On the Rates of Nonlinear and Linear Approximation

Kateřina Hlaváčková–Schindler
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07, Praha 8, Czech Republic
katka@uivt.cas.cz
Marcello Sanguineti
Department of Communications, Computer, and System Sciences
DIST–University of Genova, Via Opera Pia 13, 16145 Genova, Italy
marce@dist.unige.it

Abstract

Motivated by the remarkable performance of nonlinear approximation schemes in the solution of complex problems (e.g., functional optimization problems from control theory, text pronounciation, etc.), we discuss some issues on the rates of approximation by linear and nonlinear approximators in certain spaces of functions. We demonstrate some qualitative arguments which support better experimental performance of nonlinear approximators with respect to the linear ones. Finally, using the concept of variation of a function with respect to a set, we present bounds to the rate of the approximation error achievable by nonlinear approximators in certain spaces of functions and show that they are better than the rates achieved by linear approximation in the same spaces. This motivates using the nonlinear approximators for solving complex problems.

Keywords:
nonlinear and linear approximators, rate of approximation, variation of a function with respect to a set of functions.

1. Introduction

When approximating functions which belong to a given function space by linear and nonlinear approximators, the *curse of dimensionality* is a problem shared by upper estimates of the number of parameters, obtained through constructive proofs of the density property (see, for example, [7], [8]). On the other hand, from a general result by De Vore et al. [9] on nonlinear parametrizations in which the optimal parameters depend continuously on the function to be approximated, we know that, if we want to approximate an s-time continuously differentiable function in d variables by a function containing p parameters, we cannot get an accuracy better than $(1/p)^{s/d}$ (known as *Jackson rate*).

From this we could conclude that nonexponential upper estimates on the number of parameters cannot be achieved, if the class of functions to be approximated is defined in terms of bounds on partial derivatives. This happens unless we let the smoothness to be an increasing function $s(d)$, in such a way to cope with the dimension d in the Jackson rate. Moreover, since the Jackson lower bound is achieved in Sobolev spaces by a large variety of both linear [11] and nonlinear [10] approximators, it would seem that there is no reason for nonlinear approximators to behave better than linear ones, at least in spaces defined by smoothness conditions. However, practical applications have however shown the possibility of approximating functions of hundred of variables by nonlinear approximators with very few parameters, e.g., [12] for feed-forward neural networks). Moreover, we have succesfully used such networks for the approximate solution of complex functional optimization problems coming from control theory: stabilization of high-order strongly nonlinear dynamic systems [16], nonlinear state estimation [17], fault diagnosis for nonlinear systems with modeling uncertainties [18], etc.

The comparison of the results on the rates of approximation from the literature is made even more complex by the fact that each approximator has been introduced to approximate functions from different spaces. Barron [1] gives a better convergence rate for nonlinear approximators with respect to linear ones in a certain space of functions; however, such a comparison is made only for functions belonging to a particular space, and for a specific class of nonlinear approximators, i.e., neural networks. Extending a concept introduced by Barron [2], Kůrková [5] defined a norm, called variation of a function with respect to a set of functions, which is assigned to a given class of networks and allows the comparison of the various rates of convergence within a common framework (first results achieved by this approach are in [5], [6]). Moreover, the hypothesis made by De Vore et al. [9] to obtain the Jackson rate for nonlinear approximators, i.e., the constraint of choosing the parameters continuously with respect to the function to be approximated, deserves a detailed analysis. In principle, it may be a strong limitation: we wonder whether nonlinear approximators (which, in contrast to linear ones, have shown to be more suitable for solving complex tasks [12]), take advantage of considering also noncontinous prameters. The results of [4] confirm this intuition.

In this paper, we first discuss some qualitative arguments which support better experimental performance of nonlinear approximators with respect to the linear ones. Then, using the concept of variation of a function with respect to a set, we present some optimal bounds to the approximation error achievable by nonlinear approximators in certain spaces of functions, and we compare them to the rates obtained in the same spaces by linear approximators.
The paper is organized as follows: Section 2 contains some preliminaries on approximation and variation of a function with respect to a set, Section 3 presents qualitative arguments on linear and nonlinear approximation, and Section 4 compares the optimal bounds in the linear and nonlinear case. Section 5 contains some final remarks.

2. Preliminaries

The following definitions are in line with the definitions in [4] and [6].

When the approximating functions form a linear subspace, we call the approximation ***linear approximation***. On the contrary, the approximating functions can be members of unions of finite-dimensional subspaces generated by a given computational unit. In other words, $\mathcal{G} = \{g(.,\theta) : Y \to \mathcal{R}; \theta \in \Theta \subset \mathcal{R}^p\}$, $Y \subset \mathcal{R}^d$, is a parametrized set of functions corresponding to the computational unit g, and we consider all linear combinations of n elements of $\mathcal{G}$. This set, denoted by $span_n\mathcal{G}$, is the union of all linear subspaces spanned by n-tuples of elements of $\mathcal{G}$, i.e., $span_n\mathcal{G} := \{f \in \mathcal{X}; f = \sum_{i=1}^n w_i g_i; w_i \in \mathcal{R}, g_i \in \mathcal{G}\} = \bigcup\{span\{g_1,\ldots,g_n\}; g_i \in \mathcal{G}, i = 1,\ldots,n\}$. In this case we call the approximation ***nonlinear approximation***.

Consider any $f \in span\,\mathcal{G} := \bigcup_{n\in\mathcal{N}} span_n\mathcal{G}$. Denote $\mathcal{G}(b) := \{wg; w \in \mathcal{R}, |w| \leq b, g \in \mathcal{G}\}$. For a subset $\mathcal{G}$ of a normed linear space $(\mathcal{X}, \|.\|)$, ***$\mathcal{G}$-variation*** of $f \in \mathcal{X}$ is

$$V(f,\mathcal{G}) := \inf\{b > 0; f \in cl\,conv\,\mathcal{G}(b)\}.$$

Although the concept of $\mathcal{G}$-variation in general depends on the choice of the norm, to simplify the notation we only write $V(f,\mathcal{G})$ (note that when $\mathcal{X}$ is finite-dimensional all norms on it are equivalent, and so in that case $\mathcal{G}$-variation does not depend on $\|\cdot\|$). In [5] it is shown that $V(\cdot,\mathcal{G})$ is a norm on $\{f \in \mathcal{X}; V(f,\mathcal{G}) < \infty\}$.

3. Some Qualitative Arguments

3.1. Approximation of Real-Valued Boolean Functions

Nonlinear approximators take advantage of their adjustable parameters (for example the centers and the weights matrices in radial basis functions, the frequencies and phases in trigonometric basis functions, the thresholds in sigmoidal functions). As an example, let us consider the space $S_{P(d)}(\{0,1\}^d)$ of real-valued functions with d Boolean variables (i.e. functions $f : \{0,1\}^d \to \mathcal{R}$) such that $f \neq 0$ in at most $P(d)$ points, where P is a polynomial [15]. By using nonlinear approximators, $n = P(d)$ parametric basis functions suffice to represent $f \in S_{P(d)}(\{0,1\}^d)$, since through $\theta \in \Theta$ we can adjust $g_i(\cdot) := g(\cdot,\theta_i)$ for f. On the contrary, fixed basis functions are given in the linear case: for a selected basis, there exists a function in $S_{P(d)}(\{0,1\}^d)$ such that even $n = 2^d - P(d)$ basis functions do not suffice (if $2^d - P(d)$ functions were sufficient then the complementary function, expressible by $P(d)$ basis functions, would have 0 as its best approximation).

3.2. Continuous and Non-Continuous Nonlinear Approximation

Firstly, let us recall the definition of Kolmogorov linear $n-$width. Let $(\mathcal{X}, \|.\|)$ be a normed linear space and $\mathcal{S}$ be a subset of $\mathcal{X}$. The $n-$width in the sense of Kolmogorov (or the *Kolmogorov n-width*) of $\mathcal{S}$ in $\mathcal{X}$ is

$$d_n(\mathcal{S},\mathcal{X}) = \inf_{\mathcal{X}_n} \sup_{x\in\mathcal{S}} \inf_{y\in\mathcal{X}_n} \|x - y\|$$

where the left-most infimum is taken over all n-dimensional subspaces $\mathcal{X}_n$ of $\mathcal{X}$. The constraint to adjust the parameters of the basis functions continuously with respect to the function to be approximated [9] allows to extend the lower bounds achieved by linear $n-$widths [11] to the nonlinear case; this is achieved by defining a proper *continuous nonlinear n-width.* Nonlinear n-width as the alternative to the Kolmogorov linear $n-$width for nonlinear approximation has been first suggested in [4]. *Nonlinear n-width* of $\mathcal{S}$ in $\mathcal{X}$ is defined as

$$\delta_n(\mathcal{S},\mathcal{X}) =$$

$$\inf\{d(\mathcal{S}, span_n\mathcal{G})\}$$

where $\mathcal{G}$ is a member of a family of parametrized subsets of $\mathcal{X}$, $d(\mathcal{S},\mathcal{Y}) = \sup_{x\in\mathcal{S}} \|x - \mathcal{Y}\|$, $\mathcal{S}$ is the set of functions to be approximated, $\mathcal{Y} = span_n\mathcal{G}$ the approximating set and both $\mathcal{S}$ and $\mathcal{Y}$ are subsets of a normed linear space $(\mathcal{X}, \|.\|)$. Arguments based on continuity do not suffice to find good lower bounds on this measure. More specifically, although the continuity requirement is not an essential restriction in a number of cases [9], the results in [4] show that, for many standard types of neural networks (e.g., Gaussian radial-basis-functions and Heaviside perceptrons) and under mild hypotheses on the norm, best approximation cannot be achieved in a continuous way. The following qualitative argument [14] may help to clarify the role of non-continuous dependence of the optimal parameters.

Suppose that in $\mathcal{R}^2$ with the Euclidean norm we want to approximate the points belonging to one of the four quadrants defined by the intersection of two straight lines s and t by means of the points belonging to the union of such lines (i.e., the union of two one-dimensional linear spans). If we consider a point P on the bisecting line of the quadrant and the two segments u and v from such point perpendicular, respectively, to s and t, then the points belonging to $u \setminus \{P\}$ have their best approximation in a point of s, whereas those belonging to $v \setminus \{P\}$ are mapped into a point of t. Consequently, the best approximation mapping is not continuous (note that this is a general problem, when mapping a connected set to a non-connected one). In other words, because of the continuity requirement, by approximating the points in the quadrant through only s or only t we obtain the same result we get by approximating them through $s \cup t$: the continuity constraint does not allow to take advantage of the freedom in choosing the parameters.

3.3. Curse of Dimensionality for Linear Approximators: an Example

Let us consider the class Γ_c of functions $f : \mathcal{R}^d \to \mathcal{R}$ defined by the following bound on the average of the norm of the frequency vector weighted by its Fourier transform:

$$\Gamma_c = \left\{ f : \mathcal{R}^d \to \mathcal{R}; \int_{\mathcal{R}^d} |\underline{\omega}|_K \; |F(\underline{\omega})| \; d\underline{\omega} \leq c \right\}$$

where $K \subset \mathcal{R}^d$ is a compact set, $|\underline{\omega}|_K = \max_{\underline{x} \in K} |\underline{\omega} \cdot \underline{x}|$, F is the Fourier transform of f, and $c \geq 0, c \in \mathcal{R}$.

The following is based on a discussion from [16], where nonlinear approximators have been used to stabilize nonlinear systems. It is shown in [1] that feedforward neural networks with one hidden layer composed of n neural units representing Γ_c have the $\mathcal{L}_2$ approximation error of order $1/n$. On the contrary, there exist functions in Γ_c, for which for any $n > 0$ integer there exist a basis of n functions such that linear combinations of them can have the error of smaller order than $(1/n)^{2/d}$ [1]. The exponent $2/d$ can then cause the "curse of dimensionality". However, such a worst–case performance by linear approximators does not occur for functions characterized by a higher degree of smoothness, like functions with square-integrable partial derivatives of order up to s (hence belonging to Sobolev spaces), provided that s is the least integer greater than $1 + \frac{d}{2}$ [11]; denote such a space by $W_2^{(s)}$. It is shown in [1] that the integral $\int_{\Re^d} |\underline{\omega}| |F(\underline{\omega})| d\underline{\omega}$ is finite for these functions. Then if for $c > 0$ such that $\Gamma_c \supset W_2^{(s)}$ (i.e., $W_2^{(s)}$ is a proper subset of Γ_c), neural approximators should behave better than linear ones in the difference set $\Gamma_c \setminus W_2^{(s)}$. This motivates our choice of feedforward neural networks for finding the approximate solutions of complex problems from control theory ([16], [17], [18]).

4. Comparison of Upper and Lower Bounds for Linear and Nonlinear Approximation

In the following we consider a finite-dimensional Hilbert space $(\mathcal{X}, \|\cdot\|_2)$.

For an orthonormal basis $\mathcal{A}$ of $\mathcal{X}$, we denote by $\|.\|_{1,\mathcal{A}}$, the *l_1-norm with respect to* $\mathcal{A}$, i.e. $\|f\|_{1,\mathcal{A}} = \sum_{i=1}^m |w_i|$, where $f = \sum_{i=1}^m w_i g_i$. It easy to verify that for every $f \in \mathcal{X}$ $V(f, \mathcal{A}) = \|f\|_{1,\mathcal{A}}$, i.e. $\mathcal{A}$-variation is the l_1-norm with respect to $\mathcal{A}$ [6]. The following three theorems have been proved in [6].

Theorem 4.1
Let $(\mathcal{X}, \|.\|_2)$ be a finite-dimensional Hilbert space and let $\mathcal{A}$ be its orthonormal basis. Then for every $f \in \mathcal{X}$ and for every positive integer n there exists $f_n \in span_n\mathcal{A}$ such that

$$\|f - f_n\|_2 \leq \frac{\|f\|_{1,\mathcal{A}}}{2\sqrt{n}}.$$

This implies that $\forall \mathcal{S} \subset \mathcal{X}$, $\delta_n(\mathcal{S}, \mathcal{X}) \leq \frac{\|f\|_{1,\mathcal{A}}}{2\sqrt{n}}$, $\forall f \in \mathcal{S}$ holds for nonlinear n-width.

If the only information available about f is the value of its $\mathcal{A}$-variation, then this upper bound cannot be improved:

Theorem 4.2 *Let $(\mathcal{X}, \|.\|_2)$ be a finite-dimensional Hilbert space, n be a positive integer such that $2n \leq \dim \mathcal{X}$. Then for every* *orthonormal basis $\mathcal{A}$ of $\mathcal{X}$ there exists $f \in \mathcal{X}$ with $\|f\|_{1,\mathcal{A}} = V(f, \mathcal{A})$ so that for every $f_n \in \text{span}_n\mathcal{A}$*

$$\|f - f_n\|_2 \geq \frac{\|f\|_{1,\mathcal{A}}}{2\sqrt{n}}.$$

However, if in addition to $\|f\|_{1,\mathcal{A}}$ also $\|f\|_2$ is known, then the upper bound given by Theorem 4.1 can be improved:

Theorem 4.3 *Let $(\mathcal{X}, \|.\|_2)$ be a finite-dimensional Hilbert space and $\mathcal{A}$ be its orthonormal basis. Then for every $f \in \mathcal{X}$ and for every positive integer n there exists $f_n \in \text{span}_n\mathcal{A}$ such that*

$$\|f - f_n\|_2 \leq \frac{\|f\|_{1,\mathcal{A}}}{2\sqrt{n-1}} \left(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2}\right).$$

This implies that $\forall \mathcal{S} \in \mathcal{X}, \delta_n(\mathcal{S}, \mathcal{X}) \leq \frac{\|f\|_{1,\mathcal{A}}}{2\sqrt{n-1}} \left(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2}\right), \forall f \in \mathcal{S}$ holds for nonlinear n-width.

Theorem 4.3 yields a non-trivial upper bound on $\|f - span_n\mathcal{A}\|_2$ only if

$$\frac{\|f\|_{1,\mathcal{A}}}{2\sqrt{n-1}} \left(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2}\right) < \|f\|_2$$

which is equivalent to

$$\frac{\|f\|_{1,\mathcal{A}}}{\|f\|_2} < \sqrt{n} + \sqrt{n-1}$$

Otherwise, the trivial upper bound equal to $\|f\|_2$ (which corresponds to the approximation of f by the constant zero function) is better. Thus Theorem 4.3 gives non-trivial estimates only for functions with sufficiently large ratio between $\|f\|_{1,\mathcal{A}}$ and $\|f\|_2$. It is shown in [6] that the minimum of $\|f\|_2$ and the bound from Theorem 4.3, yield together a bound on $\|f - span_n\mathcal{A}\|_2$ which is, up to a constant factor, the best possible upper bound expressed in terms of only $\|f\|_{1,\mathcal{A}}$ and $\|f\|_2$.

Using linear approximation for orthonormal sets, we now consider the following lemma and corollary by Kůrková, which are an improvement of Theorem 1.5 by Pinkus [11], in the case of orthonormal subsets of a Hilbert space.

Lemma 4.4 *[14] Let $(\mathcal{X}, \|.\|_2)$ be a Hilbert space and $\mathcal{A}$ be its orthonormal subset of cardinality m. Then for every integer $n \geq 1$, $d_n^2(\mathcal{A}, \mathcal{X}) \geq 1 - \frac{n}{m}$.*

Corollary 4.5 *[14] Let $(\mathcal{X}, \|.\|_2)$ be a Hilbert space and $\mathcal{A}$ an orthonormal subset. Then for every integer $n \geq 1$, $d_n(\mathcal{A}, \mathcal{X}) \geq 1$.*

Comparing Lemma 4.4 with Theorem 4.1, we obtain the following proposition.

Proposition 4.6 *Let $(\mathcal{X}, \|.\|_2)$ be a finite-dimensional Hilbert space and $\mathcal{A}$ its orthonormal basis of cardinality m. Let only $\|f\|_{1,\mathcal{A}}$ be known about $f \in \mathcal{X}$. If $n\left(1 - \frac{n}{m}\right) > \frac{\|f\|_{1,\mathcal{A}}^2}{4}$ then the lower bound on $d_n(\mathcal{A}, \mathcal{X})$ from Lemma 4.4 is bigger than the upper bound on $\delta_n(\mathcal{A}, \mathcal{X})$ from Theorem 4.1.*

In other words, under the hypotheses of Proposition 4.6, nonlinear n-width is smaller than Kolmogorov's n-with, i.e. nonlinear approximation by orthonormal sets in finite-dimensional Hilbert spaces gives better results than linear approximation for $n\left(1 - \frac{n}{m}\right) > \frac{\|f\|_{1,\mathcal{A}}^2}{4}$.

Comparing Lemma 4.4 with Theorem 4.3 and taking into account the condition to get a nontrivial bound from that theorem, we obtain the following result.

Proposition 4.7
Let $(\mathcal{X}, \|.\|_2)$ be a finite-dimensional Hilbert space and $\mathcal{A}$ its orthonormal basis of cardinality m. Let $\|f\|_{1,\mathcal{A}}$ and $\|f\|_2$ be known about $f \in \mathcal{X}$. If $n > \frac{1}{\|f\|_2^2} \frac{\|f\|_{1,\mathcal{A}}^2}{4} \left(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2}\right)^2 + 1$ and $n\left(1 - \frac{n}{m}\right) > \frac{\|f\|_{1,\mathcal{A}}^2}{4} \left(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2}\right)^2 + 1$ then the bound on $d_n(\mathcal{A}, \mathcal{X})$ from Lemma 4.4 is bigger than the upper bound on $\delta_n(\mathcal{A}, \mathcal{X})$ from Theorem 4.3.

In other words, under the hypotheses of Proposition 4.7, nonlinear n-width is smaller than Kolmogorov's n-with, i.e. nonlinear approximation by orthonormal sets in finite-dimensional Hilbert spaces gives better results than linear approximation for n, m satisfying $n > \frac{1}{\|f\|_2^2} \frac{\|f\|_{1,\mathcal{A}}^2}{4} \left(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2}\right)^2 + 1$ and $n\left(1 - \frac{n}{m}\right) > \frac{\|f\|_{1,\mathcal{A}}^2}{4} \left(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2}\right)^2 + 1$. Note that for $n = m$ Lemma 4.4 gives the trivial result $d_n^2(\mathcal{A}, \mathcal{X}) \geq 0$; obviously, in this case the upper bound on nonlinear approximation can not be smaller than the lower bound (i.e. zero) on linear approximation.

Theorem 4.3 gives a nontrivial bound if $\frac{\|f\|_{1,\mathcal{A}}^2}{4\|f\|_2^2}(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2})^2 = V \cdot W < n - 1$,
where $V = \frac{1}{\|f\|_2^2}$, $W = \frac{\|f\|_{1,\mathcal{A}}^2}{4}(1 - \frac{\|f\|_2^2}{\|f\|_{1,\mathcal{A}}^2})^2$.

On the other hand, the upper bound on nonlinear n-width from Theorem 4.3 is smaller than the lower estimate on d_n from Lemma 4.4 if $\frac{\|f\|^2_{1,\mathcal{A}}}{4(n-1)}(1 - \frac{\|f\|^2_2}{\|f\|^2_{1,\mathcal{A}}})^2 < 1 - \frac{n}{m}$, i.e. $W = \frac{\|f\|^2_{1,\mathcal{A}}}{4}(1 - \frac{\|f\|^2_2}{\|f\|^2_{1,\mathcal{A}}})^2 < (n-1)\frac{m-n}{m}$.
$V.W < n - 1$ and $W < (n-1)\frac{m-n}{m}$ gives together $V \leq \frac{m-n}{m}$, i.e. $\frac{1}{\|f\|^2_2} \leq \frac{m-n}{m}$.
Then, if $\|f\|^2_2 > 1 - \frac{n}{m}$, only the condition $n > \frac{1}{\|f\|^2_2} \frac{\|f\|^2_{1,\mathcal{A}}}{4} \left(1 - \frac{\|f\|^2_2}{\|f\|^2_{1,\mathcal{A}}}\right)^2 + 1$ must be satisfied.
Note that $\|f\|^2_2 > 1 - \frac{n}{m}$ is easily fulfilled for values of m close to n.

5. Concluding remarks

The definition of a proper nonlinear n-width, analogous to the Kolmogorov n-width for the linear case, enables the comparison of the performance of linear and nonlinear approximators. This comparison supports the superiority of nonlinear approximation (as suggested by qualitative arguments), at least for orthonormal sets of approximators in finite-dimensional Hilbert spaces. The analysis in more general cases (i.e. in infinite dimensional spaces and generally non orthonormal sets of approximators) is of high interest and is the object of current research. The comparison does not seem easy as the known upper estimates in these cases are based on knowledge of variation of the function to be approximated with respect to the set of approximators ([6]).

Acknowledgements

We are grateful to Věra Kůrková for valuable discussions and remarks. Kateřina Hlaváčková-Schindler was partially supported by GA AV grant A2075606 and grant CR 201/96/0917. Marcello Sanguineti was partially supported by grant CNR 96.02472.CT07.

References

[1] Barron, A.R.: Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory* 39, 930-945, 1993.

[2] Barron, A. R.: Neural net approximation. *Proc. 7th Yale Workshop on Adaptive and Learning Systems.* K. Narendra Ed., Yale University Press, 1992.

[3] Jones, L.K.: A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training. *Annals of Statistics* 20, 608-613, 1992.

[4] Kainen, P.C., Kůrková, V., Vogt, A: Approximation by neural networks is not continuous. Submitted to *Neurocomputing*.

[5] Kůrková, V.: Dimension-independent rates of approximation by neural networks. *Computer-intensive methods in Control and Signal Processing: Curse of Dimensionality* (Eds. K. Warwick, M. Kárný). Birkhauser, Boston, 261-270, 1997.

[6] Kůrková, V., Savický, P., Hlaváčková, K.: Representations and rates of approximation of real valued Boolean functions by neural networks. *Neural Networks*, in print.

[7] Mhaskar, H.N., Micchelli, C.A.: Approximation by superposition of sigmoidal and radial basis functions. *Advances in Applied Mathematics* 13, 350-373, 1992.

[8] Kůrková, V.: Kolmogorov's theorem and multilayer neural networks. *Neural Networks*, 5, 501-506, 1992.

[9] DeVore, R. Howard, R., Micchelli, C.: Optimal nonlinear approximation. *Manuscripta Mathematica*, 469-478, 1989.

[10] Mhaskar, H.N.: Neural networks for optimal approximation of smooth and analytic functions. *Neural Computation* 8, 164-177, 1996.

[11] Pinkus, A.: *N− Widths in Approximation Theory.* Springer-Verlag, New York, 1986.

[12] Sejnowski, T.J., Rosenberg, C.: Parallel networks that learn to pronounce English text. *Complex Systems* 1, 145-168, 1987.

[13] Girosi, F.: Regularization theory, radial basis functions and networks. In *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*, Subseries F, Computer and Systems Sciences, V. Cherkassky, J.H. Friedman and H. Wechsler Eds., Springer-Verlag, 1994.

[14] From a discussion with V. Kůrková.

[15] From a discussion with P. Savický.

[16] Parisini, T., Sanguineti, M., Zoppoli, R.: Nonlinear stabilization by receding-horizon neural regulators. *International Journal of Control*, in print.

[17] Alessandri, A., Sanguineti, M., Zoppoli, R., Parisini, T.: Neural nonlinear controllers and observers: stability results. *Proceedings of WIRN-Vietri '96*, 1996, pp. 80-90.

[18] Parisini, T., Polycarpou, M., Sanguineti, M., Vemuri, A.: Robust parametric and nonparametric fault diagnosis in nonlinear input-output systems. *Proceedings of the 36th Con-*

ference on Decision and Control, 1997, pp. 4481-4482.

Computer-Intensive Methods in Control and Data Processing

Can We Beat the Curse of Dimensionality?

CMP'98

Preprints of the 3rd European IEEE Workshop CMP'98
Prague, Czech Republic, September 7-9, 1998

edited by

J. Rojíček, M. Valečková, M. Kárný, K. Warwick

The 3rd European IEEE Workshop on

Computer-Intensive Methods in Control and Data Processing

Can We Beat the Curse of Dimensionality?

September 7–9, 1998

Prague, Czech Republic

Edited by

Jiří Rojíček, Markéta Valečková, Miroslav Kárný
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic

and

Kevin Warwick
School of Engineering & Information Sciences
University of Reading, U.K.

Participating Society
IEEE Control Systems Society