Decision Region Approximation by Polynomials or Neural Networks

Kim L. Blackmore, Robert C. Williamson, *Member, IEEE*, and Iven M. Y. Mareels, *Senior Member, IEEE*

***Abstract*—We give degree of approximation results for decision regions which are defined by polynomial and neural network parametrizations. The volume of the misclassified region is used to measure the approximation error, and results for the degree of L_1 approximation of functions are used. For polynomial parametrizations, we show that the degree of approximation is at least 1, whereas for neural network parametrizations we prove the slightly weaker result that the degree of approximation is at least r, where r can be any number in the open interval $(0, 1)$.**

***Index Terms*—Classification, decision region, neural networks, polynomials, rate of approximation.**

I. Introduction

DECISION regions arise in machine learning problems of sorting or classification of data [1]. Points contained in the decision region are positively classified, and points outside the decision region are negatively classified. For a decision region $D \subset \mathbb{R}^n$, this classification can be described by the discriminant function

$$y_D(x) = \begin{cases} 1, & \text{if } x \in D \\ -1, & \text{otherwise.} \end{cases} \tag{1}$$

The learning task is to use examples of classified points to be able to correctly classify all possible points.

In neural network learning, decision boundaries are often represented as zero sets of certain functions, with points contained in the decision region yielding positive values of the function, and points outside the decision region yielding negative values [2]. In this case, the learning task is to use examples of correctly classified points to identify a parameter $a \in \mathbb{R}^m$ for which the set $\{x\colon f(a, x) \geq 0\}$, called the *positive domain of* $f(a, \cdot)$, matches the true decision region.

For the purposes of analyzing a learning algorithm, it is useful to assume that a suitable value of the parameter exists. However, there is no general reason why such an assumption is satisfied in practice. Even if there is a class of functions $f(\cdot, \cdot)$ and a parameter a such that the positive domain of $f(a, \cdot)$ matches the true decision region, there is usually no way of identifying this class *a priori*. It is therefore useful to know how well particular classes of functions can approximate decision regions with prescribed general properties. In particular, it is important to know how fast the approximation error decreases as the approximating class becomes more complicated—e.g., as the degree of a polynomial or the number of nodes of a neural network increases.

The question of approximation of functions has been widely studied. The classical Weierstrass Theorem showed that polynomials are universal approximators [3] (in the sense that they are dense in the space of continuous functions on an interval). Many other classes have been shown to be universal approximators, including those defined by neural networks [4]. Degree of approximation results tell the user how complicated a class of approximating functions must be in order to guarantee a certain degree of accuracy of the best approximation. The classical Jackson Theorem [5] is the first example of this. Hornik [6], Barron [7], Mhaskar and Michelli [8], [9], Mhaskar [10], Darken *et al*. [11], and Hornik *et al*. [12] give degree of approximation results for neural networks.

The problem of approximating sets, rather than functions, has received some attention in the literature. Approximation of (unparametrized) sets and curves has been studied for pattern recognition and computer vision purposes [13]–[15]. The approach is quite different from the approach here. Theoretical work can be grouped according to two basic approaches—namely, explicit and implicit parametrizations. "Explicit parametrization" refers to frameworks where the decision *boundary* is parametrized. For example, if the decision region is a set in $\mathbb{R}^n$, the decision boundary might be considered the graph of a function on $\mathbb{R}^{n-1}$, or a combination of such graphs. "Implicit parametrization" refers to frameworks (as used in this work) where the decision *region* is the positive domain of some function.

Most existing work is in terms of explicit parametrizations [16]. For instance, Korostelev and Tsybakov [17], [18] consider the *estimation* (from sample data) of decision regions. Although they consider nonparametric estimation, it is, in fact, the explicit rather than implicit framework as defined above (they reduce the problem to estimating functions whose graphs make up parts of the decision boundary). In a similar vein, Dudley [19] and Shchebrina [20] have determined the metric entropy of certain smooth curves.

Regarding the implicit problem, Mhaskar [10] gives a universal approximation type result for approximation by positive domains of certain neural network functions. Ivanov [21] summarizes many problems in algebraic geometry con-

Manuscript received November 15, 1994; revised May 27, 1996. This work was supported by the Australian Research Council. The material in this paper was presented in part at the Australian Conference on Neural Networks, 1994.

K. L. Blackmore is with Communications Division, DSTO, P.O. Box 1500, Salisbury SA 5108, Australia.

R. C. Williamson is with the Department of Engineering, Australian National University, Canberra ACT 0200, Australia.

I. M. Y. Mareels was with the Department of Engineering, Australian National University, Canberra ACT 0200, Australia. He is now with the Department of Electrical Engineering, The University of Melbourne, Melbourne, Australia.

Publisher Item Identifier S 0018-9448(97)02634-5.

cerned with the question of when a smooth manifold can be approximated by a real algebraic set but does not address the degree of approximation question. In work similar to that described in [21], Broglia and Tognoli [22] consider when a C^∞ function can be approximated by certain classes of functions without changing the positive domain.

In this paper, we use function approximation results to determine the degree of approximation of decision regions by positive domains of polynomial functions and neural networks. We consider the L_1 approximation of the discriminant function $y_D(x)$. This implies a bound on the L_1 distance between $y_D(x)$ and $\operatorname{sgn}(f(x))$, where $f(x)$ is the approximating polynomial or neural network function. We use a result from differential geometry to link this distance with the size of the misclassified volume. Since most learning problems can be analyzed probabilistically, the volume of the misclassified region has a natural interpretation as the probability of misclassification by the approximate decision region when the data are drawn from a uniform distribution over the input space.

The next section of this paper contains a formal statement of the degree of approximation problem for decision regions. In Section III we define a corridor around the decision boundary, and give a result concerning its volume, which is used in the later sections. Section IV contains the polynomial approximation results. Our main result is Theorem 8, which says that the volume of the misclassified region when a decision region with smooth boundary is approximated by the positive domain of a polynomial of degree d, goes to zero at least as fast as d^{-1}. By "smooth boundary" we mean essentially that the boundary is a finite union of $n-1$-dimensional manifolds. In Section V, a similar result is given for decision regions defined by neural networks results and the two results are compared. When the number of nodes in the network is chosen so that the polynomials and neural networks are defined by the same number of parameters, Theorem 11 says that the volume of the misclassified region goes to zero at least as fast as d^{-r}, where r can be made as close to (but less than) 1 as desired. This is slightly weaker than the result for polynomial decision regions. Section VI concludes the paper.

II. The Approximation Problem

We assume that a decision region is a closed subset D of a compact set $X \subset \mathbb{R}^n$, called the sample space. Points in the sample space are classified positively if they are contained in the decision region, and negatively if they are not. We wish to determine how well a decision region can be approximated by the positive domain of functions belonging to a parametrized class of functions, in the sense of minimizing the probability of misclassification. If points to be classified are chosen uniformly throughout the sample space X, the probability of misclassification is equal to the *volume of the misclassified region*, i.e., the volume of the symmetric difference of the two sets. For decision regions $D_1, D_2 \subset X$, the volume of the misclassified region is

$$V(D_1, D_2) := \operatorname{vol}(D_1 \Delta D_2) = \int_{D_1 \Delta D_2} dx.$$

For a decision region $D \subset X$ and an approximate decision region $\Sigma \subset X$, we say that Σ approximates D well if $V(D, \Sigma)$ is small; thus most points in X are correctly classified by Σ.

Typically, one is interested in approximating decision regions that belong to some class of subsets of X. Our results are for decision regions which have boundaries that are a finite union of hypersurfaces—$(n-1)$-dimensional submanifolds of $\mathbb{R}^n$.

Definition 1: A set $M \subset \mathbb{R}^n$ is an $n-1$-dimensional submanifold of $\mathbb{R}^n$ if for every $x \in M$, there exists an open neighborhood $U \subset \mathbb{R}^n$ of x and a function $f: U \to \mathbb{R}^n$ such that $f(U) \subset \mathbb{R}^n$ is open, f is a C^∞ diffeomorphism onto its image and either

1) $f(U \cap M) = f(U) \cap \mathbb{R}^{n-1}$, or
2) $f(U \cap M) = f(U) \cap \{y \subset \mathbb{R}^{n-1}: y(1) \geq 0\}$.

Here $y(1)$ denotes the first component of the vector y. The usual definition of a submanifold allows only the first case. When both cases are allowed, M is usually called a *submanifold with boundary*. We allow both cases because our consideration of decision regions confined to a compact domain implies that many interesting decision boundaries are not true submanifolds.

Definition 2: The piecewise-smooth decision regions in $X \subset \mathbb{R}^n$ are the sets in the collection

$$\mathcal{D}(X) = \{D \subset X: \partial D \text{ is a finite union of } n-1\text{-dimensional submanifolds of } \mathbb{R}^n\}$$

where ∂D denotes the boundary of D.

Allowing ∂D to be a *union* of submanifolds rather than a single submanifold means D may have (well-behaved) sharp edges. For instance, if $X = [1,1]^n$ and the decision region is the halfspace $\{x \in X: a^\top x \geq 0\}$, then the decision boundary consists of a union of up to $2n$ polygonal faces. Each of these faces is an $n-1$-dimensional submanifold (with boundary).

It is assumed that the approximating decision regions belong to a class $\mathcal{C}^d$ of subsets of X which gets progressively larger as d increases. That is, $\mathcal{C}^{d_1} \subset \mathcal{C}^{d_2}$ if $d_1 < d_2$. Typically, d is a nondecreasing function of the dimension of the parameter space. If the true decision region is D, then for any particular choice of d the minimum approximation error is $\inf_{\Sigma \in \mathcal{C}^d} V(D, \Sigma)$. Clearly, the minimum approximation error is a nonincreasing function of d. For some choices of $\mathcal{C}^d$, the minimum approximation error goes to zero as $d \to \infty$. In such cases, the classes $\mathcal{C}^d$ are said to be uniform approximators. The degree of approximation problem for uniform approximators $\mathcal{C}^d$ involves determining how quickly the minimum approximation error decreases.

The Degree of Approximation Problem: Let $X \subset \mathbb{R}^n$ be compact and for each $d > 0$ let $\mathcal{C}^d$ be a set of subsets of X such that

$$\lim_{d \to \infty} \sup_{D \in \mathcal{D}(X)} \inf_{\Sigma \in \mathcal{C}^d} V(D, \Sigma) = 0.$$

Find the largest $R \geq 0$ such that, for all sufficiently large d

$$\sup_{D \in \mathcal{D}(X)} \inf_{\Sigma \in \mathcal{C}^d} V(D, \Sigma) \leq \frac{c}{d^R} \tag{2}$$

where c is constant with respect to d.

The constant R in (2) is called the *degree of approximation* for the class $\mathcal{C}^d$ of decision regions.

III. The Delta Corridor

Let

$$B(x,\delta) := \{z \in \mathbb{R}^n : \|x - z\| \le \delta\}$$

the closed-ball with center x and radius δ, where $\|\cdot\|$ denotes the 2 norm (Euclidean distance) in $\mathbb{R}^n$.

Definition 3: The δ corridor around the decision boundary is the set

$$\partial D + \delta := \bigcup_{x \in \partial D} B(x,\delta).$$

For any $y \in \mathbb{R}^n$, $y \le \delta$

$$\partial D + y = \{x + y : x \in \partial D\} \subset \partial D + \delta.$$

The construction in Section IV of the approximating set Σ bounds the misclassified region by the volume of a δ corridor around the decision boundary. Thus in order to answer the approximation problem, it is necessary to determine the volume of the decision boundary. This requires some knowledge of the size and smoothness of ∂D. For instance, if ∂D is a space-filling curve, then the volume of *any* corridor around ∂D will be equal to the volume of X, and knowledge of the size of the corridor offers no advantage. On the other hand, if D is a ball with radius greater than the corridor size, then the volume is equal to two times the corridor size multiplied by the surface area of the ball. In order to obtain a general result for decision regions in $\mathcal{D}(X)$, we use the following definition for the area of a hypersurface [23], [24].

Definition 4: Let $\partial D \subset \mathcal{D}(X)$, and let the points $u \in \partial D$ be locally referred to parameters $u(1), \cdots, u(n-1)$, which are mapped to the Euclidean space $\mathbb{R}^{n-1}$ with the coordinates $v(1), \cdots, v(n-1)$. The surface area of ∂D is defined as

$$\text{area}\,(\partial D) := \int_{\partial D} \det(R)\, du(1) \cdots du(n-1)$$

where $R = [R_{ij}]$, $R_{ij} = \partial v(i)/\partial u(j)$. Thus area (∂D) is the volume of the image of ∂D in $\mathbb{R}^{n-1}$.

If $n = 2$, then ∂D is a curve in the plane, and area (∂D) is the length of ∂D. The size of area (∂D) increases rapidly with n, for instance, if $D = [-1,1]^n$ then area $(\partial D) = 2^{n+1}n$.

Using this definition, the volume of the corridor around a decision boundary can be bounded as follows:

Lemma 5: Let $X \subset \mathbb{R}^n$ be compact. For any $D \in \mathcal{D}(X)$ there exists $\Delta = \Delta(D) > 0$ such that

$$\text{vol}\,(\partial D + \delta) \le c\delta\ \text{area}\,(\partial D)$$

for all δ such that $0 < \delta < \Delta$.

This result is intuitively obvious, since ∂D can be locally approximated by an $n-1$-dimensional hyperplane, and the volume of the δ corridor around a piece of an $n-1$-dimensional hyperplane with area a is $2\delta a + O(\delta^2)$. A rigorous proof of Lemma 5 can be given using a result by Weyl that appears in [23].

IV. Polynomial Decision Regions

Definition 6: $\mathcal{P}_d^n$ is the space of polynomials of degree at most d in each of n variables. That is, $\mathcal{P}_d^n$ is the space of all linear combinations of $x(1)^{s_1}x(2)^{s_2}\cdots x(n)^{s_n}$ with $s_i \le d$, $s_i \in \mathbb{N}_0$. The number of parameters necessary to identify elements in $\mathcal{P}_d^n$ is $(d+1)^n$.

$\mathcal{CP}_d^n$ is the class of polynomial decision regions. Each decision region in $\mathcal{CP}_d^n$ is the positive domain of a polynomial in $\mathcal{P}_d^n$. Specifically,

$$\mathcal{CP}_d^n := \left\{ \Sigma \subset X : \exists f \in \mathcal{P}_d^n \text{ satisfying } \begin{array}{ll} f(x) \ge 0, & \text{if } x \in \Sigma \\ f(x) < 0, & \text{if } x \notin \Sigma \end{array} \right\}.$$

In this section and in Section V, $c \in \mathbb{R}$ denotes a quantity which is independent of d. Dependence of c on other variables will be indicated by, for instance, $c = c(n)$. If no such indication is given, c is an absolute constant. The exact value of c will change without notice, even in a single expression.

The following Theorem is derived from Timan [25, result 5.3.2].

Theorem 7 (Timan): Let $X = [-1,1]^n$ and $g : X \to \mathbb{R}$. If g is L_1-integrable on X then there exists a constant c such that

$$\inf_{f \in \mathcal{P}_d^n} \int_X |g(x) - f(x)|\, dx \le cn \sup_{\|y\| \le 1/(d+1)} \int_X |g(x) - g(x+y)|\, dx.$$

In the following, Theorem 7 is used to determine the degree of approximation of decision regions possessing a smooth boundary by polynomial decision regions.

Theorem 8: Let $X \subset \mathbb{R}^n$ be compact. If $D \in \mathcal{D}(X)$ then there exists a constant c such that

$$\inf_{\Sigma \in \mathcal{CP}_d^n} V(D,\Sigma) < \frac{cn\ \text{area}\,(\partial D)}{d+1}.$$

Proof: From the definition of $V(D,\Sigma)$

$$\begin{aligned} \inf_{\Sigma \in \mathcal{CP}_d^n} V(D,\Sigma) &= \inf_{f \in \mathcal{P}_d^n} \int_X |y_D(x) - \text{sgn}(g(x))|\, dx \\ &= \inf_{f \in \mathcal{P}_d^n} \int_{\{x : |y_D - g(x)| > 1\}} \cdot |y_D(x) - \text{sgn}(g(x))|\, dx \\ &\le 2 \inf_{f \in \mathcal{P}_d^n} \int_X |y_D(x) - g(x)|\, dx. \end{aligned}$$

Now y_D is L_1-integrable, so Theorem 3 applies with $g = y_D$. From the definition of the δ-corridor

$$\sup_{\|y\| \le \delta} \int_X |y_D(x) - y_D(x+y)|\, dx \le \text{vol}(\partial D + \delta).$$

Letting $\delta = 1/(d+1)$ and combining with Lemma 5 gives the result.

V. Neural Network Decision Regions

Definition 9: $\mathcal{N}_d^n$ is the space of functions defined by single hidden layer feedforward neural networks with n inputs, and

d nodes in the hidden layer. That is, $\mathcal{N}_d^n$ is the space of all linear combinations

$$\delta_0 + \sum_{i=1}^{d} \alpha_i \phi(\beta_i^\top x + \delta_i)$$

where $x, \beta_i \in \mathbb{R}^n, \alpha_i, \delta_i \in \mathbb{R}$, and $\phi(x) = (1+e^{-x})^{-1}$. In order to identify elements in $\mathcal{N}_d^n$, one must specify $1+d(n+2)$ real numbers.

$\mathcal{CN}_d^n$ is the class of neural network decision regions. Each decision region in $\mathcal{CN}_d^n$ is the positive domain of a function in $\mathcal{N}_d^n$. Specifically

$$\mathcal{CN}_d^n := \left\{ \Sigma \subset X : \exists f \in \mathcal{N}_d^n \text{ satisfying } \begin{array}{ll} f(x) \geq 0, & \text{if } x \in \Sigma \\ f(x) < 0, & \text{if } x \notin \Sigma \end{array} \right\}.$$

Theorem 10 is derived from [9, Corollary 5.2], since the degree of approximation by polynomials is the same as the degree of approximation by trigonometric polynomials in this case.

Theorem 10 (Mhaskar and Micchelli): Let $X \subset \mathbb{R}^n$ be compact and $g: X \to \mathbb{R}$. There exists a function $d: \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that if g is L_1 integrable on X then for any $p, q \in \mathbb{N}$, there exists a constant c such that

$$\inf_{f \in \mathcal{N}_{d(p,q)}^n} \int_X |g(x) - f(x)|\, dx \leq c \left(\inf_{f \in \mathcal{P}_p^n} \int_X |g(x) - f(x)|\, dx + p^n e^{-cq} \right) \int_X |g(x)|\, dx.$$

Moreover, there exists $c > 0$ such that $d(p,q) \leq cp^n q^2$ for sufficiently large p and q.

Thus the degree of approximation by neural net functions is related to the degree of approximation by polynomial functions. The upper bound in Theorem 10 is a monotonic decreasing function if both p and q are increasing.

Next we use the technique in Section IV to obtain a degree of approximation result for neural network decision regions. In order to get a fair comparison with the polynomial decision regions, we consider degree of approximation by $\mathcal{CN}_{d^n}^n$, since the number of parameters necessary to specify elements in either $\mathcal{CP}_d^n$ or $\mathcal{CN}_{d^n}^n$ is approximately $c(n)d^n$.

Theorem 11: Let $X \subset \mathbb{R}^n$ be compact. If $D \in \mathcal{D}(X)$ then for any $r \in (0,1)$ there exist constants c, $c(n,r)$ such that

$$\inf_{\Sigma \in \mathcal{CN}_{d^n}^n} V(D, \Sigma) < \frac{c \text{ area } (\partial D)}{d^r}$$

for all $d \geq c(n,r)$.

Proof: Assume $d^n = cp^n q^2 \geq d(p,q)$. From Theorem 10, the minimum misclassified region satisfies

$$\inf_{\Sigma \in \mathcal{CP}_{d^n}^n} \leq \inf_{\Sigma \in \mathcal{CP}_{d(p,q)}^n} \leq c \left(\frac{n \text{ area } (\partial D)}{p+1} + p^n e^{-cq} \right). \quad (3)$$

Now choose p and q such that the upper bound in (3) decreases to zero as quickly as possible as p and q increase. Let $q = p^\alpha$, so that $d = p^{1+\alpha}$. For any $\alpha > 0$, $p^n e^{-cp^\alpha} < p^{-1}$ is decreasing for sufficiently large p. How large p must be depends on α. Now for any $r \in (0,1)$ there exists $\alpha > 0$ such that $r = (1+\alpha)^{-1}$, so there exists a choice of p, q such that

$$\frac{n \text{ area } (\partial D)}{p+1} + p^n e^{-cq} \leq \frac{cn \text{ area } (\partial D)}{d^r}. \quad (4)$$

The result follows. ∎

Comparing Theorem 11 with Theorem 8, it can be seen that the bound on the degree of approximation is stronger for polynomial decision regions than neural network decision regions. This is a consequence of the function approximation results that are used, and in the absence of lower bounds on the rate of function approximation for neural networks, it is possible that a stronger upper bound may be found in the future. Moreover, lower bounds on the degree of approximation for decision regions are needed in order to determine whether polynomial decision regions or neural network decision regions are better approximators of decision regions.

VI. Concluding Remarks

We have given degree of approximation results for implicit decision region approximation which are similar to Jackson's Theorem for polynomial function approximation. The approximating decision regions are defined by the positive domains of polynomial functions or feedforward neural networks. These results support our intuition that classes of functions which are good function approximators tend to be good implicit decision region approximators.

Many open problems remain—the most pressing being "What conditions give better degree of approximation?" In function approximation, higher order smoothness of the approximated function gives a better degree of approximation. For instance, in Theorem 7 if the pth derivative is Lipschitz-continuous, then the degree of approximation is at least $p+1$. We would expect that there exist restrictions on the decision region to be approximated, D, which will guarantee a better degree of approximation than our results suggest. Moreover, we would expect that there would be a series of successively tighter restrictions on D which would guarantee successively better degree of approximation results.

However, it is not clear what the right conditions are. Bounding the curvature of the boundary of D will not affect the degree of approximation using our argument, since all information about the decision boundary other than its area affects only higher order terms in the approximation bound, not the degree of approximation obtained in Theorem 8. Perhaps the number of connected components in D is the condition we need. Or perhaps the curvature properties of the decision boundary are important, but a tighter method of bounding $V(D, \Sigma)$ than the volume of the corridor size is needed. Maybe a completely different proof technique is needed to get higher degree of approximation results.

Acknowledgment

The authors wish to thank the reviewers and associate editor of this journal for their comments, and P. Bartlett for helpful discussions motivating this research. In particular, H. Mhaskar pointed out in reviews that the use of L_1-function

approximation results would give a much simpler derivation and a tighter bound than the L_∞ technique we originally proposed.

REFERENCES

[1] J. Sklansky and G. Wassel, *Pattern Classifiers and Trainable Machines*. New York: Springer-Verlag, 1981.

[2] K. Blackmore, R. Williamson, and I. Mareels, "Learning nonlinearly parametrized decision regions," *J. Math. Syst., Estim., Contr.*, vol. 6, no. 1, pp. 129–132, 1996. (Full manuscript [On line]. Available FTP: ftp://trick.ntp.springer.de/jmsec/88289.ps.)

[3] G. Lorentz, *Approximation of Functions*. New York: Holt, Rinehart and Winston, 1966.

[4] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.

[5] E. Cheney, *Introduction to Approximation Theory*, 2nd ed. New York: Chelsea, 1982.

[6] K. Hornik, M. Stinchcombe, and H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward neural networks," *Neural Networks*, vol. 3, pp. 551–560, 1990.

[7] A. Barron, "Universal approximation bounds for superposition of a sigmoidal function," *IEEE Trans. Inform. Theory*, vol. 39, pp. 930–945, 1993.

[8] H. Mhaskar and C. Micchelli, "Approximation by superposition of sigmoid and radial basis functions," *Adv. Appl. Math.*, vol. 13, pp. 350–373, 1992.

[9] ———, "Degree of approximation by neural and translation networks with a single hidden layer," *Adv. Appl. Math.*, vol. 16, pp. 151–183, 1995.

[10] H. Mhaskar, "Approximation properties of a multilayered feedforward artificial neural network," *Adv. Comput. Math.*, vol. 1, pp. 61–80, 1993.

[11] C. Darken, M. Donahue, L. Gurvits, and E. Sontag, "Rate of approximation results motivated by robust neural network learning," in *Proc. 6th ACM Conf.on Computational Learning Theory*, 1993, pp. 103–109.

[12] K. Hornik, M. Stinchcombe, H. White, and P. Auer, "Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives," Tech. Rep. NC-TR-95-004, NeuroCOLT Tech. Rep. Ser., Jan. 1995.

[13] A. Bengtsson and J.-O. Eklundh, "Shape representation by multiscale contour approximation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, pp. 85–93, 1991.

[14] P. Kenderov and N. Kirov, "A dynamical systems approach to the polygonal approximation of plane convex compacts," *J. Approx. Theory*, vol. 74, pp. 1–15, 1993.

[15] J.-S. Wu and J.-J. Leou, "New polygonal approximation schemes for object shape representation," *Pattern Recogn.*, vol. 26, pp. 471–484, 1993.

[16] N. Korneichuk, "Approximation and optimal coding of plane curves," *Ukranian Math. J.*, vol. 41, no. 4, pp. 429–435, 1989.

[17] A. Korostelev and A. Tsybakov, "Estimation of the density support and its functionals," *Probl. Inform. Transm.*, vol. 29, no. 1, pp. 1–15, 1993.

[18] ———, *Minimax Theory of Image Reconstruction* (Lecture Notes in Statistics, vol. 82). New York: Springer, 1993.

[19] R. Dudley, "Metric entropy of some classes of sets with differentiable boundaries," *J. Approx. Theory*, vol. 10, pp. 227–236, 1974.

[20] N. Shchebrina, "Entropy of the space of twice smooth curves in $\mathbb{R}^{n+1}$," *Math. Acad. Sci. USSR*, vol. 47, pp. 515–521, 1990.

[21] N. Ivanov, "Approximation of smooth manifolds by real algebraic sets," *Russian Math. Surv.*, vol. 37, no. 1, pp. 1–59, 1982.

[22] F. Broglia and A. Tognoli, "Approximation of C^∞ functions without changing their zero-set," *Ann. Inst. Fourier, Grenoble*, vol. 39, no. 3, pp. 611–632, 1989.

[23] H. Weyl, "On the volume of tubes," *Amer. J. Math.*, vol. 61, pp. 461–472, 1939.

[24] M. Berger and B. Gostiaux, *Differential Geometry: Manifolds, Curves, and Surfaces*. New York: Springer-Verlag, 1988.

[25] A. Timan, *Theory of Approximation of Functions of a Real Variable*. New York: Macmillan, 1963.