \pi_{\mathcal{F}}}{\log n}$$
$$\mathrm{vc}^*(\mathcal{F}) = \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}}}{\log n}$$

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 1.8. Work in a structure M. Fix a finite collection of formulas $\Psi(x, y) = \{\phi_i(x, y)\}$.

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

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 1.9.

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

$$\mathrm{vc}(\Psi) = \text{degree of polynomial growth of } \pi_\Psi(n) = \limsup_{n\to\infty} \frac{\log \pi_\Psi}{\log n}$$

One can check that 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:

Lemma 1.10. *ϕ is NIP if and only 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 1.11. For a given NIP structure M, define the <u>vc-function</u>

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

As before this definition is elementary, so it only depends on the theory of M. One can easily check the following bounds:

Lemma 1.12 (Lemma 3.22 in [1])**.**

$$\mathrm{vc}(1) \geq 1$$

$$\mathrm{vc}(n) \geq n\,\mathrm{vc}(1)$$

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

2. P-ADIC NUMBERS

P-adic numbers are often studied in the language of Macintyre $\mathcal{L}_{Mac} = \{0, 1, +, -, \cdot, P_n\}$. which is a language of fields together with unary predicates $\{P_n\}_{n\in\mathbb{N}}$ interpreted by

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

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

Theorem 2.1 (???)**.** *$(\mathbb{Q}_p, \mathcal{L}_{Mac})$ has quantifier elimination.*

There is also the following cell decomposition result:

Theorem 2.2 (???)**.** *Any formula $\phi(t, x)$ in $(\mathbb{Q}_p, \mathcal{L}_{Mac})$ with t singleton decomposes into the union of the following cells:*

$$\{(t, x) \in K \times D \mid \operatorname{val} a_1(x) \;\Box_1\; \operatorname{val}(t - c(x)) \;\Box_2\; \operatorname{val} a_2(x), t - c(x) \in \lambda P_n\}$$

where D is a cell of a smaller dimension, $a_1(x), a_2(x), c(x)$ are $\emptyset$-definable, $\Box$ is $<, \leq$ or no condition, and $\lambda \in \mathbb{Q}_p$.

In [1], Aschenbrenner, Dolich, Haskell, Macpherson, and Starchenko show that this structure has $\mathrm{vc}(n) \leq 2n - 1$, 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 $\operatorname{val} a \leq \operatorname{val} b$, 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 $Q_{m,n}$ 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 full structure $(\mathbb{Q}_p, \mathcal{L}_{Mac})$. The following cell decomposition result is provided by [2]:

Theorem 2.3. *Any formula $\phi(t, x)$ in $(\mathbb{Q}_p, \mathcal{L}_{aff})$ with t singleton decomposes into the union of the following cells:*

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

where D is a cell of a smaller dimension, $a_1(x), a_2(x), c(x)$ are linear polynomials, $\square$ is $<$ or no condition, and $\lambda \in \mathbb{Q}_p$.

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

References

[1] M. Aschenbrenner, A. Dolich, D. Haskell, D. Macpherson, S. Starchenko, *Vapnik-Chervonenkis density in some theories without the independence property*, I, preprint (2011)

[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`