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 using a cell decomposition result of Leenknegt I 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 a NIP theory we can define the VC function

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

where $vc(n)$ 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 hope 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. Cell Decomposition

Definition 1.1. Let

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

It is a subgroup of the multiplicative group of $\mathbb{Q}_p$ with finitely many cosets.

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

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

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

$$\{(t, x) \in K \times D \mid \mathrm{val}\, a_1(x) \square_1 \mathrm{val}(t - c(x)) \square_2 \mathrm{val}\, a_2(x), t - c(x) \in \lambda Q_{n',m'}\}$$

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

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

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 the 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 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 the $\bar{x}$ components 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 the cosets of $Q_{n,m}$. □

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

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

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 the 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

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

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 the polynomial 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 over B then they have the same ϕ-type over B. 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

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

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*

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

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

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

Then

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

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

Proof. Fix some i, J. For some c_i

$$\vec{p}_j = \sum_{i\in J} c_i \vec{p}_i$$
$$\vec{p}_j \cdot \vec{x} = \sum_{i\in J} c_i \vec{p}_i \cdot \vec{x}$$

We have

$$\mathrm{val}\,(c_i \vec{p}_i \cdot \vec{x}) = \mathrm{val}\,(c_i) + \mathrm{val}\,(\vec{p}_i \cdot \vec{x}) > \mathrm{val}\,(c_i) + \alpha$$

Pick $\gamma = -\max \mathrm{val}\,(c_i)$ or 0 if all those values are positive. Then we have

$$\begin{aligned}
&\mathrm{val}\,(c_i \vec{p}_i \cdot \vec{x}) > \alpha - \gamma && \text{for all } i \in J\\
&\sum_{i\in J} c_i \vec{p}_i \cdot \vec{x} > \alpha - \gamma
\end{aligned}$$

This shows that we can pick such γ for a given choice of i, J, but independent from $\alpha, \vec{x}$. To get a choice independent from i, J, go over all such eligible choices (of which there are finitely many as I is finite), pick γ for each, and then take the maximum of those values. □

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

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

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 $B(t_1, \mathrm{val}(t_1 - t_2))$ 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. Let $c \in \mathbb{Q}_p$. It lies in the tree in one of the unique intervals $B(c_L, \alpha_L) \backslash B(c_U, \alpha_U)$. Define $F(c)$, the floor of c to be α_L.

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

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

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.*

Proof. Clear from the construction. □

Definition 2.8. 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 the 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 $(c_\alpha, c_{\alpha+1}, \dots c_\beta)$.

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

Lemma 2.9. *For n, m there exists $D = D(n, m) \in \mathbb{Z}$ such that for any $x, y, a \in \mathbb{Q}_p$ if*

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

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

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

$$a \restriction [\mathrm{val}\, a, \mathrm{val}\, a + (m + n)] = b \restriction [\mathrm{val}\, b, \mathrm{val}\, b + (m + n)]$$

If a, b are similar then

$$a \in Q_{n,m} \leftrightarrow b \in Q_{n,m}$$

Moreover for any $\lambda \in \mathbb{Q}_p$, if a, b are similar we would also have $a/\lambda, b/\lambda$ are similar. Thus if a, b are similar, then they belong in the same coset of $Q_{n,m}$. If we pick $D = n + m$ then conditions of the lemma force $x - c, y - c$ to be similar. □

The following construction is along the lines of lemmas 7.3, 7.5 of [1].

Definition 2.10. For two balls $B(a, \alpha), B(b, \beta)$ let $\gamma = \min(\alpha, \beta, \mathrm{val}(a - b))$. Define the distance between those two balls to be $|\alpha - \gamma| + |\beta - \gamma|$. In $\mathbb{Q}_p$ value group is discrete and residue field is finite, so there are finitely many balls at a fixed distance from a given ball. Near balls of $B(a, \alpha)$ are defined to be balls with distance $\mathcal{D}$ from $B(a, \alpha)$. Enumerate those as:

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

Near balls partition the space

$$\{b \in \mathbb{Q}_p \mid |\mathrm{val}(a - b) - \alpha| \le D\}$$

Definition 2.11. 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 of $B(c_L, \alpha_L)$ 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.

Lemma 2.12. *If a, a' are in the same interval and have the same interval type then they have the same tree type.*

Proof. First part of the tree type definition is satisfied as a, a' are in the same interval, so we only need to demonstrate that the corresponding Q-cosets match. Pick any element of our tree $c_i(b)$. We want to show that $a - c_i(b), a' - c_i(b)$ are in the same Q-coset.

Suppose a is in one of the near balls. As a' has the same interval type, it has to be in the same near ball. By definition of the near ball we then have $\mathrm{val}(a - c_i(b)) = \mathrm{val}(a' - c_i(b)) < \mathrm{val}(a - a') - D$. Thus by Lemma 2.10 we have $a - c_i(b), a' - c_i(b)$ in the same Q-coset.

Now, suppose both a, a' aren't in any near balls. Label their interval as $B(c_L, \alpha_L) \backslash B(c_U, \alpha_U)$. Then we have

$$\alpha_L + D < \mathrm{val}(a - c_U) < \alpha_U - D$$

$$\alpha_L + D < \mathrm{val}(a' - c_U) < \alpha_U - D$$

as otherwise one (both) of them would be in one of the near balls. We have either $\mathrm{val}(c_U - c_i(b)) \ge \alpha_U$ or $\mathrm{val}(c_U - c_i(b)) \le \alpha_L$ as otherwise it would contradict the definition of an interval.

Suppose it is the first case $\mathrm{val}(c_U - c_i(b)) \ge \alpha_U$. Then

$$\mathrm{val}(a - c_i(b)) = \mathrm{val}(a - c_U) < \alpha_U - D \le \mathrm{val}(c_U - c_i(b)) - D$$

so by Lemma 2.10 we have $a - c_i(b), a - c_U$ are in the same Q-coset. By a parallel argument we have $a' - c_i(b), a' - c_U$ are in the same Q-coset. As we are assuming a, a' have the same tree type it implies that $a - c_U, a' - c_U$ 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 $\mathrm{val}(c_U - c_i(b)) \le \alpha_L$. Then

$$\mathrm{val}(a - c_i(b)) = \mathrm{val}(c_U - c_i(b)) \le \alpha_L < \mathrm{val}(a - c_U) - D$$

so by Lemma 2.10 we have $a - c_i(b), c_U - c_i(b)$ are in the same Q-coset. By a parallel argument we have $a' - c_i(b), c_U - c_i(b)$ 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. □

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 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 $\{F(a_i)\}_{i \in I}$. 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

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

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.

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 maps $\bar{a}$ 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 2.13. Bt records the 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 the interval for a_i, a_i', so those have to belong to the same interval. Now suppose i is dependent. Pick the largest $j < i$ such that j is independent. We have $F(a_i) \le F(a_j)$ and $F(a_i') \le F(a_j')$. 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 the independent indices less than i. We have

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

as for the 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, finishing the proof of the claim. □

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

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

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 2.8 c, c' have the same Ψ-type. Thus the number of possible Ψ-types is bounded by the size of the range of g times the number of possible partitions

$$(\text{number of partitions}) \cdot |Bt|^{|I|} \cdot |Pt|^{|J|} \cdot |Ct|^{|I-J|}$$

We have

$$|\,\mathrm{Bt}\,| = N_D + \text{number of cosets of } Q|\,\mathrm{Pt}\,| \leq N \cdot I^2 \text{ (the only component dependent on } N)$$
$$|\,\mathrm{Ct}\,| = p^\gamma$$

and there are at most $|I|!$ many partitions of Sg. This gives us a bound

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

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 $|J| \leq |x|$ and the bound follows. □

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

Proof. Previous lemma implies that $\mathrm{vc}(\phi) \leq \mathrm{vc}(\Psi) \leq |x|$. 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] E. Leenknegt. *Reducts of p-adically closed fields*, Archive for Mathematical logic, 53(3):285-306, 2014

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