Approximating the Volume of Definable Sets

Pascal Koiran*
LIP, ENS Lyon - CNRS
46, allée d'Italie
69364 Lyon Cedex 07
France
e-mail: koiran@lip.ens-lyon.fr

Abstract

The first part of this paper deals with finite-precision arithmetic. We give an upper bound on the precision that should be used in a Monte-Carlo integration method. Such bounds have been known only for convex sets; our bound applies to almost any "reasonable" set.

In the second part of the paper, we show how to construct in polynomial time first-order formulas that approximately define the volume of definable sets. This result is based on a VC dimension hypothesis, and is inspired from the well-known complexity-theoretic result "BPP $\subseteq \Sigma_2$".

Finally, we show how these results can be applied to sets defined by systems of inequalities involving polynomial or exponential functions. In particular, section 5 contains an application to a problem of structural complexity in the Blum-Shub-Smale model of computation over the reals.

1 Introduction

The Monte-Carlo integration method has been known for several decades, and has been studied extensively from the theoretical and practical point of views. The volume $\mu(E)$ of a set $E \subset [0,1]^p$ is approximated by $v = \sum_{i=1}^{k} 1_E(x_i)/k$, where the points $x_1, \ldots, x_k$ are drawn from the uniform distribution on $[0,1]^p$. One way to analyze this method is by Hoeffding's inequality (see section 2): for any $\epsilon > 0$, the probability that $|v - \mu(E)| \geq \epsilon$ is bounded by $2e^{-2k\epsilon^2}$.

One problem with this method is the assumption that we can actually generate random real numbers. In practice one has to work with finite precision numbers. In this paper, we would like to address the following problem: given some relevant information on the set E and an accuracy ϵ, what precision should be used ? This problem seems to be important both in theory and in practice, but surprisingly there is very little material on that question in the literature.

Working in finite precision amounts to estimating the proportion of the vertices of a finite grid that fall into the set. Let $h = 1/N$ be the grid size: the set of N^p vertices is $\mathcal{V} = \{(i_1/N, \ldots, i_p/N);\ i_1, \ldots, i_p = 0, 1, 2, \ldots, N-1\}$ (and the number of bits is roughly $\log N$). Let μ_h be the probability measure which is uniformly distributed on $\mathcal{V}$. Again by Hoeffding's inequality, the finite-precision Monte-Carlo method produces an estimate v such that $|v - \mu_h(E)| < \epsilon$ with probability at least $1 - 2e^{-2k\epsilon^2}$. Hence it just remains to bound the difference $|\mu(E) - \mu_h(E)|$ between the Lebesgue measure and the discrete measure. In section 3, we show that $|\mu(E) - \mu_h(E)| \leq ph\kappa(E)$, where $\kappa(E)$ is the maximum number of connected components of the intersection $L \cap E$ between E and an axis-parallel line L. Similar results have been known for convex sets (for such sets, $\kappa(E) = 1$). See for instance Proposition 3 in the appendix of the Dyer-Frieze-Kannan paper [6]. Their proof seems to rely heavily on convexity; our result is based on a different idea.[1]

Clearly, $\kappa(E)$ is reasonably small for many natural and man-made objects. In section 5, we give bounds on this parameter for sets that are defined by systems of polynomial equations or inequations. As an application, we show that at the polynomial-time level, continuous randomization is equivalent to discrete randomization in the Blum-Shub-Smale model of com-

*A part of this work was done when the author was visiting DIMACS at Rutgers University. He was supported by an INRIA fellowship and by Esprit working group 8556 (NeuroCOLT).

[1] Their main concern was not to obtain a grid size estimate, but rather to compute in polynomial time an approximate volume with *relative* error ϵ; here we work only with absolute error bounds.

0272-5428/95 $04.00 © 1995 IEEE

putation over the reals. We also discuss the case of systems where the exponential function occurs.

In section 4, we turn our attention to a problem which is more theoretical in nature, namely, defining the volume of "reasonable" sets by first-order logical formulas. It is well-known that many geometric concepts can be defined by formulas with a bounded number of quantifier alternations. For instance, in a theory which includes at least addition, multiplication and order, the diameter δ of a set E can be defined as follows:

$$\begin{aligned}&(\forall x, y \in E \ \sum_{i=1}^{p}(x_i - y_i)^2 \leq \delta)\wedge \\ &(\forall \epsilon > 0 \ \exists x, y \in E \ \sum_{i=1}^{p}(x_i - y_i)^2 \geq \delta - \epsilon).\end{aligned} \tag{1}$$

We would like to do the same thing for the volume. One problem here is that in the theory of the reals with addition, multiplication and order, the "volume" of an object as simple as the unit disk cannot be defined since π is a transcendent number (by quantifier elimination, only algebraic numbers are definable). Hence we shall deal with approximate rather than exact volumes (see section 4 for a precise definition). Our main result is that when the family of translates (modulo 1) of E has finite Vapnik-Chervonenkis (VC) dimension, it is possible to construct in polynomial time a formula which defines an approximate volume for E. One important idea is borrowed from the proof of the complexity-theoretic result "BPP $\subseteq \Sigma_2$" (see e.g. [1]), and makes it possible to "derandomize" a probabilistic sampling procedure. For technical reasons, Theorem 5 actually provides a Σ_3 formula. This section can therefore be considered as a continuation of [5], where a real-number-model analogue of "BPP $\subseteq$ P/poly" was established. In Theorem 4, we also give a counterexample which shows that a key step in our proof fails if the VC dimension hypothesis is removed. Finally, section 6 provides bounds that make it possible to apply Theorem 5 to a variety of concrete examples: sets that are defined by systems of inequalities involving polynomial functions, or polynomial and exponential functions. In particular, we generalize the polynomial VC dimension bounds of [8] to a larger class of circuits made of arithmetic, exponentiation, and sign gates.

In section 2, we introduce some notations and recall a few results that will be used throughout the paper.

2 Preliminaries

Semi-algebraic sets are a natural class of sets to which our results apply. They are defined as follows.

Definition 1 *A set $S \subseteq \mathbb{R}^n$ is semi-algebraic if it can be defined by a boolean formula $B(P_1, \ldots, P_m)$, where the atomic predicates P_i are of the from "$p_i(x) = 0$" or "$p_i(x) > 0$" and $p_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is a multivariate polynomial.*

The Chernoff bounds are a standard tool in the analysis of probabilistic algorithms. In this paper we use Hoeffding's inequality instead (see Appendix B of [12] for a proof).

Theorem 1 (Hoeffding) *Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with zero means and bounded ranges: $a_i \leq Y_i \leq b_i$. For each $\eta > 0$,*

$$\Pr\{|Y_1 + \ldots + Y_n| \geq \eta\} \leq 2\exp\left[-2\eta^2 / \sum_{i=1}^{n}(b_i - a_i)^2\right].$$

In this paper we work only with i.i.d Bernouilli random variables, so we only need the following special case.

Corollary 1 *Let $X_1, \ldots, X_n$ be i.i.d random variables with range $[0,1]$ and mean EX. For each $\epsilon > 0$,*

$$\Pr\{|\frac{X_1 + \ldots + X_n}{n} - EX| \geq \epsilon\} \leq 2e^{-2n\epsilon^2}.$$

Proof. Apply Theorem 1 to $Y_i = X_i - EX$. □

One advantage of this bound is that it applies to arbitrary (unbounded) values of ϵ.

The remainder of this section can be skipped by readers interested only in the Monte-Carlo integration results (sections 3 and 5). The Vapnik-Chervonenkis dimension of a family of sets is defined as follows.

Definition 2 *Let $\mathcal{F}$ be a class of subsets of a set X. We say that $\mathcal{F}$* shatters *a set $A \subseteq X$ if for every subset $E \subseteq A$, there exists $S \in \mathcal{F}$ such that $E = S \cap A$. The VC dimension of $\mathcal{F}$ is the cardinality of the largest set that is shattered by $\mathcal{F}$.*

Sauer's Lemma is a key property of classes of finite VC dimension. For a proof and additional references, see for instance [4].

Lemma 1 (Sauer) *Let $F \subseteq 2^X$ be a class of finite VC dimension d. For any set $A \subseteq X$ of cardinality $m \geq d$, there are at most $(em/d)^d$ distinct sets of the form $S \cap A$ ($S \in F$).*

The following result is from [4]. Similar bounds have been established by other authors, including Vapnik and Chervonenkis [15].

Theorem 2 *Let $\mathcal{F}$ be a family of subsets of a set X, and let P be an arbitrary probability distribution on X. For $S \in \mathcal{F}$, let $P(S) = \Pr\{x \in S\}$. Let us draw m points $x_1, \ldots, x_m$ independently from P. Assume that $\mathcal{F}$ has finite VC dimension d. If the sample size m is at least*[2]

$$\max\left(\frac{4}{\epsilon}\log\frac{2}{\delta}; \frac{8d}{\epsilon}\log\frac{13}{\epsilon}\right)$$

then

$$\Pr\{(x_1, \ldots, x_m); \exists S \in \mathcal{F} \mid P(S) \geq \epsilon \text{ and } x_1 \notin S, \ldots, x_m \notin S\} \leq \delta.$$

As is often the case in computational geometry, the only property of VC classes that will actually be used is the existence of "small" ϵ-nets (see [11] for optimal constants).

Corollary 2 *With the notations of Theorem 2, if $m \geq \frac{8d}{\epsilon}\log\frac{13}{\epsilon}$ there exists a set $\{x_1, \ldots, x_m\} \subseteq X$ (called an ϵ-net) such that $S \cap \{x_1, \ldots, x_m\} \neq \emptyset$ for every $S \in \mathcal{F}$ such that $P(S) \geq \epsilon$.*

Proof. Apply Theorem 2 with, e.g., $\delta = 1/2$. □

For $x \in \mathbb{R}$, $y = x \bmod 1$ denotes the unique number $y \in [0, 1[$ such that $x - y \in \mathbb{Z}$ (equivalently, $y = x - \lfloor x \rfloor$).

Given $t \in I^n$, the map $\phi_t : I^n \to I^n$ is defined as follows: $\phi_t(x) = x \oplus t = (x + t) \bmod 1$ (the "mod 1" operation is applied componentwise).

Note that ϕ_t is a measure-preserving map, and that I^n endowed with the $\oplus$ operation is (isomorphic to) the additive group $(\mathbb{R}/\mathbb{Z})^n$. Subtraction in this group will be denoted by $\ominus$. For $x \in I^n$ and $S \subseteq I^n$, we shall also use the notation $x * S = \{x * y; y \in S\}$, where $*$ stands for $\oplus$ or $\ominus$.

3 How fine grids should be

As explained in the introduction, $\kappa(E)$ is the maximum number of connected components of the intersection $E \cap L$ of E with any axis-parallel line L. In particular, in dimension 1, $\kappa(E)$ is the number of connected components of E. The proof of Theorem 3 is by induction on the dimension of the space. We start with the base case $p = 1$. Throughout the paper, I is the unit interval $[0, 1]$.

Lemma 2 *Let $E \subseteq I$ be a measurable set with $\kappa(E)$ finite: $|\mu_h(E) - \mu(E)| \leq h\kappa(E)$.*

Proof. The set E is the union of $\kappa(E)$ disjoint intervals $I_1, \ldots, I_{\kappa(E)}$. It is sufficient to consider the case where E is an interval since $|\mu_h(E) - \mu(E)| \leq \sum_{j=1}^{\kappa(E)} |\mu_h(I_j) - \mu(I_j)|$.

If $E \cap \mathcal{V} = \emptyset$, E is included between two consecutive points of $\mathcal{V}$. Hence $0 \leq \mu(E) \leq h$ and $\mu_h(E) = 0$. Otherwise, let $\{hi, h(i+1), \ldots, hj\} = \mathcal{V} \cap E$: $\mu_h(E) = (j - i + 1)h$. Since $E \subseteq]h(i-1), h(j+1)[$, $\mu(E) \leq (j - i + 2)h$. Hence $\mu(E) \leq \mu_h(E) + h$. Since $[hi, hj] \subseteq E$, $\mu(E) \geq (j - i)h = \mu_h(E) - h$. □

Theorem 3 *Let $E \subseteq I^p$ be a measurable set with $\kappa(E)$ finite: $|\mu_h(E) - \mu(E)| \leq ph\kappa(E)$.*

Proof. Assume that $p \geq 2$ and that the result is true in dimension $p - 1$. Let f be the characteristic function of E. By definition, the difference between the discrete measure and the Lebesgue measure is

$$|\mu_h(E) - \mu(E)| = \left| h^p \sum_{i_1, \ldots, i_p = 0}^{N-1} f(hi_1, \ldots, hi_p) - \int_{I^p} f(x)dx \right|.$$

This difference can be bounded by A+B, where

$$A = \left| \int_{I^p} f(x)dx - h \sum_{i=0}^{N-1} \int_{I^{p-1}} f(hi, x_2, \ldots, x_p)dx_2 \ldots dx_p \right|$$

and

$$B = \left| h \sum_{i=0}^{N-1} \int_{I^{p-1}} f(hi, y)dy - h^p \sum_{i=0}^{N-1} \sum_{i_2, \ldots, i_p = 0}^{N-1} f(hi, hi_2, \ldots, hi_p) \right|$$

Here, y stands for $x_2, \ldots, x_p$. The second term can be bounded by

$$h \sum_{i=0}^{N-1} \left| \int_{I^{p-1}} f(hi, y)dy - h^{p-1} \sum_{i_2, \ldots, i_p = 0}^{N-1} f(hi, hi_2, \ldots, hi_p) \right|.$$

For any i, the intersection $F = E \cap \{x_1 = hi\}$ of E with the hyperplane $\{x_1 = hi\}$ satisfies $\kappa(F) \leq \kappa(E)$. Hence by induction hypothesis, $B \leq (hN)[(p-1)h\kappa(F)] \leq (p-1)h\kappa(E)$. By inverting the summation and integration symbols, the first term can be bounded as follows:

$$\begin{aligned} A &= \left| \int_{I^{p-1}} \left[\int_0^1 f(x_1, y)dx_1 - h \sum_{i=0}^{N-1} f(hi, y) \right] dy \right| \\ &\leq \sup_{y \in I^{p-1}} \left| \int_0^1 f(x_1, y)dx_1 - h \sum_{i=0}^{N-1} f(hi, y) \right|. \end{aligned}$$

[2] log denotes the logarithm to base 2 and ln the natural logarithm.

For any $(\lambda_2,\ldots,\lambda_p) \in I^{p-1}$, the intersection $G = E \cap \{x_2 = \lambda_2,\ldots,x_p = \lambda_p\}$ of E with the line $\{x_2 = \lambda_2,\ldots,x_p = \lambda_p\}$ satisfies $\kappa(G) \leq \kappa(E)$. Hence by Lemma 2, $A \leq h\kappa(E)$. Therefore $|\mu_h(E) - \mu(E)| \leq h\kappa(E) + h(p-1)\kappa(E) = hp\kappa(E)$. □

4 Defining the volume of definable sets

We shall work in a first-order theory of the reals which includes at least addition, subtraction and order. Other function symbols are also allowed. For instance, in section 6 we consider theories that include multiplication, or multiplication and exponentiation. We would like to construct formulas that approximate the volume of a set E which is definable in this theory. This is made precise in Definition 3. The main assumption on E, introduced in Lemma 6, is that the family $\{x \ominus E; x \in I^p\}$ has finite VC dimension.

Definition 3 *Let E be subset of $\mathbb{R}^p$, and let $F(v)$ be a formula where the free variable v lives in $\mathbb{R}$. We say that $F(v)$ defines an ϵ-approximate volume when the two following properties hold:*

1. *If $F(v)$ is satisfied then $|v - \mu(E)| < \epsilon$.*
2. *If $|v - \mu(E)| \leq \epsilon/4$ then $F(v)$ is satisfied.*

One might be tempted to ask in this definition that $F(v)$ be satisfied if and only if $|v-\mu(E)| < \epsilon$. This does not work in general for the same algebraicity problem as explained in the introduction. Indeed, for the unit disk the set of satisfying v's should be $]\pi - \epsilon, \pi + \epsilon[$. This is impossible if ϵ is rational, or even algebraic.

Note that when F is decidable, one can compute an ϵ-approximate volume $v \in V = \{0, \epsilon/2, \epsilon, 3\epsilon/2, \ldots, \lfloor 2/\epsilon \rfloor \epsilon/2, 1\}$ by deciding whether the $2 + \lfloor 2/\epsilon \rfloor$ formulas $F(v)$ $(v \in V)$ are satisfied. Indeed, by property 2, at least one of these formulas must be satisfied. By property 1, any such satisfied formula provides an ϵ-approximate volume. For instance, it follows from Theorem 5 and the results of [2] or [13] that an ϵ-approximate volume can be computed in polynomial space when E is a semi-algebraic set defined by a system of polynomial (in)equations with rational coefficients (see section 6 for more details on the case of semi-algebraic sets). The bounds in the next section imply that it is much more efficient to use the randomized polynomial-time algorithm of section 3.

The basic ideas of the construction are roughly that:

- the set of random samples that provide an accurate volume estimate is "large" (Lemma 5, property 2);
- the set of random samples that provide a bad volume estimate is "small" (Lemma 5, property 1);
- A set of random samples is "large" if and only if a "small" number of its translations can cover the whole space (Lemmas 3 and 4).

Lemma 3 *For any $S \subseteq I^n$ and any sequence $t_1,\ldots,t_m$ of points of I^n such that $m < 1/\mu(S)$, $\bigcup_{i=1}^{m} S \oplus t_i \neq I^n$.*

Proof. Obvious since the ϕ_{t_i}'s are measure-preserving. □

Lemma 4 *Let $S \subseteq I^n$ be such that the family of sets $\mathcal{F} = \{x \ominus S; x \in I^n\}$ has finite VC dimension d. There are points $t_1,\ldots,t_m \in I^n$ such that $\bigcup_{i=1}^{m} S \oplus t_i = I^n$ as soon as:*

$$m \geq \frac{8d}{\mu(S)} \log \frac{13}{\mu(S)}.$$

Proof. The condition $\bigcup_{i=1}^{m} S \oplus t_i = I^n$ is equivalent to $F \cap \{t_1,\ldots,t_m\} \neq \emptyset$ for every $F \in \mathcal{F}$. Thus the result follows from Corollary 2 since $\mu(F) = \mu(S)$ for every $F \in \mathcal{F}$. □

One can ask whether a VC dimension hypothesis is really necessary here. It turns out that positive measure alone is not a sufficient condition to obtain a result similar to Lemma 4. In fact, one can prove the following statement.

Theorem 4 *Let $S \subseteq I$ be a set whose complement in I is open dense (for instance, take S to be a fat Cantor set of positive measure). I cannot be covered by a countable union of sets of the form $x \ominus S$ $(x \in I)$.*

Proof. The complement of a set of the form $x \ominus S$ is also open dense. Hence by the Baire category theorem, the complement of a countable union of sets of the form $x \ominus S$ is dense, and, in particular, is not empty. □

Lemma 5 *Let $E \subseteq I^p$. Given $k \in \mathbb{N}$ and $v, \alpha \in I$, let*

$$S_{v,\alpha} = \left\{ x \in (I^p)^k; \left| \frac{1}{k} \sum_{i=1}^{k} 1_E(x_i) - v \right| \leq \alpha \right\} \qquad (2)$$

(here, x stands for $(x_1,\ldots,x_k)$.) The two following properties hold:

1. *if* $\mu(S_{v,\epsilon/2}) > 2e^{-k\epsilon^2/2}$ *then* $|v - \mu(E)| < \epsilon$;

2. *if* $|v - \mu(E)| \leq \epsilon/4$ *then* $\mu(S_{v,\epsilon/2}) \geq 1 - 2e^{-k\epsilon^2/8}$.

Proof. For $\alpha \in I$, let

$$S_\alpha = \left\{ x \in (I^p)^k; \left| \frac{1}{k} \sum_{i=1}^{k} 1_E(x_i) - \mu(E) \right| < \alpha \right\}.$$

By Hoeffding's inequality, $\mu(S_\alpha) \geq 1 - 2e^{-2k\alpha^2}$.

For property 1, assume that $|v - \mu(E)| \geq \epsilon$. This implies that $S_{v,\epsilon/2} \cap S_{\epsilon/2} = \emptyset$, hence $\mu(S_{v,\epsilon/2}) \leq 2e^{-k\epsilon^2/2}$. For property 2, if $|v - \mu(E)| \leq \epsilon/4$ then $S_{\epsilon/4} \subseteq S_{v,\epsilon/2}$, whence the result. □

Corollary 3 *With the notations of Lemma 5, assume that the the VC dimension of the family of sets* $\mathcal{F} = \{x \ominus S_{v,\epsilon/2}; x \in I^{kp}\}$ *is bounded by* d. *Consider the property* (C_m) *defined as follows: "there exist* m *points* $t_1, \ldots, t_m \in I^{kp}$ *such that* $I^{kp} = \bigcup_{i=1}^{m} S_{v,\epsilon/2} \oplus t_i$*".*

If the conditions

$$m \geq (16 \log 26) d \qquad (3)$$

and

$$k > \max(2 \ln 2m, 8 \ln 4)/\epsilon^2 \qquad (4)$$

are satisfied, the two following properties hold:

1. *If* (C_m) *holds then* $|v - \mu(E)| < \epsilon$.

2. *If* $|v - \mu(E)| \leq \epsilon/4$ *then* (C_m) *holds.*

Proof. If (C_m) holds then $\mu(S_{v,\epsilon/2}) \geq 1/m$ by Lemma 3. Hence $\mu(S_{v,\epsilon/2}) > 2e^{-k\epsilon^2/2}$ by (4). This implies that $|v - \mu(E)| < \epsilon$ by Lemma 5.

If $|v - \mu(E)| \leq \epsilon/4$ then $\mu(S_{v,\epsilon/2}) \geq 1 - 2e^{-k\epsilon^2/8}$ by Lemma 5. Hence $\mu(S_{v,\epsilon/2}) \geq 1/2$ by (4). This implies (C_m) by Lemma 4. □

Note that (C_m) can be informally expressed as a Σ_2 formula , i.e., there is one block of existential quantifiers followed by one block of universal quantifiers):

$$\exists t_1, \ldots, t_m \in I^{kp}\ \forall x \in I^{kp} \bigvee_{i=1}^{m} [(x \ominus t_i) \in S_{v,\epsilon/2}]. \quad (5)$$

In order to apply the previous result, we now relate the VC dimension of $\mathcal{F}$ to that of the family $\mathcal{F}' = \{x \ominus E; x \in I^p\}$.

Lemma 6 *Let* d' *be the VC dimension of the family* $\mathcal{F}' = \{x \ominus E; x \in I^p\}$. *Then for any* $v, \alpha \in I$, *the VC dimension* d *of the family*

$$\mathcal{F} = \{x \ominus S_{v,\alpha}; y \in I^{kp}\}$$

satisfies $d \leq Cd'k \log k$, *where* C *is a universal constant.*

Proof. Given $x \in I^{kp}$, a point $y = (y_1, \ldots, y_k) \in I^{kp}$ is in $x \ominus S_{v,\alpha}$ if and only if $x \ominus y \in S_{v,\alpha}$. This means that:

$$\left| \frac{1}{k} \sum_{i=1}^{k} 1_E(x_i \ominus y_i) - v \right| \leq \alpha$$

or equivalently,

$$\left| \frac{1}{k} \sum_{i=1}^{k} 1_{x_i \ominus E}(y_i) - v \right| \leq \alpha.$$

By Sauer's lemma, for each i the number of subsets of the form $x_i \ominus E$ that can be induced on a set of I^p of cardinality $m \geq d'$ is at most $(em/d')^{d'}$. Hence the number of subsets induced by $\mathcal{F}$ on a set $X \subseteq I^{kp}$ of cardinality $m \geq d'$ is at most $(em/d')^{d'k}$. If X is shattered, $(em/d')^{d'k}$ must be at least 2^m, and thus $m \leq kd' \log e + kd' \log(m/d')$. This implies that $m/2 \leq kd' \log e$ or $m/2 \leq kd' \log(m/d')$. In the first case, we have a linear upper bound on m/d', which is good enough for our purpose. In the second case, $(m/d')/\log(m/d') \leq 2k$. This implies that $m/d' = O(k \log k)$. □

Theorem 5 *Let* $E \subseteq I^p$ *be such that the family* $\mathcal{F}' = \{x \ominus E; x \in I^p\}$ *has finite VC dimension* d'. *Then the following polynomial-size formula* $F(v)$ *defines an* ϵ*-approximate volume as soon as* $k \geq (C_1/\epsilon^2) \log(d'/\epsilon)$ *and* $m = \lceil C_2 d' k \log k \rceil$ *(*C_1 *and* C_2 *are universal constants):*

$$\exists t_1, \ldots, t_m \in I^{kp}\ \forall x \in I^{kp} \exists y, z \in I^{kp} \bigvee_{i=1}^{m} F_i$$

where F_i *stands for:*

$$(y = x \ominus t_i) \wedge \bigwedge_{j=1}^{k} (z_j = 1_E(y_j)) \wedge \left| \frac{1}{k} \sum_{j=1}^{k} z_j - v \right| \leq \epsilon/2.$$

In this theorem, the notation $y = x \ominus t$ is an abbreviation for

$$\bigwedge_{j=1}^{k} [(x_j - t_j \geq 0 \wedge y_j = x_j - t_j) \vee (x_j - t_j < 0 \wedge y_j = x_j - t_j + 1)]$$

and $z = 1_E(y)$ is an abbreviation for

$$(y_j \in E \wedge z_j = 1) \vee (y_j \notin E \wedge z_j = 0).$$

Of course, the predicate $|X| \leq \epsilon/2$ can be written as $(X \leq \epsilon/2) \wedge (X \geq -\epsilon/2)$. Note that if we have in our first-order language a function symbol for the operator $\ominus$ on $\mathbb{R}$ and for the characteristic function of E, we can get rid of the quantified variables y and z, and $F(v)$ can be replaced by (5). Note also that if we can use quantified circuits instead of quantified formulas, these variables are again unnecessary.[3]

Proof of Theorem 5. Let us show that conditions (3) and (4) can be satisfied with the the choices of k and m made in this theorem. By Lemma 6, condition (3) will be satisfied for $C_2 = (16 \log 26)C$. Taking (3) into account, condition (4), is asymptotically equivalent to: $k > 2\ln(2\lceil C_2 d' k \log k \rceil)/\epsilon^2$. Since $k > \log k$ for $k \geq 2$, it suffices that $k \geq 2\ln(4C_2 d')/\epsilon^2 + (4 \ln k)/\epsilon^2$. This new condition will be satisfied if

$$k/2 \geq 2\ln(4C_2 d')/\epsilon^2 \qquad (6)$$

and $k/2 \geq (4 \ln k)/\epsilon^2$. The latter constraint will be satisfied if

$$k \geq (C'/\epsilon^2)\ln(1/\epsilon^2), \qquad (7)$$

for some constant C'. The result follows from (6) and (7). □

5 Bounds on $\kappa(E)$

We start with a bound on the parameter $\kappa(E)$ of section 3 for semi-algebraic sets.

Theorem 6 *Let $E \subseteq \mathbb{R}^p$ be defined by a boolean formula containing s distinct atomic predicates, where each predicate is a polynomial equality or inequality of degree at most D: $\kappa(E) \leq Ds + 1$.*

Proof. The intersection of E with any line L can again be defined by s polynomial (in)equalities in one variable of degree at most D. Together these polynomials have at most Ds roots. The complement of these roots is a union of at most $Ds + 1$ intervals. Each root and each interval is either included in E or included in its complement. The maximum number of connected components is therefore obtained by alternating inclusion in E and $\complement_E$. □

Since we have to consider axis-parallel line only, in this theorem D can be replaced by the maximum degree with respect to any variable (for instance, the maximum degree with respect to any variable of $P(x, y) = x^2 y$ is 2).

Together with Theorem 3, this result has an interesting application to the study of randomized algorithms in the model of computation over the reals introduced by Blum, Shub & Smale [3]. There are several ways to introduce randomization in this model. Perhaps the two simplest are *discrete randomization*, where random elements are drawn from the uniform distribution on $\{0, 1\}$, just as on an ordinary Turing machine; and *continuous (uniform) randomization*, where random elements are drawn from the uniform distribution on $[0, 1]$. As usual, an input is considered accepted (resp. rejected) if the probability of acceptance (resp. rejection) is at least, say, 2/3. In the terminology of this paper, this means that the volume of the set of random elements leading to acceptation (or rejection) should be at least 2/3. According to Theorems 3 and 6, this volume can be efficiently estimated by a discrete Monte-Carlo method. Hence at the polynomial-time level, continous randomization is not more powerful than discrete randomization (the details will appear in a forthcoming book by Blum, Cucker, Shub & Smale). The converse is also true since one can test whether a given element $x \in [0, 1]$ is smaller or larger than 1/2.

One can give exponential upper bounds on $\kappa(E)$ for sets whose definition involves the exponential function. This is a rather straightforward consequence of the results of Khovanskii [9] also used (in a much more sophisticated way) by Karpinski & Macintyre to obtain VC dimension bounds (see Theorem 8 in the next section). We do not give more details here because these bounds seem to be gross overestimates. There is actually an exponential gap between known upper and lower bounds. A sharp bound is available in the special case of exponential polynomials, i.e., functions of the form $f(x) = \sum_{i=1}^{n} \alpha_i e^{\lambda_i x}$. It is well-known (and easily provable by induction on n) that f cannot have more than $n - 1$ zeros. This bound is definitely not exponential. Sigmoidal neural networks are another special case of interest. Here one does not have a sharp bound, even for networks with a single hidden layer (these devices compute functions of the form $f(x) = \sum_{i=1}^{n} a_i \sigma(w_i x - \theta_i)$, where $\sigma(x) = 1/(1+e^{-x})$). In this case one can also obtain an upper bound by elementary methods (reduce to the same denominator and apply the result for exponential polynomials); unfortunately, this bound remains exponential. In con-

[3] A formula is a circuit with fan-out 1; a circuit can always be transformed into a formula, but this can cause an exponential blowup in size.

trast, the best lower bound currently known is only linear in n [14]. We conclude this section on a positive note: our exponential upper bounds may be far from optimal, but the number of digits required by Theorem 3 for Monte-Carlo integration is only logarithmic in $\kappa(E)$, so it remains polynomial (it is only logarithmic for semi-algebraic sets).

6 VC dimension bounds

In order to apply the results of section 4 to a semi-algebraic set, it is necessary to bound the VC dimension of families of semi-algebraic sets. The following result is from [7].

Theorem 7 (Goldberg-Jerrum) *Let $\Phi(x,y)$ be a boolean formula containing s distinct atomic predicates, where each predicate is a polynomial equality or inequality over $n+k$ variables (representing $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^k$, respectively) of degree at most D. For any $y \in \mathbb{R}^k$, let $F_y \subseteq \mathbb{R}^n$ be the set of instances x such that $\Phi(x,y)$ is satisfied. The family $\{F_y; y \in \mathbb{R}^k\}$ has VC dimension at most $2k\log(8eDs)$.*

Let $E \subseteq I^p$ be a semi-algebraic set defined by a boolean formula involving s atomic predicates $P_1, \ldots, P_s$ that are polynomial equalities or inequalities of degree at most D. By the following lemma, the VC dimension of the set $\mathcal{F}' = \{x' \ominus E; x' \in I^p\}$ is $d' \leq 2p(p + \log(16eDs))$. This bound on d' can be plugged directly into Theorem 5.

Lemma 7 *Let $E \subseteq I^p$ be a semi-algebraic set that can be defined by a boolean formula with s distinct atomic predicates, where each predicate is a polynomial equality or inequality of degree at most D. The family $\mathcal{F}' = \{x \ominus E; x \in I^p\}$ has VC dimension $d' \leq 2p(p + \log(16eDs))$.*

Proof. A point $t \in I^p$ is in $x \ominus E$ if and only if $x \ominus t \in E$. Each component $(x \ominus t)_i$ of $x \ominus t$ is either equal to $x_i - t_i$ or $x_i - t_i + 1$ depending on the sign of $x_i - t_i$. The condition $x \ominus t \in E$ can thus be expressed as a boolean formula with $p + 2^p s$ distinct atomic predicates (the first p predicates are of the form "$x_i - t_i \geq 0$"; the $2^p s$ other predicates are obtained from those of E by considering the 2^p possible outcomes of the tests "$x_i - t_i \geq 0$"). These predicates are of degree at most D, hence $d' \leq 2p\log(8eD(p + 2^p s)) \leq 2p(p + \log(16eDs))$ by Theorem 7. □

It is possible to deal with sets whose definition involves the exponential function thanks to the recent work of Karpinski and Macintyre [8]. One of their main result is as follows.

Theorem 8 (Karpinski-Macintyre) *Let $\Phi(x,y)$ be a boolean formula containing s distinct atomic predicates, where each predicate is an equality or inequality over $n+k$ variables (representing $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^k$, respectively). We assume that the terms in Φ are polynomials of degree at most D in x, y and no more than than q subterms $\exp(g(x,y))$, where g is linear. For any $y \in \mathbb{R}^k$, let $F_y \subseteq \mathbb{R}^n$ be the set of instances x such that $\Phi(x,y)$ is satisfied. The family $\{F_y; y \in \mathbb{R}^k\}$ has VC dimension at most $q(q-1)/2 + q\log(2k+2) + k(\log 3 + \log(2k+2) + 17\log s)$.*

By the same argument as for Lemma 7, one can give an upper bound of $q(q-1)/2 + q\log(2k+2) + k(\log 3 + \log(2k+2) + 17\log(p + 2^p s))$ on the VC dimension of the family $\mathcal{F}' = \{x \ominus E; x \in I^p\}$ for a set E which is defined by a boolean formula containing s atomic predicates, where each predicate is a polynomial equality or inequality of degree at most D over $x \in \mathbb{R}^p$ and no more than q linear-exponential subterms of the form $\exp(g(x))$.

It is also possible to give a polynomial VC dimension bound for circuits (rather than just formulas) over the reals made of sign, addition, multiplication, and exponentiation gates. Such a bound can already be found in [8] for a restricted class of circuits called *sigmoidal neural networks*. In the remainder of this section, we sketch a proof of this generalization.

One remark in [8] states that a polynomial bound applies to circuits made of other gates functions than the standard sigmoid $\sigma(x) = 1/(1+e^{-x})$: any Pfaffian function can be used, including of course the identity and exponential functions (these two cases are actually simpler than the standard-sigmoidal case). Also, gates that compute a polynomial function of their inputs before applying σ can be allowed. This makes it possible to deal with circuit of addtion, multiplication and exponentiation gates. Unfortunately, the Karpinski-Macintyre theory does not apply directly with the sign function s since it is not C^∞. However, for any finite set S of inputs, all sign gates in a circuit can be replaced by standard-sigmoidal gates since

$$\lim_{\gamma \to +\infty} \frac{1}{1 + e^{-\gamma x}} = s(x)$$

for any $x \in \mathbb{R}$. (before applying this transformation, one has to make sure that for all inputs $x \in S$, the input to all sign gates in the circuit is non-zero; this condition can be enforced by adding a threshold to the sign gates. A similar construction is detailed in [10].)

Thus one can conclude by applying the VC dimension bound for circuits made of addition, multiplication, exponentiation and standard-sigmoidal gates.

It is not much harder to include the inverse function $i(x) = 1/x$ since it can be simulated in the same sense by the C^∞ function $i_\gamma(x) = (\tanh \gamma x)^2/x$ when γ is large enough (note that $i_\gamma(0) = 0$). Then one can check that the proof technique of [8] for sigmoidal networks applies to circuits using i_γ-gates with a few straightforward modifications.

7 Final Remarks

It is possible to construct volume-defining formulas as in section 4 with $\kappa(E)$ used as complexity parameter instead of the VC dimension. By Theorem 3 we just have to "derandomize" a coin-tossing approximation algorithm. This can be done as in the standard proof that BPP $\subseteq \Sigma_2$. Quantification in the resulting formulas is over boolean rather than real variables.

Finitiness of the Vapnik-Chervonenkis dimension is equivalent (for boolean functions) to the existence of a distribution-free uniform law of large numbers. However, in section 4 we worked only with the uniform distribution on the unit cube. The VC dimension hypothesis can thus be replaced by a less restrictive metric entropy hypothesis. It would be interesting to find out whether for natural (e.g., semi-algebraic) sets one can give better bounds on the metric entropy than those that follow from the VC dimension bounds.

Acknowledgments

Eduardo Sontag suggested a simplification in the proof of the VC dimension bound for circuits using sign and division gates. An anonymous referee pointed out that my application of the Chernoff bounds was erroneous (the present paper is hopefully correct).

References

[1] J.L. Balcázar, J. Diáz, and J. Gabarró. *Structural Complexity I.* EATCS Monographs on Theoretical Computer Science. Springer-Verlag, 1988.

[2] S. Basu, R. Pollack, and M.-F. Roy. On the combinatorial and algebraic complexity of quantifier elimination. In *Proc. 35th IEEE Symposium on Foundations of Computer Science*, pages 632–641, 1994.

[3] L. Blum, M. Shub, and S. Smale. On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines. *Bulletin of the American Mathematical Society*, 21(1):1–46, July 1989.

[4] A. Blumer, A Ehrenfeucht, D. Haussler, and M. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1990.

[5] F. Cucker, M. Karpinski, P. Koiran, T. Lickteig, and K. Werther. On real Turing machines that toss coins. In *Proc. 27th ACM Symposium on Theory of Computing*, pages 335–342, 1995.

[6] M. Dyer, A. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM*, 38(1):1–17, 1991.

[7] P. Goldberg and M. Jerrum. Bounding the Vapnik-Chervonenkis dimension of concept classes parameterized by real numbers. *Machine Learning*, 18:131–148, 1995.

[8] M. Karpinski and A. Macintyre. Polynomial bounds for VC dimension of sigmoidal neural networks. In *Proc. 27th ACM Symposium on Theory of Computing*, pages 200–208, 1995.

[9] A. G. Khovanskii. *Fewnomials*, volume 88 of *Translations of Mathematical Monographs.* American Mathematical Society, 1991.

[10] P. Koiran and E. D. Sontag. Neural networks with quadratic VC dimension. NeuroCOLT Technical Report 95-044, 1995. this report is available online: `http://www.dcs.rhbnc.ac.uk/neurocolt.html`.

[11] J. Komlós, J. Pach, and G. Wöginger. Almost tight bound for ϵ-nets. *Discrete & Computational Geometry*, 7:163–173, 1992.

[12] D. Pollard. *Convergence of Stochastic Processes.* Springer Verlag, 1984.

[13] J. Renegar. On the computational complexity and geometry of the first-order theory of the reals. parts I, II, III. *Journal of Symbolic Computation*, 13(3):255–352, March 1992.

[14] E. D. Sontag. Feedforward nets for interpolation and classification. *Journal of Computer and System Sciences*, 45:20–48, 1992.

[15] V. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theoret. Prob. and its Appl.*, 16(2):264–280, 1971.