NEUROCOMPUTING

ELSEVIER

Neurocomputing 29 (1999) 47–56

www.elsevier.com/locate/neucom

Approximation by neural networks is not continuous

Paul C. Kainen[a], Věra Kůrková[b,*], Andrew Vogt[a]

[a]*Georgetown University, Washington, DC, USA*

[b]*Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věži 2, P.O. Box 5, 182 07 Prague 8, Czech Republic*

Received 12 January 1998; accepted 5 June 1998

Abstract

It is shown that in a Banach space X satisfying mild conditions, for its infinite, linearly independent subset G, there is no continuous best approximation map from X to the n-span, $span_n\, G$. The hypotheses are satisfied when X is an $\mathscr{L}_p$-space, $1 < p < \infty$, and G is the set of functions computed by the hidden units of a typical neural network (e.g., Gaussian, Heaviside or hyperbolic tangent). If G is finite and $span_n\, G$ is not a subspace of X, it is also shown that there is no continuous map from X to $span_n\, G$ within any positive constant of a best approximation. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Nonlinear approximation; One-hidden-layer neural networks; Rates of approximation; Continuous selection; Metric projection; Proximinal set; Chebyshev set; n-width; Geometry of Banach spaces

1. Introduction

In a typical approximation scheme, the approximating functions are members of a parametrized family; their complexity can be measured by the length of a parameter vector. This corresponds, for example, to the degree of a polynomial or rational function, the number of knots in a spline, or the number of hidden units in a neural network.

* Corresponding author. Tel.: 004202-6605-3231; fax: 004202-858-5789.

E-mail address: vera@uivt.cas.cz (V. Kůrková)

PII: S0925-2312(99)00111-3

For a given length of parameter vector, the approximating functions often form a linear subspace (as in the polynomial case). In contrast, the approximating functions in neural network theory are members of *unions of finite-dimensional subspaces* generated by hidden unit functions. More precisely, if $G = \{g(.,a): D \to \mathscr{R}; a \in \mathscr{A}\}$, $D \subset \mathscr{R}^d$, is a parametrized set of functions corresponding to a type of computational unit, then a one-hidden-layer network with a single linear output unit and n hidden units computing functions from G can generate as its input/output functions all linear combinations of n elements of G. This set, denoted by $span_n G$, is the union of all subspaces spanned by n-tuples of elements of G.

In recent years, various authors derived upper bounds on rates of approximation by neural networks. Some of these upper bounds were achieved using *continuous* approximation operators on sets of functions defined by a smoothness condition (see e.g. [17]).

Continuity of an approximation operator is a great advantage in the classical (linear) theory since it allows one to estimate worst-case error using methods of algebraic topology (see e.g. [12,16,18]). Extending these algebraic-topological proof techniques to nonlinear approximation, DeVore et al. [4] utilized a concept of continuous nonlinear width. They measured the worst-case error in approximation of elements in a compact subset using a parametrized set of functions as the approximants when the parameters are selected continuously. For neural networks, however, the difference in the sort of subset determined by the families of approximating functions constrains the domain of applicability of classical methods.

In this paper we study the existence of continuous best approximation in neural networks. We show that for certain standard types of neural networks, such as Heaviside perceptrons or Gaussian radial-basis-functions, best approximation with n hidden units cannot be achieved in a continuous way. This has important practical consequences; in particular, numerical stability of computation is not guaranteed, even under "low amplitude" assumptions. A theoretical consequence is that estimates of worst-case errors of approximation that exploit continuity of the approximation operator cannot be applied to neural networks.

The paper is organized as follows. In Section 2, we recall basic concepts from approximation theory such as metric projection, best approximation, Chebyshev set, and continuous selection. In Section 3, we investigate approximation from subsets of normed linear spaces. It is shown that continuous best approximations do not exist under some mild hypotheses on the norm and the subsets. If the subsets are finite unions of finite-dimensional spaces, then it is not possible to come within any positive constant of the best approximation in a continuous way. Section 4 applies the results of Section 3 to parametrized sets of functions corresponding to neural networks, and Section 5 briefly summarizes the implications.

2. Preliminaries

Let $\mathscr{R}$ denote the set of real numbers. In the following, a linear space X always means a linear space over $\mathscr{R}$. The *dimension* of a linear space is the cardinality of any basis.

For x, y in a linear space X we denote by $[x, y]$ the *closed* (*line*) *segment* connecting x and y, i.e. $[x, y] := \{x + \lambda(y - x)\colon \lambda \in [0,1]\}$. Similarly, $(x, y] = [x, y] - \{x\}$. Recall that a subset $Y \subseteq X$ is *convex* if $[x, y] \subseteq Y$ whenever $x, y \in X$. A subset Y is called *positively homogeneous* if for all $a > 0$, $aY := \{ay\colon y \in Y\} = Y$. For $G \subseteq X$ we denote by *span* G the linear span of G, and for n a positive integer $span_n\, G$ denotes the set of all linear combinations of at most n elements of G, i.e., $span_n\, G := \{x \in X\colon x = \sum_{i=1}^{n} w_i g_i,\ w_i \in \mathscr{R},\ g_i \in G\}$. Thus $span_n\, G = \bigcup\{span\{g_1, \ldots, g_n\}\colon g_1, \ldots, g_n \in G\}$. By $B[x, r]$ we denote the closed ball of radius r centered at x; i.e., $B[x, r] := \{y \in X\colon \|x - y\| \le r\}$.

A normed linear space $(X, \|.\|)$ is said to be *uniformly convex* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in X$ $\|x\| = \|y\| = 1$ and $\|x - y\| > \varepsilon$ implies $\|x + y\| < 2 - \delta$. It is *strictly convex* if for all $x, y \in X$ $\|x\| = 1$, $\|y\| = 1$ and $\|x + y\| = 2$ implies $x = y$. Strict convexity is equivalent to the following condition: whenever for three points $x, y, z \in X$ the triangle inequality becomes equality, i.e. $\|x - z\| = \|x - y\| + \|y - z\|$, then $y \in [x, z]$ (see e.g. [16]). Uniform convexity implies strict convexity and $\mathscr{L}_p$-spaces are uniformly convex for $p \in (1, \infty)$ (see [5]).

A Banach space X is a complete normed linear space. The set of continuous linear functionals $f\colon X \to \mathscr{R}$ is also a Banach space, called the *dual* and denoted by X^*. X is *reflexive* if it is isometrically isomorphic to $X^{**} := (X^*)^*$ under the canonical mapping $x \to x'$, where $x'(f) := f(x)$. X is *smooth* if for each x such that $\|x\| = 1$, there is a unique $f \in X^*$ such that $f(x) = 1$ and $f(y) < 1$ for all $y \in Y$ with $\|y\| < 1$.

Let X be a normed linear space and let Y be a non-empty subset of X. For $x \in X$, let $\|x - Y\| = \inf_{y \in Y}\|x - y\|$. The functional $x \mapsto \|x - Y\|$ is uniformly continuous; see e.g. [20, p. 391].

Let $\mathscr{P}(Y)$ denote the set of all subsets of Y. The set-valued mapping $P_Y\colon X \to \mathscr{P}(Y)$ defined by $P_Y(x) := \{y \in Y\colon \|x - y\| = \|x - Y\|\}$ is called the *metric projection* or *projection of X onto Y* and $P_Y(x)$ is called the *projection of x onto Y*. If Y is closed, resp. convex, then for all $x \in X$, $P_Y(x)$ is closed, resp. convex. Note that we do not assume that $P_Y(x)$ is non-empty.

When $P_Y(x)$ is non-empty for all x in X, then Y is said to be *proximinal* (or an *existence set*). Y is called a *Chebyshev set* if for all $x \in X$ the projection $P_Y(x)$ is a singleton. In this case we denote by p_Y the unique projection mapping $p_Y\colon X \to Y$. We shall also write $x_Y = p_Y(x)$.

Let $F\colon X \to \mathscr{P}(Y)$ be a set-valued function. A *selection for F* is a function $f\colon X \to Y$ such that for all $x \in X, f(x) \in F(x)$. A function $\phi\colon X \to Y$ is called a *best-approximation operator for Y* if it is a selection for P_Y. A set Y is proximinal if and only it has a best-approximation operator; it is Chebyshev if and only if it has a unique best approximation operator. General references on best approximation are Singer [20] and Vlasov [24]. A continuous selection for P_Y is thus a best-approximation operator for Y which is continuous at each $x \in X$. A best-approximation operator ϕ has $\phi(x) \in P_Y(x)$ and satisfies $\|x - \phi(x)\| = \|x - Y\|$ for every $x \in X$.

3. Non-existence of continuous best approximations

In this section we study approximation from subsets of a normed linear space and determine geometric conditions on the subset and the space which prevent the existence of a continuous best approximation.

The following lemma says that metric projection acts as regularly as possible on segments joining points in X to their nearest neighbors in a subset Z.

Lemma 3.1. *Let X be a strictly convex normed linear space, and let Z be any subset of X. Let $x \in X$, $x \notin Z$, and $z \in P_Z(x)$. Then $P_Z([z,x)) = \{z\}$.*

Proof. If $y \in (x,z]$, then $\|y - Z\| = \|y - z\|$ by the triangle inequality and the fact that z is in $P_Z(x)$. Thus $z \in P_Z(y)$. If $z' \in P_Z(y)$, then $\|x - z'\| \leq \|x - y\| + \|y - z'\| = \|x - y\| + \|y - z\| = \|x - z\|$. Hence, z' belongs to $P_Z(x)$, the last inequality is an equality, and by strict convexity y lies in $(x, z']$. Since the intervals $[x,z]$ and $[x,z']$ have the same length and both contain $[x,y]$, $z = z'$. Therefore, $P_Z(y) = \{z\}$. □

Our first theorem shows that in a strictly convex normed linear space, existence of a continuous best approximation implies its uniqueness. A point x in X is called a *splitting point* if and only if the cardinality of $P_Z(x)$ is larger than 1. Such a point is also called a *point of nonuniqueness*; see Rice [19].

Theorem 3.2. *Let X be a strictly convex normed linear space, and let Z be any subset of X. If Z is not Chebyshev, then P_Z does not have a continuous selection.*

Proof. Z must be proximinal or no best-approximation mapping from X to Z could exist. If Z is not Chebyshev, there is a splitting point $x \in X$ with $z_1 \neq z_2$ both in $P_Z(x)$. Let $V := [z_1, x] \cup [x, z_2]$. By Lemma 3.1 any selection must map $V - \{x\}$ into $\{z_1, z_2\}$. Hence, the selection cannot be continuous on V. □

In particular, the selection fails to be continuous at the splitting point x and fails to be directionally continuous along one of the two directions $[x, z_1]$ and $[x, z_2]$.

The above proof shows further that if for some ε there is a continuous function $\phi : X \to Z$ satisfying $\|\phi(x) - P_Z(x)\| \leq \varepsilon$ for all x in X, then $\sup\{diameter(P_Z(y)) : y \in X\} \leq 2\varepsilon$.

Continuity is not guaranteed even on neighborhoods of zero if Z satisfies an additional condition. Recall that a neighborhood of zero means a set whose interior contains zero.

Theorem 3.3. *Let X be a strictly convex normed linear space, and let Z be any non-Chebyshev subset of X that is positively homogeneous. Then in any neighborhood N of zero the restriction of P_Z to N has no continuous selection.*

Proof. Note that for all $a > 0$ and any positively homogeneous Z, $P_Z(ax) = aP_Z(x)$. Using the fact that, for $a > 0$, $\|ax - Z\| = a\|x - Z\|$ for all $x \in X$ ([20, p. 148]), we show that $P_Z(ax) \subseteq aP_Z(x)$. Indeed, suppose $y \in P_Z(ax)$ or equivalently, $\|y - ax\| = \|Z - ax\|$. We need to show that $a^{-1}y \in P_Z(x)$. But this follows from $\|a^{-1}y - x\| = a^{-1}\|y - ax\| = a^{-1}\|Z - ax\| = \|Z - x\|$. As x and $a > 0$ were arbitrary, the reverse inclusion holds as well; hence, the equation holds.

If Z were not proximinal, then for some x_0, $P_Z(x_0) = \emptyset$. So, for any neighbourhood N of zero, we could choose a so that $ax_0 \in N$ and $P_Z(ax_0) = \emptyset$. Thus no selection could exist for the restriction of P_Z to N.

If Z is proximinal, let x be a splitting point with z_1 and z_2 in $P_Z(x)$, $z_1 \neq z_2$. For a suitable scalar a, az_1, az_2, and ax are in a convex neighborhood of zero inside N. Then $[az_1, ax]$ and $[az_2, ax]$ are in N, ax is a splitting point, and the argument in Theorem 3.2 can be applied. □

The problem of convexity of Chebyshev sets in a general Banach space has been studied extensively (see e.g. [10,8]). Closed convex sets in reflexive strictly convex Banach spaces are Chebyshev (see, e.g., Singer [20, pp. 111, 364]). Bunt [2] showed that the converse (Chebyshev implies convex) holds in finite-dimensional Hilbert spaces. It is an open question whether the converse is true in the infinite-dimensional Hilbert case.

Vlasov [22,23,20, p. 368] gave two useful sufficient conditions for convexity of a Chebyshev subset of a Banach space. Recall that a subset Z is boundedly compact if its intersection with every closed ball is compact.

(V1) *In a Banach space with strictly convex dual, every Chebyshev subset with continuous metric projection is convex.*

(V2) *In a smooth Banach space, any boundedly compact Chebyshev set is convex.*

Theorem 3.4. *Let X be a strictly convex Banach space with a strictly convex dual and let Z be any non-convex subset of X. Then $P_Z : X \to \mathscr{P}(Z)$ does not have a continuous selection.*

Proof. If Z is not Chebyshev, by Theorem 3.2, no continuous selection exists. If Z is Chebyshev with continuous projection, then by V1, Z would be convex. Hence, there can be no continuous selection from P_Z. □

Theorem 3.5. *Let X be a strictly convex Banach space with a strictly convex dual and let Z be any positively homogeneous subset that is not convex. Then in any neighborhood N of zero, the restriction of P_Z has no continuous selection.*

Proof. If Z is not Chebyshev, Theorem 3.3 yields the conclusion. If Z is Chebyshev, by V1, continuity of the metric projection P_Z would imply that Z is convex. So $P_Z = p_Z$ fails to be continuous at some point x. For some scalar a, ax is in N, and p_Z is not continuous at ax. □

In the important special case when Z is the union of finitely many finite-dimensional subspaces, metric projection has a kind of "robust" non-continuity.

Theorem 3.6. *Let X be a smooth, strictly convex Banach space, and let Z be a finite union of finite-dimensional subspaces that is not itself a subspace. Then for every $\varepsilon > 0$, there is* *no continuous function* $\phi : X \to Z$ *satisfying either of the following conditions*:

(i) *For all* $x \in X$, $\|\phi(x) - P_Z(x)\| \leq \varepsilon$;

(ii) *For all* $x \in X$, $\|x - \phi(x)\| \leq \|x - Z\| + \varepsilon$.

Proof. Let $Z = \bigcup \mathscr{S}$, where $\mathscr{S}$ is a finite family of finite-dimensional subspaces of X. Since X is strictly convex, each subspace $T \in \mathscr{S}$ is Chebyshev and for each x in X there is a unique nearest point x_T in T. So $\|x - Z\| = \|x - x_T\|$ for one or more of the members T of $\mathscr{S}$.

Since Z is a union of subspaces but not itself a subspace, Z cannot be convex. However, Z is boundedly compact and hence, by Vlasov's second result (V2 above), Z is not Chebyshev.

Any function ϕ satisfying condition (i) above automatically satisfies (ii) for the same ε. Hence, it suffices to find a point x and an ε so that if ϕ satisfies (ii), then ϕ is not continuous. Larger values of ε are obtained by scaling up by a positive constant, using the homogeneity of Z.

Choose a splitting point x, and let $x_1, \ldots, x_n$ be the distinct members of $P_Z(x)$ corresponding to distinct subspaces $T_1, \ldots, T_n$ in $\mathscr{S}$, $n > 1$. Choose $\varepsilon_1 > 0$ such that $\{B[x_i, \varepsilon_1]\colon i = 1, \ldots, n\}$ is a disjoint family. Again using bounded compactness of Z, we can choose $\varepsilon_2 > 0$ so that $B[x, \|x - Z\| + \varepsilon_2] \cap Z \subseteq \bigcup_{i=1}^{n} B[x_i, \varepsilon_1]$.

Assume $\phi : X \to Z$ satisfies condition (ii) with $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$. By Lemma 3.1, $\phi([x, x_i]) \subseteq B[x, \|x - Z\| + \varepsilon] \cap Z$. Since for each i $\phi(x_i) \in B[x_i, \varepsilon]$ by condition (ii), ϕ maps the connected set $\bigcup_{i=1}^{n} [x, x_i]$ onto a disconnected set, and so ϕ cannot be continuous. $\square$

Other types of nearly best approximation in addition to (i) and (ii) can also be considered [4]. Suppose $\phi : X \to Z$ satisfies:

(iii) For all $x \in X$, $\|x - \phi(x)\| \leq (1 + \varepsilon)\|x - Z\|$.

Then the argument of Theorem 3.6 will show that ϕ cannot be continuous if ε is sufficiently small, but one can find such maps which are continuous and satisfy (iii) for a large enough ε. This is related to the idea of a *retraction* which is a continuous function from a space to some subset which is the identity when restricted to the subset. If ϕ satisfies (iii), then $\phi(x) = x$ for every x in Z.

Suppose X is two-dimensional Euclidean space, and let Z be the union of the two coordinate axes. The retraction of X onto Z whose preimages are half-lines of slope ± 1 satisfies (iii) with $\varepsilon = \sqrt{2} - 1$.

For ε sufficiently small, however, the proof of Theorem 3.6 shows that such retractions do not exist. For if x is a splitting point and $\|y - \phi(y)\| \leq (1 + \varepsilon)\|y - Z\|$ for all y in $\bigcup_{i=1}^{n} [x, x_i]$, then $\|y - \phi(y)\| \leq \|y - Z\| + \varepsilon\|x - Z\|$.

In contrast to Theorem 3.6, when Z is a finite-dimensional linear manifold, DeVore et al. [4, Theorem 2.1] showed that for any Banach space X and any $\varepsilon > 0$, there does exist a continuous function $\phi : X \to Z$ satisfying (ii). Note also that the set-valued function P_Z is upper-semicontinuous whenever Z is boundedly compact ([20, p. 386], [1, p. 39]).

4. Application to neural networks

One-hidden-layer neural networks with a single linear output unit compute functions of the form

$$\sum_{i=1}^{n} w_i g(x, a_i), \tag{1}$$

where n is the number of hidden units, w_i are output weights and $g: D \times \mathscr{A} \to \mathscr{R}$ is the function of the hidden units with parameters $a_i \in \mathscr{A}$ and input vectors $x \in D \subseteq \mathscr{R}^d$. For example, perceptrons with an activation function $\psi: \mathscr{R} \to \mathscr{R}$, have $\mathscr{A} = \mathscr{R}^{d+1}$ and $g(x,(v,b)) = \psi(v \cdot x + b)$. For radial-basis-functions with radial function ψ, $\mathscr{A} = \mathscr{R}^d \times (0, \infty)$ and $g(x,(v,b)) = \psi(b\|x - v\|)$.

The set of functions of the form (1) with arbitrary output weights is equal to $span_n\, G$, where $G = \{g(\cdot, a): D \to \mathscr{R},\ a \in \mathscr{A}\}$. Being a union of homogeneous spaces, $span_n\, G$ is homogeneous. To apply Theorems 3.4 and 3.5, we investigate when $span_n\, G$ is convex.

Lemma 4.1. *Let X be a linear space and G a linearly independent subset. If n is less than the cardinality of G, then $span_n\, G$ is not convex.*

Proof. Let $\{g_1, \ldots, g_{n+1}\}$ be a subset of $n+1$ distinct elements of G. Choosing $x = (1/n+1)\sum_{j=1}^{n} g_j$ and $y = (1/n+1)\sum_{j=2}^{n+1} g_j$, we have $x, y \in span_n\, G$. Now, $u = (1/2)(x + y)$ is a non-trivial convex combination of $n+1$ elements of G. However, by the independence of G, u is not in $span_n\, G$. $\square$

Since $span_n\, G$ is a union of subspaces, it is convex if and only if it is a subspace. The following is an immediate consequence of Theorem 3.4 and Lemma 4.1.

Theorem 4.2. *Let Ω be a measurable subset of $\mathscr{R}^d$; let G be any infinite, linearly independent subset of $\mathscr{L}_p(\Omega)$, $1 < p < \infty$; and let N be a neighborhood of zero in $\mathscr{L}_p(\Omega)$. Then for n a positive integer, there is no continuous best approximation from N to $span_n\, G$.*

In fact, many neural network architectures produce families of hidden unit functions which satisfy the hypotheses of 4.2 or can be made to do so by restricting their parameter sets in a natural way.

The question of linear independence of parametrized sets of functions representing neural networks was first considered by Hecht-Nielsen [7] who investigated multiple global minima in parameter spaces. He conjectured that for a perceptron network with hyperbolic tangent as its activation, the neural network's resulting (input/output) function would determine network parametrization uniquely, up to sign flips and permutations of hidden units. This was proved by Sussmann [21]. Thus, with τ denoting hyperbolic tangent, a parametrized family $F_\tau = \{\tau(v \cdot x + b);\ (v,b) \in \mathscr{A}\}$, where $\mathscr{A} \subset \mathscr{R}^d \times \mathscr{R}$ contains for each pair of sign-flipped hidden unit parameters (v, b) and $(-v, -b)$ only one element, is linearly independent.

For the Heaviside activation function $\vartheta: \mathscr{R} \to \mathscr{R}$ with $\vartheta(t) = 0$ for $t < 0$ and $\vartheta(t) = 1$ for $t \geq 0$, Chui et al. [3] proved linear independence of the set of functions on the

d-cube computable by Heaviside perceptrons that contain for each pair of characteristic functions of complementary half-spaces only one representative.

Kůrková and Neruda [15] proved the linear independence of Gaussian radial-basis-functions $\{\gamma(b\|x - v\|)\colon v \in \mathscr{R}^d,\ b > 0\}$, where $\gamma(t) = \exp(-t^2)$. Characterization of linearly independent families for different types of activation functions was given by Kainen et al. [9], while Kůrková and Kainen [14] developed some general theory for corresponding functional equations.

In digital implementations of neural networks, one always has a finite set G; hence, $span_n\, G$ is a finite union of finite-dimensional subspaces.

Theorem 4.3. *Let Ω be a measurable subset of $\mathscr{R}^d$; let G be a linearly independent subset of $\mathscr{L}_p(\Omega)$, $1 < p < \infty$; and let Z be a finite union of finite-dimensional subspaces spanned by members of G, with Z not itself a subspace. Then, for each* $\varepsilon > 0$, *there is no continuous function* $\phi\colon \mathscr{L}_p(\Omega) \to Z$ *satisfying conditions* (i) *or* (ii) *above, and there is no such continuous function satisfying condition* (iii) *for* ε *sufficiently small.*

5. Discussion

The theory of neural networks overlaps, as the above arguments suggest, with approximation theory. For example, in the uniform norm best approximation by rational functions of a given degree fails to be continuous [1, p. 115], and we have shown that continuity also fails for neural networks in the $\mathscr{L}_p$-case, $1 < p < \infty$.

Questions concerning existence of best approximation, with or without constraints on the parameters, can be investigated by the methods used here. Relevant results include those of Gurvits and Koiran [6] and Kůrková [13] on compactness and closure of certain sets of Heaviside functions.

Suitability of an approximation scheme can be measured by the worst-case approximation error. When K, the set of functions to be approximated, and Y, the approximating set, are both subsets of a normed linear space $(X, \|.\|)$, recall that the *deviation of* K *from* Y is defined as $d(K, Y) = \sup_{x \in K}\|x - Y\|$.

Kolmogorov [11] defined the (linear) *n-width* of a (usually compact) subset K of X to be the least possible deviation from any n-dimensional subspace of X, i.e., $\inf\{d(K, Y)\colon Y$ is an n-dimensional subspace of $X\}$.

We suggest corresponding "neural network n-widths" of a set K: consider, for example, a perceptron-n-width $\inf_{\psi}\{d(K, span_n\, G_{\psi})\}$, where G_{ψ} is the parametrized family associated with perceptrons with activation function ψ. To find good lower bounds on these measures of complexity, arguments based on continuity will not suffice.

Acknowledgements

Collaboration of V. Kůrková and A. Vogt was supported by the National Research Council NAS COBASE grant. V. Kůrková was also partially supported by GA AV grant A2030602.

References

[1] D. Braess, Nonlinear Approximation Theory, Springer, Berlin, 1980.

[2] L.N.H. Bunt, Bijdrage tot de theorie der konvekse puntverzamelingen, Thesis University, Groningen, Amsterdam, 1934.

[3] C.K. Chui, X. Li, H.N. Mhaskar, Neural networks for localized approximation, Math. Comput. 63 (1994) 607–623.

[4] R. DeVore, R. Howard, C. Micchelli, Optimal nonlinear approximation, Manuscripta Math. 63 (1989) 469–478.

[5] A. Friedman, Foundations of Modern Analysis, Dover, New York, 1982.

[6] L. Gurvits, P. Koiran, Approximation and learning of convex superpositions, J. Comput. System Sci. 55 (1) (1997) 161–170.

[7] R. Hecht-Nielsen, On the algebraic structure of feedforward network weight spaces, in: Advanced Neural Computers, Elsevier, Amsterdam, 1990, pp. 129–135.

[8] R. Huotari, W. Li, Continuities of metric projection and geometric consequences, J. Approx. Theory 90 (1997) 319–339.

[9] P.C. Kainen, V. Kůrková, V. Kreinovich, O. Sirisengtaksin, Uniqueness of network parametrization and faster learning, Neural, Parallel Sci. Comput. 2 (1994) 459–466.

[10] V. Klee, Convexity of Chebyshev sets, Math. Annalen 142 (1961) 202–304.

[11] A.N. Kolmogorov, Uber die beste Annaherung von Funktionen einer gegenen Funktionenklasse, Ann. Math. 37 (1936) 107–111.

[12] M.G. Krein, M.A. Krasnoselski, D.P. Milman, On deficiency numbers of linear operators in Banach spaces and on some geometric problems, Sbornik Trudov Inst. Math. Akad. Nauk SSSR 11 (1948) 97–112.

[13] V. Kůrková, Approximation of functions by perceptron networks with bounded number of hidden units, Neural Networks 8 (1995) 745–750.

[14] V. Kůrková, P.C. Kainen, Singularities of finite scaling functions, Appl. Math. Lett. 9 (1996) 33–37.

[15] V. Kůrková, R. Neruda, Uniqueness of functional representations by Gaussian basis function networks, Proceedings of ICANN'94, Springer, London, 1994, pp. 471–474.

[16] G.G. Lorentz, Approximation of Functions, Halt, Reinhart, and Winston, New York, 1966.

[17] H.N. Mhaskar, Versatile Gaussian networks, in: Proceedings of IEEE Workshop on Nonlinear Image and Signal Processing, IEEE, 1995, pp. 70–73.

[18] A. Pinkus, n-Width in Approximation Theory, Springer, Berlin, 1985.

[19] J.R. Rice, The Approximation of Functions, vol. 2, Nonlinear and Multivariate Theory, Addison-Wesley, Reading, MA, 1969.

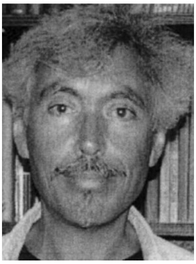
[20] I. Singer, Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces, Springer, Berlin, 1970.

[21] H.J. Sussmann, Uniqueness of the weights for minimal feedforward nets with a given input–output map, Neural Networks 5 (1992) 589–594.

[22] L.P. Vlasov, On Chebyshev subsets of a Banach space, Doklady Akad. Nauk SSSR 141 (1961) 19–20.

[23] L.P. Vlasov, Almost convex and Chebyshev sets, Math. Notes Acad. Sci. USSR 8 (1970) 776–779.

[24] L.P. Vlasov, Approximative properties of sets in normed linear spaces, Uspehi Mat. Nauk 28 (6) (1973) 3–66.



Paul Kainen is Adjunct Associate Professor at Georgetown University. He has a Ph.D. in algebraic topology from Cornell University. His interests include geometry, functional approximation, topological biology and quantum computing.

Věra Kůrková received her Ph.D. in topology from Charles University in Prague. She is the leader of a research group in the mathematical theory of neural networks at the Institute of Computer Science of the Czech Academy of Sciences. Here research interests are approximation of functions by neural networks and nonlinear approximation theory.

Andrew Vogt teaches at Georgetown University. He has a Ph.D. in mathematics from the University of Washington in Seattle. His interests include, functional analysis, the mathematical foundations of quantum theory, and statistics.