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

ANTON BOBKOV

ABSTRACT. [1] computed a bound $2n+1$ for the VC function in p-adic numbers, but it is not known to be optimal. I investigate a C-minimal additive reduct of p-adic numbers and using techniques of [2] I compute an optimal bound n for that structure.

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

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

Where $vc(n)$ measures complexity of definable sets in an n-dimensional space. 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's not known whether it's optimal. This same bound would hold in any reduct of p-adic numbers, so one may hope that the simplified structure of the reduct would allow a better bound. In [2], Leenknegt provides a cell decomposition result for the C-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. Cell Decomposition

We work with the reduct of p-adic numbers in the language $\mathcal{L}_R = \left\{\mathbb{Q}_p, \{Q_{n,m}\}_{n,m\in\mathbb{N}}, +, -, \{\bar{c}\}_{c\in K}\right\}$, where $\bar{c}$ is a scalar multiplication by c, and $Q_{n,m}$ is a unary predicate

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

[2] provides a cell decomposition result for this structure. Any formula $\phi(t,x)$ with t singleton decomposes as the union of the following cells:

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

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

Lemma 1.1. *For a formula $\phi(x)$ with $x = (t, \bar{x})$ there exists a family of formulas $\Psi'(x)$*

$$\begin{aligned}
&\operatorname{val}(q_i(x)) < \operatorname{val}(q_j(x)) && i,j \in I\\
&\operatorname{val}(q_i(x)) \in \lambda_k Q_{n,m} && i \in I, k \in K\\
&\bar{x} \in D_l && l \in L
\end{aligned}$$

with I, K, L finite, D_l cells, q_i linear polynomials, $\lambda_k \in \mathbb{Q}_p$, and $Q = Q_{n,m}$ for some n,m. Moreover we have that if $a, a' \in Q_p^{|x|}$ agree on all formulas from Ψ' then they agree on ϕ.

Proof. To see that apply cell decomposition theorem to $\phi(t, \bar{x})$. Let q_i enumerate all of the following polynomials $a_1(\bar{x}), a_2(\bar{x}), t - c(\bar{x})$ that show up in the cells Let D_l be the smaller cells for $\bar{x}$ component that appear in the cells. Choose n, m large enough to cover all n', m' that come up in the cells for $Q_{n',m'}$. Choose λ_k to go over all cosets of $Q_{n,m}$. $\square$

Applying this lemma inductively to smaller cells that appear we obtain a family $\Psi(x)$

$$\begin{aligned}
&\operatorname{val}(q_i(x)) < \operatorname{val}(q_j(x)) && i,j \in I\\
&\operatorname{val}(q_i(x)) \in \lambda_k Q_{n,m} && i \in I, k \in K
\end{aligned}$$

with I, K finite, q_i linear polynomials, $\lambda_k \in \mathbb{Q}_p$, and $Q = Q_{n,m}$ for some n, m. Moreover whenever $a, a' \in Q_p^{|x|}$ agree on all formulas from Ψ then they agree on ϕ.

Now fix a formula $\phi(x; y)$ for finding an upper bound of its VC-density. Using the result above we can construct a family of formulas $\Psi(x; y)$ which can be now written as

$$\begin{aligned}
&\operatorname{val} p_i(x) - c_i(y) < \operatorname{val} p_j(x) - c_j(y) && i,j \in I\\
&\operatorname{val} p_i(x) - c_i(y) \in \lambda_k Q && i \in I, k \in K
\end{aligned}$$

where I, K finite, p_i a homogeneous linear polynomials in x, c_i is a linear polynomial in y, $\lambda_k \in \mathbb{Q}_p$, and $Q = Q_{n,m}$ for some n, m (to do this we simply split polynomials q_i into its x part and into its y part including the constant term). Now for any parameter set B we have that if a, a' have the same Ψ-type then they have the same ϕ-type. Thus it suffices to bound VC-density for Ψ.

2. Key Lemmas and Definitions

Definition 2.1. A tuple $p \in \mathbb{Q}_p^{|x|}$ can be viewed as a vector $\vec{p}$, treating $\mathbb{Q}_p^{|x|}$ as a vector space over $\mathbb{Q}_p$.

We may rewrite our collection of formulas $\Psi(x, y)$ as

$$\begin{aligned}
&\operatorname{val}(\vec{p}_i \cdot \vec{x}) - c_i(y) < \operatorname{val}(\vec{p}_j \cdot \vec{x}) - c_j(y) && i,j \in I\\
&\operatorname{val}(\vec{p}_i \cdot \vec{x}) - c_i(y) \in \lambda_k Q && i \in I, k \in K
\end{aligned}$$

Lemma 2.2. *Suppose we have a collection of vectors $\{\vec{p}_i\}_{i\in I}$ with each $\vec{p}_i \in \mathbb{Q}_p^{|x|}$. Pick a subset $J \subset I$ and $j \in I$ such that*

$$\vec{p}_j \in \operatorname{span}\{\vec{p}_i\}_{i\in J}$$

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

$$\operatorname{val}(\vec{p}_i \cdot \vec{x}) > \alpha \text{ for all } i \in J$$

Then

$$\operatorname{val}(\vec{p}_j \cdot \vec{x}) > \alpha - \gamma$$

for some $\gamma \in \mathbb{Z}^{\geq 0}$. Moreover γ can be chosen independent of choice of $J, j, \vec{x}, \alpha$ depending only on $\{\vec{p}_i\}_{i\in I}$ independent of their order.

Definition 2.3. For $c \in \mathbb{Q}_p, \alpha \in \mathbb{Z}$ we define an open ball

$$B(c, \alpha) = \{c' \in \mathbb{Q}_p \mid \operatorname{val}(c' - c) \leq \alpha\}$$

Definition 2.4. Suppose we have a finite $T \subset \mathbb{Q}_p$. We view it as a tree as follows. Branches through the tree are elements of T. With this tree we associate open balls $B(t_1, \operatorname{val}(t_1 - t_2))$ for all $t_1, t_2 \in T$. An interval is two balls $B(t_1, v_1) \supset B(t_2, v_2)$ with no balls in between. An element $a \in \mathbb{Q}_p$ belongs to this interval if $a \in B(t_1, v_1) \backslash B(t_2, v_2)$. There are at most $2|T|$ different intervals and they partition the entire space.

Fix a parameter set B of size N.

Consider a tree $T = \{c_i(b) \mid b \in B, i \in I\}$ It has at most $O(N) = N \cdot |I|$ many intervals. Denote the set of all intervals as Pt. For the remainder of the paper we work with this tree.

Definition 2.5. $a, a' \in \mathbb{Q}_p^{|x|}$ have the same Ψ-type if they have the same Ψ type over B.

Definition 2.6. $x, x' \in \mathbb{Q}_p$ have the same tree type if

- $x + c_i(b)$ is in the same Q-coset as $x' + c_i(b)$ for all $i \in I, b \in B$
- $\operatorname{val}(x + c_i(b)) < \operatorname{val}(x + c_j(b))$ iff $\operatorname{val}(x' + c_i(b)) < \operatorname{val}(x' + c_j(b))$ for all $i, j \in I, b \in B$

Lemma 2.7. *Let $a, a' \in \mathbb{Q}_p^{|x|}$. If $p_i(a), p_i(a')$ have the same tree type for all $i \in I$, then a, a' have the same Ψ-type.*

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

Lemma 2.8. *For n, m there exists $D = D(n,m) \in \mathbb{Z}$ such that for any $x, y, a \in \mathbb{Q}_p$ if*

$$\mathrm{val}(x-a) = \mathrm{val}(y-a) < \mathrm{val}(x-y) - D$$

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

Next lemma is along the lines of lemma 7.5 of [1].

Lemma 2.9. *Using D from the previous lemma define an enumeration of near balls*

$$B_1(c,\alpha), B_2(c,\alpha), \dots B_{N_D}(c,\alpha)$$

Definition 2.10. Let $c \in \mathbb{Q}_p$. It lies in our tree in one of the intervals $B(c_L,\alpha_L)\backslash B(c_U,\alpha_U)$. Suppose c lies in one of the near balls corresponding to $B(c_L,\alpha_L)$ or $B(c_U,\alpha_U)$. Then define its interval type to be the index of that near ball. Otherwise define its interval type to be the coset of $c - c_U$ of Q. Denote the space of all the possible branch types Bt. We have

$$|\,\mathrm{Bt}\,| = N_D + \text{number of cosets of } Q$$

depending only on Ψ, independent from B.

Lemma 2.11. *If c, c' are in the same interval and have the same interval type then they have the same tree type.*

Definition 2.12. For $c \in \mathbb{Q}_p$ and $\alpha, \beta \in \mathbb{Z}$ let $c \restriction [\alpha,\beta] \in \mathbb{Z}/p\mathbb{Z}^{\beta-\alpha}$ be the record of coefficients of c for valuations between α, β. More precisely write c in its power series form

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

Then $c \restriction [\alpha,\beta]$ is just $(c_\alpha, c_{\alpha+1}, \dots c_\beta)$.

Definition 2.13. Let $c \in \mathbb{Q}_p$. It lies in our tree in one of the intervals $B(c_L,\alpha_L)\backslash B(c_U,\alpha_U)$. Define $F(c)$, the floor of c to be α_L.

3. Main Proof

Fix γ corresponding to $\{\vec{p}_i\}_{i\in I}$ according to Lemma 2.2.

Definition 3.1. Denote $\mathbb{Z}/p\mathbb{Z}^\gamma$ as Ct.

Definition 3.2. Let $f : \mathbb{Q}_p^{|x|} \longrightarrow \mathbb{Q}_p^I$ with $f(\bar{c}) = (p_i(\bar{c}))_{i\in I}$. Define segment space Sg to be the image of f.

Given a tuple $(a_i)_{i\in I}$ in segment space look at corresponding floors $\{F(a_i)\}_{i\in I}$. Those are ordered as elements of $\mathbb{Z}$ Partition the segment space by order type of $\{F(a_i)\}$. Work in a fixed partition Sg$'$. After relabeling we may assume that

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

Consider (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.

Now, we define the following function

$$g : \mathrm{Sg}' \longrightarrow \mathrm{Bt}^I \times \mathrm{Pt}^J \times \mathrm{Ct}^{I-J}$$

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

For all a_i record its interval type $\in$ Bt giving the first component.

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

For the third component do the following computation. Pick a_i with i dependent. Let j be the largest independent index with $j < i$. Record $a_i \restriction [F(a_j) - \gamma, F(a_j)]$.

Lemma 3.3. *For $\bar{a}, \bar{a}' \in \mathrm{Sg}'$ if $g(\bar{a}) = g(\bar{a}')$ then a_i, a_i' have the same tree type for all $i \in I$.*

Proof. For each i we show that a_i, a_i' are in the same interval and have the same interval type, so the conclusion follows by Lemma 2.11. Bt records interval type of each element, so if $g(\bar{a}) = g(\bar{a}')$ then a_i, a_i' have the same interval type for all $i \in I$. Thus it remains to show that a_i, a_i' lie in the same interval for all $i \in I$. Suppose i is an independent index. Then by construction, Pt records interval for a_i, a_i', so those have to belong to the same interval. Now suppose i is dependent. Pick 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 they are mapped to the same interval (using the earlier part of the argument as j is independent).

Claim 3.4. $\mathrm{val}(a_i - a_i') > F(a_j) - \gamma$

Proof. Let $\vec{x}, \vec{x}' \in \mathbb{Q}_p^{|x|}$ be some elements with

$$\vec{p}_k \cdot \vec{x} = a_k$$
$$\vec{p}_k \cdot \vec{x}' = a_k' \text{ for all } k \in I$$

It is always possible to do that as $\bar{a}, \bar{a}' \in \mathrm{Sg}'$. Let J' be the set of independent indices less than i. We have

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

as for independent indices a_k, a_k' lie in the same interval.

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

J' and i match the requirements of Lemma 2.2 so we conclude

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

as needed. □

moreover because a_i, a_i' have the same image in Ctcomponent under g we can conclude that

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

As $F(a_i) \leq F(a_j)$, a_i, a_i' have to lie in the same interval. □

Corollary 3.5. *$\Psi(x,y)$ has 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 $p_i(c)$ has the same tree type as $p_i(c')$ for all $i \in I$. Then by Lemma 2.7 c, c' have the same Ψ-type. Thus the number of possible Ψ-types is bound by the size of the range of g times the number of possible partitions.

$$|\,\mathrm{Ct}\,| = p^\gamma$$
$$|\,\mathrm{Pt}\,| \leq N \cdot I^2 \text{ (the only component dependent on } N)$$

Moreover we need at most $|I|!$ many partitions of Sg. This gives us

$$|I|! \cdot |Bt|^{|I|} \cdot (N \cdot |I|^2)^{|J|} \cdot p^{\gamma|I-J|} = O(N^{|J|})$$

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 $|J| \leq |x|$ as needed. □

Corollary 3.6. *In the language $\mathcal{L}_R$ we have* $\mathrm{vc}(n) = n$.

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. □

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] insert citation

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