Approximating Volumes and Integrals Via the VC - Dimension

Marek Karpinski (Bonn)* Angus Macintyre (Oxford)†

December 8, 1995

Abstract

We introduce a new powerful method of approximating the volume (and integrals) for a vast number of geometric body classes defined by boolean combinations of Pfaffian conditions. The method depends on the VC-Dimension of the class of bodies. The resulting approximation algorithms are quite different in spirit from the other up to now known methods, and give randomized solutions even for such seemingly untouchable problems of statistical physics like computing the volume of sets defined by the systems of exponential (or more generally Pfaffian) inequalities.

*Dept. of Computer Science, University of Bonn, 53117 Bonn. Email: `marek@theory.cs.uni-bonn.de`

†Mathematical Institute, University of Oxford, Oxford OX1 3LB. Email: `ajm@maths.ox.ac.uk`

1 Introduction

In [KM95] we gave a powerful method for estimating VC-dimension of classes defined by Boolean combinations of C^∞ (infinitely differentiable) conditions $f(\overline{v}) > 0$. In this paper we exploit these results to give good randomized polynomial time algorithms for *absolute approximation* of various volumes and integrals defined "in the Pfaffian category". These algorithms are quite different in spirit from the well-known ones of Dyer, Freeze and Kannan [DFK91] and in many concrete cases are faster.
In addition we consider the issue of formulas for ε- approximation of volume of definable sets (and integrals of definable functions), and we do this not only in *real* analysis but also in p-adic analysis. Due to our current ignorance of good bounds for VC-dimension in p-adic algebra or analysis, we cannot yet give efficient randomized algorithms in that setting.
The first of our results were obtained in October-November 1994 just prior to the Dagstuhl meeting on Neural Computing. In the real case they overlap with these of Pascal Koiran [K95], with whom we had at Dagstuhl some technical discussions on approximate definition of volume via VC-dimension.

2 Measure and VC-Dimension

2.1 We assume familiarity with the basics about VC-dimension [AB92], [L92], [GJ93], with our paper [KM95], and the basics on approximating the volume of [DF88], [DFK91]. We refer to [H50] for the most basic notions of the measure theory, and to [H76] for the basic notions from differential topology.
Our usual setting will involve a set X and a class $\mathcal{C}$ of subsets of X, of finite VC-dimension d (VC-Dim$(\mathcal{C}) = d$). Model theory provides a wide selection of such X and $\mathcal{C}$. If Γ is a theory without the *independence* property (cf. [L92]), and $\mathcal{M}$ is any model $\mathcal{C}$, we get X and $\mathcal{C}$ as follows:
$X = \mathcal{M}^k$, some k;
$\mathcal{C}$ is the class of all subsets of $\mathcal{M}^k$ defined by $\phi(v_1, \ldots, v_k, \beta_1, \ldots, \beta_\ell)$, for a fixed formula ϕ, and $\tilde{\beta}$ ranging over $\mathcal{M}^\ell$.
For this see [L92], where (astronomical) bounds on VC-Dim$(\mathcal{C})$ are established. (We note that one can generalize significantly. X could be a set *interpretable* in $\mathcal{M}$. In analytic situations, this would give us access to manifolds, and not just affine spaces).
There are two notable classes of examples:

1. Γ is o-minimal [L92]. The largest current example, with enormous expressive power, is the theory of the real exponential field, with primitives for all restricted analytic functions [DMM94];

2. $\mathcal{T}$ is the theory of p-adic field with primitives for all restricted analytic functions [DD88]. That this does not have the independence property

follows from [De84], [DD88] and [L92].

A conspicuous difference between (1) and (2) is that in (1) one has good bounds for VC-Dim($\mathcal{C}$) in many cases [KM95], whereas in (2) one has no good bounds so far.
There are also important cases where one has good bounds for VC-dimension without knowing any general result about lack of the independence property. The most important examples are in [KM95] for $\mathcal{C}$ defined by boolean combinations of $f(\bar{v},\bar{\beta}) > 0$, for f being Pfaffian [K91]. At present we do not know any general o-minimality result for unbounded Pfaffian functions.

2.2 We have stressed the real and p-adic cases, because $\mathbb{R}$ and $\mathbb{Q}_p$, being locally compact groups, carry Haar measures (cf. [H50]). For convenience, fix, in each case, such a measure μ giving measure 1 to unit ball.
Lemma 1 Suppose that $\mathcal{M}$ is either

i) $\mathbb{R}$ with some o-minimal structure, or

ii) $\mathbb{Q}_p$ with the structure of restricted analytic functions.

Then every definable subset of $\mathcal{M}^k$ is μ-measurable.
Proof By cell-decomposition [KPS86], [DD88] every definable subset is Borel. $\square$

2.3 The basis for the randomized approximation algorithms below is the fundamental [V82], [BEHW90]:
Theorem 2 Suppose $\mathcal{C}$ is a class of subsets of X, of VC-dimension d. Let P be a probability distribution on X, and P^n the product distribution on X^n. Suppose P satisfies the condition (#) (to be given below). Then if $n \geq max(\frac{4}{\varepsilon}\log\frac{2}{\delta}, \frac{8d}{\varepsilon}\log\frac{13}{\varepsilon})$

$$P^n\left(\{(x_1,\ldots,x_n) : sup_{C\in\mathcal{C}}\left|\frac{1}{n}\sum_i \chi_c(x_i) - P(C)\right| < \varepsilon\}\right) > 1-\delta$$

(Here χ_c is the characteristic function of C, and (#) ensures that the set displayed above is measurable).
The condition (#) is just that for each n the functions

$sup_{C\in\mathcal{C}}\left|\frac{1}{n}\sum_i \chi_c(x_i) - P(C)\right|$, and

$sup_{C\in\mathcal{C}}\left|\frac{1}{n}\sum_i \chi_c(x_i) - \frac{1}{n}\sum_i \chi_c(y_i)\right|$

are measurable.
Lemma 3 (#) holds for o-minimal $\mathcal{M}$, and for the p-adic subanalytic case.

Proof Lemma 1. □
A notable special case of Theorem 2, called ε-nets, used in [K95], is:
Theorem 4 (Assumptions as in Theorem 3) If $n \geq \frac{8d}{\varepsilon} \log \frac{13}{\varepsilon}$ there is a subset $\{x_1, \ldots, x_n\}$ of X such that for all $C \in \mathcal{C}$ with $P(C) \geq \varepsilon$,
$C \cap \{x_1, \ldots, x_n\} \neq \emptyset$.

3 A More General Version of Koiran's "Volume Approximation Formula"

3.1 Koiran [K95] works relative to a first-order structure on the $\mathbb{R}$ including $+, -, <$ as primitives. Let L be the language involved.
Koiran considers a measurable $E \subseteq \mathbb{R}^p$, a formula $\varphi(v)$ of L, and $\varepsilon > 0$ in $\mathbb{R}$.
Definition 1 φ defines an ε-approximate volume for E if

a) $\mathbb{R} \models \varphi(r) \longrightarrow |r - \mu^p(E)| < \varepsilon$

b) $|r - \mu^p(E)| < \frac{\varepsilon}{4} \longrightarrow \mathbb{R} \models \varphi(r)$

He points out the well-known fact that even for semi-algebraic E one cannot demand
$\mathbb{R} \models \varphi(r)| \longleftrightarrow |r - \mu^p(E)| < \varepsilon \qquad (*)$
if φ is semi-algebraic, because π is undefinable in the semi-algebraic category. One can give more elaborate examples, for richer structure on $\mathbb{R}$ (using, e.g. the undefinability results in [DMM94]). There is however the intriguing possibility that for some huge o-minimal structure on $\mathbb{R}$ (*) is possible for definable E. This may have something to do with the error function.
Koiran [K95], by two methods, proves results of the following shape:

If E is definable, there is a formula $\varphi(v)$ which defines an ε-approximate volume for E.
We will elaborate these methods, and compare them. The second method (that of Koiran's Theorem 4) [K95] was known to us, and is related to more general model theoretic considerations. The first is very elementary.

3.2 First Method It appears that Koiran here does not wish to appeal to o-minimality. He defines $\kappa(E)$ as the maximum number of connected components of $E \cap L$, for any axis parallel line. It is however worth observing that you cannot have all $\kappa(E)$ finite, for definable E, withput Γ being o-minimal. On the other hand, if E is semi-Pfaffian, $\kappa(E)$ is finite and can be useful estimated [K91].
In both cases one has a uniformity for $\kappa(E)$.
Lemma 5 Suppose either a) $\mathcal{M}$ is o-minimal, and $\Phi(v_1, \cdots, v_p, w_1, \cdots, w_l)$ is arbitrary;
or b) Φ is defined by a Boolean combination of Pfaffian primitives.

For $\tilde{\beta} \in \mathcal{M}^l$, let $E_{\tilde{\beta}}$ be defined by $\Phi(\bar{v}, \tilde{\beta})$. Then there is an integer K_Φ depending only on Φ such that for any line L parallel to an axis in $\mathbb{R}^p$, and any $\tilde{\beta}$.
$E_{\tilde{\beta}} \cap L$ has $\leq K_\Phi$ connected components.
Proof For (a), see [KPS86], for (b) [K91].
Now let $I = [0,1]$, and assume $E \subset I^P$, E μ-measurable, and $\kappa(E)$ finite. Let N be an integer, and put on I^P the grid of size $h = \frac{1}{N}$. Let μ_h be the probability measure on I^P uniformly distributed on the N^P vertices of the grid. By a simple Fubini argument Koiran proves the following, which for him generalizes the case of converse $\kappa(E) = 1$.
$\mid \mu_h(E) - \mu(E) \mid \leq ph\kappa(E)$.
Proof. See [K95].

We now quickly consider algorithmic aspects of this. To calculate $\mu_h(E)$ one must simple N^P points.

Fix $\epsilon > 0$, and let N be chosen so $\frac{pk}{N} < \frac{\epsilon}{4}$. Then $\mid \mu_h(E) - \mu(E) \mid < \frac{\epsilon}{4}$.
Suppose $\mid \mu_h(E) - r \mid < \frac{\epsilon}{2}$. Then $\mid r - \mu(E) \mid < \epsilon$
Conversely, suppose $\mid r - \mu(E) \mid < \frac{\epsilon}{4}$. Then $\mid r - \mu_h(E) \mid < \frac{\epsilon}{4} + \frac{\epsilon}{4} = \frac{\epsilon}{2}$.
So $\mid r - \mu_h(E) \mid < \frac{\epsilon}{2}$.
ϵ - approximates $\mu(E)$, and is clearly first-order.
Moreover, in terms of conventional logical classifications, when E is defined by $\Phi(\tilde{v}, \tilde{\beta})$ as done earlier, we note that $\mid r - \mu_h(E) \mid < \frac{\epsilon}{2}$ is both Σ_m and Π_m in the $\tilde{\beta}$, of Φ is both Σ_m and Π_m in $\tilde{v}, \tilde{\beta}$.

But, the formula is very long. A routine calculation shows that its length is of order $2^{N^P} \cdot N^P \cdot \mid \Phi \mid$, and so is exponential in $\frac{\kappa}{\epsilon}$.
We summarize:
Theorem 7 (Notation as above).

(i) If Φ is both Σ_m and Π_m, then given ϵ there is a formula $\varphi(v, w_1, \ldots, w_e)$, also both Σ_m and Π_m, such, for any $\tilde{\beta}$, $\varphi(v, \tilde{\beta})$ defines an ϵ–approximate volume for $\Phi_{\tilde{\beta}}$.

(ii) φ has length of order $2^{N^P} \cdot N^P \cdot \mid \Phi \mid$, where N is $\frac{4p\kappa}{\epsilon}$.

Remarks

a) (i) is clearly of theoretical interest.

b) That the length bound in (ii) is exponental in $\frac{1}{\epsilon}$ makes it useless for the algorithmic purpose of approximating the volume of $E_{\tilde{\beta}}$, at least if one adopts any conventional technique of looking for an r – satisfying $\mid r - \mu^P(E_{\tilde{\beta}}) \mid < \epsilon$. Moreover, the size of N makes the method useless even as a randomized algorithm.

c) In the literature one is interested mainly in algorithms polynomial in P, the dimension of the ambient space [DF88, DFK91].

d) In many interesting cases K may be reasonably small. This is certainly so in the semi–algebraic case, if Φ is given as quantifier – free. Suppose $\Phi(\tilde{v}, \tilde{w})$ is a Boolean combination of conditions $f_i(\tilde{v}, \tilde{w}) > 0$,$i \leq s$. Let d_{ij} be the v_j-degree of f_i. Then it is easily seen that

$$max_j 2(\sum_{i \leq s} d_{ij}) + 1$$

If Φ is quantifier – free in the language of exponentation, Hardy's method [DF88, DFK91] can be applied to give bounds which are not too satisfactory.

In the quantifier – free Pfaffian case, Khovanskii's method applies, as in [KM95].

3.3 The Translation Method We maintain the preceding notation.
The translation method is more subtle, and very congenial to a modeltheorist who defines generic types in terms of group translation [P83]. The idea is of course to derandomize the sampling algorithm by relating *generic* and *of large measure*.
As before, the procedure is uniform in $\tilde{\beta}$, so we write E for $E_{\tilde{\beta}}$.
We can interpret $(\frac{\mathbb{R}}{\mathbb{Z}})^P$ in the L-Structure, using representations in $[0, 1[$. Let $\oplus, \ominus$ be the interpretations of the group operation, and the inverse. On $([0, 1[)^P$ these operations are quantifier-free defineable, using +, -, <.
Following Koiran [K95], define, for $k \in \mathbb{N}$, and $\nu, \alpha \in I$
$S_{\nu,\alpha} = \{(\chi_1, \ldots \chi_k) \in (I^P)^k :\mid \frac{1}{k} \sum_{i=1}^{k} \chi_E(\chi_i) - \nu \mid \leq \alpha\}$
(There is a hidden (uniform) dependence on $\tilde{\beta}$. If E is Σ_n und Π_n, so is $S_{\nu,\alpha}$.)
Let Θ be the family of all $\chi \ominus S_{\nu,\alpha}$, $\chi \in I^k P$. This is a family of sets in $(I^P)^k$, indexed by ν, α, β, χ.
Note that the length of (the natural definition of) $S_{\nu,\alpha}$ is of order $k * 2^k * \mid \Phi \mid$, as is that of $\chi \ominus S_{\nu,\alpha}$.
Finally, fix an m, and consider the natural formula C_m expressing that a union of m translates of $S_{\nu,\alpha}$ covers $(I^p)^k$. This formula is Σ_{n+2}, if E is both Σ_n and Π_n. Moreover, the first block of $\exists$'s has length m, and the first block of $\forall$'s has length pk. Inside this there is a disjunction of m formulas $S_{\nu,\alpha} \oplus t_i$, essentially of length that of $S_{\nu,\alpha}$.
Korian's Theorem 4 [K95] is (essentially):
Theorem 8 Let $d \geq$ VC–dim Θ. Let $m \geq (16 \log\, 26)d$, and $k \geq max\frac{(12ln\, m, 48ln\, 2)}{\epsilon^2}$. Then $C_m(\nu, \frac{\epsilon}{2})$ defines an ϵ–approximate volume for $E_{\tilde{\beta}}$.
Remarks

(a) The most obvious issue is what is gained by using C_m, logically more complex than the formula of Theorem 7. The length of C_m is essentially a constant times $m + pk + k * 2^k * \mid \Phi \mid$, so, with minimal choice of m, k a constant times $d + \frac{p \log\, d}{\epsilon^2} + \frac{p \log\, d}{\epsilon^2} * d^{\frac{12}{\epsilon^2}} * \mid \Phi \mid$.

(b) d can in some cases be estimated. In [K95] Koiran shows that $d \leq c * d' * k\, log\, k$, where C is an absolute constant and d' is the VC–dimension of the family $\chi \ominus E_{\tilde{\beta}}$. Note that this estimate is valid *before* we impose the constraints in Theorem 8.

d' can be nicely estimated in some cases where Φ is quantifier–free, and in the semi–algebraic case quite generally, using refined results on quantifier–elemination [GJ93]. Let us briefly recall the most notable result [KM95]:

Suppose $\Phi(\tilde{\nu}, \tilde{w})$ is a Boolean combination of s many formulas $f_i(\tilde{\nu}, \tilde{w}) > 0, \quad 1 \leq i \leq s$ where each f_i is a polynomial of degree $\leq \Delta$ in $\tilde{\nu}, \tilde{w}$ and q many (q independent of i) $\alpha(\tilde{\nu}, \tilde{w})$ which occur in a Pfaffian chain of length q and degree D. Then the VC-dimension of $E_{\tilde{\beta}}$ is bounded by

$$\begin{gathered} 2(ql)(ql-1) + 2l\, log\, \Delta + 2l\, log\, (l\Delta + lD + 1) \\ +2ql\, log\, l + 2ql\, log\, (l\Delta + lD + 1) \\ +l(16 + 2\, log\, s). \end{gathered}$$

$\circledcirc$

Note that this does not depend on the length of $\tilde{\nu}$!

To deal with the VC–dimension d of family $\chi \ominus E_{\tilde{\beta}}$ some minor adjustments are needed, mainly to take account of the interpretation of $\frac{\mathbb{R}}{\mathbb{Z}}$ via [0,1[. We omit constantly the details of the argument, but the only change needed in $\circledcirc$ is to replace s by $s + 4p$.

(c) The main point to notice vis-a-vis $\mid C_m \mid$ is that there is a term exponential in $\frac{1}{\epsilon^2}$, but in the Pfaffian case there is no exponential dependence on p. So there is a slight gain over the earlier formula. *But*, if one tries, in the semi–algebraic case, to eliminate quantifiers à la Renegar [R92], the cost is great, because the block of $\forall$'s has length kp, leading to a time cost exponential in p and $\frac{1}{\epsilon^2}$.

(d) Though it seems to be of no practical importance, it is interesting to note that for fixed ϵ, and $\mid \nu - \mu(E_{\tilde{\beta}}) \mid < \frac{\epsilon}{4}$ the block of m $\exists$'s can always be instantiated by a tuple *independent of ν and $\tilde{\beta}$*. (This follows easily by the ϵ–net argument.) The (unsolved) problem is to find this tuple.

3.4 p–adic version Delon [De89] proved that the field $\mathbb{Q}_p$ does not have the independence property, and from this, using [DD88] and [L92], one easily deduces that $\mathbb{Q}_p$ with the subanalytic structure of van den Dries–Denef also does

not have the independence property. This of course gives finiteness of VC–dimension for families of sets interpretable in $\mathbb{Q}_p$ with subanalytic structure. But, till now, even for the pure field $\mathbb{Q}_p$, one does not have *good* bounds. This is, to some extent, connected with the subtlety of notions like (definably) connected component.
Let V be the unit ball of $\mathbb{Q}_p$ and μ the normalized Haar measure with $\mu(V) = 1$. All p–adic subanalytic sets are μ–measurable, by quantifier-elimination. So the problem of closely approximating the measure of a definable set is natural. But some modifications must be made to Koiran's definitions, since the values of the measure are not elements of the model. That is, you cannot just substitute real r for ν in $\mid \nu - \mu(E) \mid < \epsilon$ and make sense inside $\mathbb{Q}_p$.
We propose the following modification. We follow the notations of [K95], except that we replace I by V. V is now an additive group, so we avoid the details of interpretability. We work in a power $((\mathbb{Q}_p)^t)^k$ as before (replacing p by t for obvious reasons!). The relative frequency $\frac{1}{k} * \sum_{i=1}^{k} \chi_E(\chi_i)$ is one of the rationals $\frac{d}{k}$ $0 \leq j \leq k$. The least distance between these numbers is $\frac{1}{k}$, and,if $E \subseteq V$, $\mu(E)$ is no more than $\frac{1}{k}$ from some $\frac{j}{k}$. If $\frac{1}{k} < \frac{\epsilon}{4}$, then $\mu(E)$ is within $\frac{\epsilon}{4}$ of some *unique* $\frac{j}{k}$.
Now we can just follow Koiran's notation [K95] prior to his Theorem 4. Observe that his argument there works even better in the p–adic case. It just does not give a formula of the intended language. All we can do is to define (in the language), for each $j \leq k$ the set $\tilde{S}_j$ of $\bar{\chi}$ in $((\mathbb{Q}_p)^P)^k$ whose relative frequency is $\frac{j}{k}$. Choose the unique j such that $\mid \frac{j}{k} - \mu(E) \mid < \frac{\epsilon}{2}$. So, as in [K95] $\mu(S_{\frac{j}{k},\frac{\epsilon}{2}}) > 1 - e^{-\frac{k\epsilon^2}{48}}$, so $\mu(\tilde{S}_j) > 1 - e^{-\frac{k\epsilon^2}{48}}$ so by Koiran's argument (for m, k as there) m translates of $\tilde{S}_j$ cover $(V^t)^k$. Conversely, as in [K95], if m translates of $\tilde{S}_j$ cover $(V^t)^k$, $\mu(\tilde{S}_j) \geq \frac{1}{m} > e^{-\frac{k\epsilon^2}{12}}$. So $\mu(S_{\frac{j}{k},\frac{\epsilon}{2}}) > e^{-\frac{k\epsilon^2}{12}}$ as in [K95].
So we have proved, with the same notation as in Theorem 8.
Theorem 9 If m translates of $\tilde{S}_j$ cover $(V^t)^k$ then $\mid \frac{j}{k} - \mu(E) \mid < \epsilon$, and if $\mid \frac{j}{k} - \mu(E) \mid < \frac{\epsilon}{4}$ then m translates of $\tilde{S}_j$ cover $(V^t)^k$.

Remarks

(a) This has for now even less computational significance than Theorem 7 , since we don't know how to calculate the VC-dimensions.

(b) We don't know any analogue of Lemma 6, though we hope to find one.

4 Integrals

4.1 In the real case, the preceding methods are readily adapted to "express" certain integrals. The relevant integrals are $\int_{\mathcal{A}} f d\mu^P$, $\mathcal{A}$ a definable subset of I^P, and f a definable function on $\mathcal{A}$ with $\mid f(\bar{x}) \mid \leq 1$, $\bar{x} \in \mathcal{A}$. By splitting

$\mathcal{A}$ into the definable sets where f is negative or nonnegative, we reduce to case where $0 \le f(\bar{x}) \le 1$ on $\mathcal{A}$. Then $\int_{\mathcal{A}} f d\mu^P = \mu^{P+1}((\bar{x}, y) : 0 \le y \le f(\bar{x})\ , \bar{x} \in \mathcal{A})$, thereby making available all the results of Section 3.
Note that definable functions are measurable, in o–minimal theories!

4.2 The preceding of course covers the case of lengths of definable curves in o–minimal theories.

4.3 It is however much less clear how to deal with various areas of curved surfaces, if these are not surfaces of revolution.

4.4 p**–adic integrals** There is a considerable body of work on rationality of p–adic Poincare' series, or equivalently rationality in P^{-s} of $\int_{\mathcal{A}} \mid f(\bar{x}) \mid^s d\mu^t$, where $\mathcal{A}$, f are p–adic definable, and s is a real variable. See particularly [De89], [DD88].
For *rational* s one can use the method of 4.1, combined with the ideas of Section 3, to get "expressibility" in the p–adic analytic language. Here there is an interesting problem of taking account of *variation in* s. We hope to return to this in a later publication.

5 Randomized Algorithms

5.1 For us the main significance of our results [KM95] is that they lead to powerful randomized algorithms in a wide range of geometric problems with Pfaffian data. There is a noticeable difference between this approach and that of, say, Dyer–Freeze–Kannan [DFK91] who have as data blackboxes for arbitrary convex bodies, and are concerned with *relative* approximation. One of their desiderata is an algorithm running in time polynomial in p, the dimension of the ambient space, and this establishes a point of comparison with our method, where convexity plays no role.
We restrict to sets definied by quantifier–free conditions in the language with primitives for all total Pfaffian functions.

5.2 For us a typical Θ is defined by a Boolean combination of conditions $f_i(\tilde{\nu}, \tilde{w}) > 0$ as in [KM95], where the f_i are polynomials of degree $\le \Delta$ in $\tilde{\nu}, \tilde{w}$ and q many Pfaffian functions occuring in a chain of length q and degree $\le$ D.
Our bound for VC-dim(Θ) [KM95], following [K91, p. 91] is

$$q(q-1) + 2l \log \Delta + 2(l+q) \log (\Delta + \mathrm{D}) + 2l \log (l+1) + (16 + 2 \log\ sl),$$

in which k doesn't occur. Let d be the above estimate. Then if a sample of size $m \ge \max(\frac{4}{\epsilon} \log \frac{2}{\delta}, \frac{8d}{\epsilon} \log \frac{13}{\epsilon})$ is chosen, and its relative frequency, for C in Θ, is tested, this is within ϵ of $\mu(C)$ with probability $\ge 1 - \delta$.

So here one has a *randomized* algorithm, taking time mp (with blackbox convention) giving an absolute error $< \epsilon$ with probability $\geq 1 - \delta$. If we unpack d, and so m, we see that the time of the algorithm, say for $\delta = \frac{1}{4}$, is (up to absolute constants):

i) linear in p;

ii) $\frac{1}{\epsilon} \log(\frac{1}{\epsilon})$ in ϵ;

iii) quadratic in q;

iv) $l \log l$ in l.

This result, which we now summarize, seems to us more than competitive with those of [DFK91]:

Theorem 10 (Notation as above.) There is a randomized algorithm for absolute ϵ–approximation of a volume of semi-Pfaffian sets, with running time of order $p * q^2 \frac{l}{\epsilon} \log\ l \log\ (\frac{1}{\epsilon})$.

5.3 Analogue for Integrals Since, as remarked in **???**, integrals
$\int_A f \, d\mu^p$, where $0 \leq f \leq 1$
are special cases of volumes, there is an obvious analogue of Theorem 10 for integration of Pfaffian f over semi-Pfaffian A.

5.4 Doing better It seems likely that one can use the deep results of [V82, Chapter 7] to get randomized algorithms based on computation of means

$$\frac{1}{n} \sum f(x_i) \chi_A(x_i) \quad ,$$

instead of via 5.3. In the o–minimal case one has of course the finitness of capacity (cf. [V82]). We have not looked closely at the relative merits of this and the idea of 5.3.

5.5 Surface areas, etc. In contrast to the situation for expressibility, where translation invariance was crucial, the randomized algorithm method works for Borel measures on compact Θ-manifolds where Θ is a geometric category [DMM94], and in an analogous Pfaffian situation. In the latter case, fast algorithms will exist, via [KM95].

Acknowledgment

We thank Martin Dyer, Mark Jerrum, Pascal Koiran, Eduardo Sontag, and Andy Yao for many stimulating discussion.

References

[AB92] M. Anthony and N. Biggs, Computational Learning Theory: An Introduction, Cambridge University Press, 1992.

[AS93] M. Anthony and J. Shawe-Taylor, A Result of Vapnik with Applications, Discrete Applied Math. **47** (1993), pp. 207–217.

[BEHW90] A. Blumer, A. Ehrenfeucht, D. Haussler and M. Warmuth, Learnability and the Vapnik-Chervonenkis Dimension, J. ACM **36** (1990), pp. 929–965.

[BT90] A. Borodin, P. Tiwari, On the Decidability of Sparse Univariate Polynomial Interpolation, Proc. 22nd ACM STOC (1990), pp. 535–545.

[D92] L. van den Dries, Tame Topology and 0-minimal Structures, preprint, University of Illinois, Urbana, 1992; to appear as a book.

[DD88] J. Denef and L.P.D. van den Dries, p-adic and real subanalytic Sets, Annals of Mathematics 128 (1988), 79–138.

[De84] J. Denef, The Rationality of the Poincaré Series associated to the p-adic Points on a Variety, Invent. Math. 77 (1984), pp. 1–23

[De89] F. Delon, Définissabilité avec Paramètres Extérieurs dans $\mathbb{Q}_p$ et $\mathbb{R}$, Proc. AMS **106** (1909), pp. 193–198.

[DF88] M. E. Dyer and M. Frieze, On the Complexity of Computing the Volme of a Polyhedron, SIAM J. Comput. **17** (1988), pp. 967–974.

[DFK91] M. Dyer, A Frieze and R. Kannan, A Random Polynomial-Time Algorithm for Approximating the Volume of Convex Bodies, J. ACM **83** (1991), pp. 1–17

[DMM94] L. van den Dries, A. Macintyre and D.Marker, The Elementary Theory of Restricted Analytic Fields with Exponentation, Annals of Mathematics **140** (1994), pp 183-205.

[GJ93] P. Goldberg and M. Jerrum, Bounding the Vapnik Chervonenkis Dimension of Concept Classes Parametrized by Real Numbers. Machine Learning, 1994 (to appear). A preliminary version appeared in Proc. 6th ACM Workshop on Computational Learning Theory, pp. 361–369, 1993.

[H50] P. Halmos, Measure Theory, Chelsey-New York, 1950.

[H12] G. H. Hardy, Properties of Logarithmic-Exponential Functions, Proc. London Math. Soc. 10 (1912), pp. 54–90.

[H92] D. Haussler, Decision Theoretic Generalizations of the PAC Model for Neural Net and other Learning Applications, Information and Computation **100**, (1992), pp. 78–150.

[HKP91] J. Hertz, A. Krogh and R. G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, 1991.

[H76] M. W. Hirsch, Differential Topology, Springer-Verlag, 1976.

[K91] A. G. Khovanski, Fewnomials, American Mathematical Society, Providence, R.I., 1991.

[K95] P. Koiran, Approximating the Volume of Definable Sets, to appear in Proc. 36th IEEE FOCS (1995).

[KM94] M. Karpinski and A. Macintyre, Polynomial Bounds for VC Dimension of Sigmoidal Neural Networks, Proc. 27th ACM STOC (1995), pp.200-208.

[KM95] M. Karpinski and A. Macintyre, Bounding VC Dimension for Neural Networks: Progress and Prospects (Invited Lecture), Proc. EuroCOLT'95, Lecture Notes in Artificial Intelligence Vol.**904**, Springer-Verlag, 1995, pp. 337-341.

[KW93] M. Karpinski and T. Werther, VC Dimension and Uniform Learnability of Sparse Polynomials and Rational Functions, SIAM J. Computing **22** (1993), pp 1276–1285.

[KPS86] J. Knight, A. Pillay and C. Steinhorn, Definable Sets and Ordered Structures II, Trans. American Mathematical Society **295** (1986), pp.593-605.

[KS95] P. Koiran and E. D. Sontag, Neural Networks with Quadratic VC Dimension to appear in Advances in Neural Information Processing Systems (NIPS '95), 1995.

[L92] M. C. Laskowski, Vapnik-Chervonenkis Classes of Definable Sets, J. London Math. Society **45** (1992), pp 377–384.

[LS90] L. Lovasz and M. Simonovits, The Mixing Rate of Markov Chains, an Isoperimetric Inequality, and Computing the Volume, Proc. 31 IEEE FOCS (1990), pp. 346–355

[M93a] W. Maass, Perspectives of Current Research about the Complexity of Learning on Neural Nets, in: Theoretical Advances in Neural Computation and Learning, V. P. Roychowdhury, K. Y. Siu, A. Orlitsky (Editors), Kluwer Academic Publishers, 1994, pp. 295–336.

[M93b] W. Maass, Bounds for the Computational Power and Learning Complexity of Analog Neural Nets, Proc. 25th ACM STOC (1993), pp. 335–344.

[M94] W. Maass, Neural Nets with Superlinear VC-Dimension, Proc. of the International Conference on Artificial Neural Networks 1994 (ICANN '94), Springer (Berlin 1994), pp. 581–584; journal version appeared in Neural Computation **6** (1994), pp. 875–882.

[MSS91] W. Maass, G. Schnitger and E. D. Sontag, On the Computational Power of Sigmoidal versus Boolean Threshold Circuits, Proc. 32nd IEEE FOCS (1991), pp. 767–776.

[MS93] A. J. Macintyre and E. D. Sontag, Finiteness results for Sigmoidal Neural Networks, Proc. 25th ACM STOC (1993), pp.325–334.

[M64] J. Milnor, On the Betti Numbers of Real Varieties, Proc. of the American Mathematical Society **15** (1964), pp 275–280.

[M65] J. Milnor, Topology from the Differentiable Viewpoint, Univ.Press, Virginia, 1965.

[P83] B. Pouzat, Groupes stables, avec types géneriques réguliers, J. S. L. 48 (1983), pp. 339–355

[R92] J. Renagar, On the Computational Complexity and Geometry of the First-Order Theory of the Reals, Parts I, II, III, J. of Symb. Comput. **3** (1992), pp. 255–352.

[S42] A. Sard, The Measure of the Critical Points of Differentiable Maps, Bull. Amer. Math. Soc. **48** (1942), pp. 883–890.

[S-T94] J. Shawe-Taylor, Sample Sizes for Sigmoidal Neural Networks, Preprint, University of London, 1994, to appear in Proc. ACM COLT, 1995.

[S92] E. D. Sontag, Feedforward Nets for Interpolation and Classification, J. Comp. Syst. Sci. **45** (1992), pp. 20–48.

[TV94] G. Turan and F. Vatan, On the Computation of Boolean Functions by Analog Circuits of Bounded Fan-in, Proc. 35th IEEE FOCS (1994), pp. 553–564.

[V82] V. Vapnik, Estimation of Dependencies Based on Empirical Data, Springer Series in Statistics, 1982

[W68] H. E. Warren, Lower Bounds for Approximation by Non-linear Manifolds, Trans. of the AMS **133** (1968), pp. 167–178.

[W94] A. J. Wilkie, Model Completeness Results of Restricted Pfaffian Functions and the Exponential Function; to appear in Journal of the AMS, 1994.