

Surrogate Modelling of Solutions of Integral Equations by Neural Networks

Věra Kůrková

► **To cite this version:**

Věra Kůrková. Surrogate Modelling of Solutions of Integral Equations by Neural Networks. 8th International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2012, Halkidiki, Greece. pp.88-96, 10.1007/978-3-642-33409-2_10 . hal-01521408

HAL Id: hal-01521408

<https://hal.inria.fr/hal-01521408>

Submitted on 11 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Surrogate Modelling of Solutions of Integral Equations by Neural Networks

Věra Kůrková

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, Prague 8, Czech Republic
vera@cs.cas.cz

Abstract. Surrogate modelling of solutions of integral equations by neural networks is investigated theoretically. Estimates of speed of convergence of suboptimal surrogate solutions to solutions described by Fredholm theorem are derived.

Keywords: surrogate modelling by neural networks, Fredholm integral equations, rates of approximation, model complexity

1 Introduction

Surrogate modelling of functional relationships by neural networks has been successfully applied to empirical functions for which experimental evaluations are too expensive or to functions, which are described by known analytic formulas but numerical calculations of such formulas are too time consuming. For construction of surrogate models, results of such experiments or calculations merely at some selected points of the domains of the functions are needed. Values obtained by such expensive or time consuming procedures are then used as training samples of data for neural networks. The networks trained on such samples are called surrogate models. For example, neural networks have been successfully used in chemistry as surrogate models of empirical functions assigning to compositions of chemicals measures of quality of catalyzers produced by reactions of these chemicals or in biology as surrogate models of empirical functions classifying structures of RNA [1, 2].

In the case of empirical functions, results obtained by surrogate modelling can only be used as suggestions to be confirmed by additional experiments as no other than empirical knowledge of the functions is available. In the case of functions whose mathematical expressions are known, although complicated and difficult to calculate numerically, these expressions might enable to estimate how close to these functions their surrogate models are. In particular, theoretical analysis can provide estimates of model complexity of neural networks needed for a sufficient accuracy of approximation of the functions which are replaced by these networks.

Recently, possibilities of neural networks in surrogate modelling of solutions of Fredholm integral equations were explored. Fredholm equations play an important role in many problems in applied science and engineering. They arise in image restoration, differential problems with auxiliary boundary conditions,

potential theory and elasticity, etc. (see, e.g., [3–5]). Numerical calculations of values of solutions of Fredholm equations are time consuming as they involve computations of complicated expressions in terms of infinite Liouville-Neumann series with integrals coefficients. Thus various methods of approximate solutions of Fredholm equations have been developed. Traditional approach employed polynomial or wavelet interpolations of numerically calculated values at certain collocation points [6, Chapter 11]. As an alternative to the traditional approach, surrogate models formed by neural networks with perceptrons [7] and Gaussian radial-basis units [8] were proposed and explored experimentally. These experiments were motivated by higher flexibility of neural networks than flexibility of polynomials. Neural networks have more adjustable parameters than linear models as in addition to coefficients of linear combinations of basis functions, also inner coefficients of computational units are optimized during learning. Thus they are sometimes called variable-basis approximation schemas in contrast to traditional linear approximators which are called fixed-basis approximation schemas. In some cases, especially in the case of functions of large numbers of variables, neural networks achieve better approximation rates than linear models with much smaller model complexity [9, 10]. Motivated by the experimental exploration [8, 7] of surrogate models of solutions of Fredholm equations by neural networks, Gnecco et al. [11] initiated a theoretical analysis of this modelling. In [11], they estimated approximation errors in supremum norm for surrogate solutions by networks with kernel units induced by the kernels of the equations.

In this paper, we investigate surrogate solutions of Fredholm integral equation by networks with general computational units. Taking advantage of results from nonlinear approximation theory and integral representations of functions, we estimate how well surrogate solutions computable by neural networks can approximate exact solutions of Fredholm equations described by the Fredholm alternative theorem. We show that such surrogate solutions converge to the exact solutions and we give upper bounds on decrease of approximation errors measured in \mathcal{L}^2 -norm with increasing model complexity of approximating networks.

The paper is organized as follows. In section 2, we introduce Fredholm integral equations and their surrogate solutions by neural networks. In section 3, using results from nonlinear approximation theory, we derive estimates of rates of approximation of solutions of Fredholm equations by neural networks. Section 4 is a brief discussion.

2 Solutions of Fredholm Equations

Solving an *inhomogeneous Fredholm integral equation of the second kind* on a domain $X \subseteq \mathbb{R}^d$ for a given $\lambda \in \mathbb{R} \setminus \{0\}$, $K : X \times X \rightarrow \mathbb{R}$, and $f : X \rightarrow \mathbb{R}$ is a task of finding a function $\phi : X \rightarrow \mathbb{R}$ such that for all $x \in X$

$$\phi(x) - \lambda \int_X \phi(y)K(x, y) dy = f(x). \quad (1)$$

The function ϕ is called a *solution*, f *data*, K a *kernel*, and a λ *parameter* of the equation (1).

Recall that tasks of finding *unknown causes of known consequences* have been studied in applied science under the name *inverse problems*. Formally, an inverse problem is defined by a linear operator $A : \mathcal{X} \rightarrow \mathcal{Y}$ between two function spaces. It is a task of finding for $g \in \mathcal{Y}$ (called *data*) some $f \in \mathcal{X}$ (called *solution*) such that $A(f) = g$.

Let T_K denotes the integral operator with a kernel $K : X \times X \rightarrow \mathbb{R}$ defined for every ϕ in a suitable function space as

$$T_K(\phi)(x) := \int_X \phi(y) K(x, y) dy. \quad (2)$$

Thus solving the Fredholm equation (1) is an inverse problem described by the linear operator $I_{\mathcal{X}} - \lambda T_K$, where $I_{\mathcal{X}}$ is the identity operator. Existence and uniqueness of solutions of this inverse problem follows from the Fredholm alternative theorem (see, e.g., [12, p.499]).

Theorem 1. *Let \mathcal{X} be a Banach space, $T : \mathcal{X} \rightarrow \mathcal{X}$ be a compact operator, and $I_{\mathcal{X}}$ be the identity operator. Then the operator $I_{\mathcal{X}} + T : \mathcal{X} \rightarrow \mathcal{X}$ is one-to-one if and only if it is onto.*

Theorem 1 implies that when T is a compact operator and $1/\lambda$ is not its eigenvalue (i.e., there is no $\phi \in \mathcal{X}$ for which $T(\phi) = \frac{\phi}{\lambda}$), then the operator $I_{\mathcal{X}} - \lambda T$ is invertible (one-to-one and onto) [13, p. 112], [14, Section 1.3].

Fredholm theorem can be applied to spaces $(\mathcal{C}(X), \|\cdot\|_{\text{sup}})$ of bounded continuous functions on $X \subseteq \mathbb{R}^d$ with the supremum norm $\|f\|_{\text{sup}} = \sup_{x \in X} |f(x)|$ and to spaces $(\mathcal{L}^2(X), \|\cdot\|_{\mathcal{L}^2})$ of square integrable functions with the norm $\|f\|_{\mathcal{L}^2} = (\int_X f(x)^2 dx)^{1/2}$. The following proposition is well-known and easy to check (see, e.g., [13, p. 112]).

Proposition 1. *(i) If $X \subset \mathbb{R}^d$ is compact and $K : X \times X \rightarrow \mathbb{R}$ is continuous, then $T_K : \mathcal{C}(X) \rightarrow \mathcal{C}(X)$ is a compact operator.*

(ii) If $X \subset \mathbb{R}^d$ and $K \in \mathcal{L}^2(X \times X)$, then $T_K : \mathcal{L}^2(X) \rightarrow \mathcal{L}^2(X)$ is a compact operator.

So by the Fredholm theorem, when $1/\lambda$ is not an eigenvalue of T_K and the assumptions of the Proposition 1 (i) or (ii) are satisfied, then for every f in $\mathcal{C}(X)$ or $\mathcal{L}^2(X)$, resp., there exists unique solution ϕ of the equation (1). Moreover,

$$\phi(x) = f(x) - \lambda \int_X f(y) R_K^\lambda(x, y) dy, \quad (3)$$

where $R_K^\lambda : X \times X \rightarrow \mathbb{R}$ is called a *resolvent kernel* [14]. Although a formula expressing the resolvent kernel is known, it is not suitable for efficient computation as it is expressed as an infinite Neumann series in powers of λ with coefficients in the form of iterated kernels [15, p.140]. Numerical calculations of values of solutions of Fredholm equations based on (3) are quite computationally demanding and so various methods of finding surrogate solutions of (1) have been explored.

A traditional approach employed linear methods such as polynomial interpolation of approximations $\bar{\phi}(x_1), \dots, \bar{\phi}(x_m)$ (computed by numerical methods) of the values $\phi(x_1), \dots, \phi(x_m)$ of the solution ϕ at suitable points x_1, \dots, x_m from the domain X . In linear methods, approximating sets of functions are n -dimensional subspaces of the form

$$\text{span}\{g_1, \dots, g_n\} := \left\{ \sum_{i=1}^n w_i g_i \mid w_i \in \mathbb{R} \right\}, \quad (4)$$

where $G = \{g_n \mid n \in \mathbb{N}_+\}$ is a set of functions with a *fixed linear ordering*. Nested linear subspaces of the form (4) can be generated, e.g., by sets of powers of increasing degrees or trigonometric functions with increasing frequencies.

A recent alternative to linear methods employs neural networks. Recall that one-hidden-layer networks with one linear output unit compute input-output functions from sets of the form

$$\text{span}_n G := \left\{ \sum_{i=1}^n w_i g_i \mid w_i \in \mathbb{R}, g_i \in G \right\}, \quad (5)$$

where the set G is sometimes called a *dictionary* [16] and n is the *number of hidden computational units*. This number can be interpreted as a measure of *model complexity* of the network. In contrast to linear approximation, the approximation scheme formed by sets $\text{span}_n G$ is called *variable-basis approximation* (the set G has no fixed ordering).

Often, dictionaries are parameterized families of functions modelling computational units, i.e., they are of the form

$$G_K(X, Y) := \{K(\cdot, y) : X \rightarrow \mathbb{R} \mid y \in Y\}, \quad (6)$$

where $K : X \times Y \rightarrow \mathbb{R}$ is a function of two variables, an input vector $x \in X \subseteq \mathbb{R}^d$ and a parameter $y \in Y \subseteq \mathbb{R}^s$. When $X = Y$, we write briefly $G_K(X)$.

In some contexts, K is called a *kernel*. However, the above-described computational scheme includes fairly general computational models, such as functions computable by perceptrons, radial or kernel units, Hermite functions, trigonometric polynomials, and splines. For example, with

$$K(x, y) = K(x, (v, b)) := \sigma(\langle v, x \rangle + b)$$

and $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ a sigmoidal function, the computational scheme (5) describes one-hidden-layer *perceptron networks*. *RBF units* with an activation function $\beta : \mathbb{R} \rightarrow \mathbb{R}$ are modeled by kernel

$$K(x, y) = K(x, (v, b)) := \beta(v\|x - b\|).$$

Typical choice of β is the Gaussian function. *Kernel units* used in support vector machine have the form $K(x, y)$ where $K : X \times X \rightarrow \mathbb{R}$ is a symmetric positive semidefinite function [12].

Various learning algorithms optimize parameters y_1, \dots, y_n of the computational units as well as coefficients of their linear combinations w_1, \dots, w_n so that network input-output functions $\sum_{i=1}^n w_i K(\cdot, y_i)$ from the set $\text{span}_n G_K$ fit well to training samples. In the case of surrogate solutions of Fredholm equations, training samples are the above described sets $\{(x_1, \bar{\phi}(x_1)), \dots, (x_m, \bar{\phi}(x_m))\}$ of input-output pairs formed by selected points of the domain together with numerical approximations of values of the solution ϕ at these points.

3 Rates of Convergence of Surrogate Solutions

Estimates of model complexity of one-hidden-layer networks approximating solutions of Fredholm equations follow from inspection of upper bounds on rates of variable-basis approximation. Such rates have been studied in mathematical theory of neurocomputing for various types of dictionaries G and norms measuring approximation errors such as Hilbert-space norms and the supremum norm (see, e.g., [17, 18]). Typically, such bounds have the form $\frac{\xi(h)}{\sqrt{n}}$, where n is the number of network units and $\xi(h)$ depends on a certain variational norm of the function h to be approximated, which is tailored to the dictionary of the network units. Variational norm is defined quite generally for any bounded nonempty subset G of a normed linear space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$. It is called G -variation, denoted $\|\cdot\|_G$, and defined for all $f \in \mathcal{X}$ as

$$\|f\|_{G, \mathcal{X}} := \inf \{c > 0 \mid f/c \in \text{cl}_{\mathcal{X}} \text{conv}(G \cup -G)\},$$

where the closure $\text{cl}_{\mathcal{X}}$ is taken with respect to the topology generated by the norm $\|\cdot\|_{\mathcal{X}}$ and conv denotes the *convex hull*. So G -variation depends on the ambient space norm, but when it is clear from the context, we write merely $\|f\|_G$ instead of $\|f\|_{G, \mathcal{X}}$.

The concept of variational norm was introduced by Barron [19] for sets of characteristic functions. In particular, for the set of characteristic functions of half-spaces, the variational norm is tailored to the dictionary of functions computable by Heaviside perceptrons. Barron's concept was generalized in [20, 21] to the concept of variation with respect to an arbitrary bounded set of functions and applied to various dictionaries of computational units such as Gaussian RBF or kernel units [22].

The following theorem from [21] is a reformulation of results by Maurey [23], Jones [24], Barron [17] in terms of G -variation. It gives an upper bound on rates of approximation by sets of the form $\text{span}_n G$. For a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$, $g \in \mathcal{X}$ and $A \subset \mathcal{X}$, we denote by $\|g - A\|_{\mathcal{X}} := \inf_{f \in A} \|g - f\|_{\mathcal{X}}$ the *distance* of g from A .

Theorem 2. *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a Hilbert space, G its bounded nonempty subset, $s_G = \sup_{g \in G} \|g\|_{\mathcal{X}}$, $f \in \mathcal{X}$, and n be a positive integer. Then*

$$\|h - \text{span}_n G\|_{\mathcal{X}}^2 \leq \frac{s_G^2 \|h\|_G^2 - \|h\|_{\mathcal{X}}^2}{n}.$$

Theorem 2 implies that for every $\varepsilon > 0$ and n satisfying

$$n \geq \left(\frac{s_G \|h\|_G}{\varepsilon} \right)^2,$$

a network with n units computing functions from the dictionary G approximates the function h within ε . So the size of the G -variation of the function h is a critical factor influencing model complexity of networks approximating h within an accuracy ε . Generally, it is not easy to estimate G -variation. However, for the special case of functions with integral representations in the form of “infinite networks”, variational norms are bounded from above by the \mathcal{L}^1 -norms of “output-weight” functions of these networks.

Theorem 3. *Let $X \subseteq \mathbb{R}^d$, $Y \subseteq \mathbb{R}^s$, $w \in \mathcal{L}^1(Y)$, $K : X \times Y \rightarrow \mathbb{R}$ be such that $G_K(Y) = \{K(\cdot, y) \mid y \in Y\}$ is a bounded subset of $(\mathcal{L}^2(X), \|\cdot\|_{\mathcal{L}^2})$, and $h \in \mathcal{L}^2(X)$ be such that for all $x \in X$, $h(x) = \int_Y w(y) K(x, y) dy$. Then*

$$\|f\|_{G_K(Y)} \leq \|w\|_{\mathcal{L}^1(Y)}.$$

Theorem 3 gives an upper bound on variation with respect to the dictionary G_K induced by the kernel K of the integral operator T_K which maps an “output-weight” function w of the “infinite network” $\int_Y w(y) K(\cdot, y) dy$ to the function h . However, to apply Theorem 2 to approximation of solutions of Fredholm equations by surrogate models formed by networks with units from a general dictionary G , we need upper bounds on G -variation for functions expressed by an integral formula with the kernel K . The next proposition describes a relationship between variations with respect to two sets, G and F . Its application to G and G_K gives an upper bound on G . The proof of the next proposition follows easily from the definition of variational norm.

Proposition 2. *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a normed linear space, F and G its bounded subsets such that $c_{G,F} := \sup_{g \in G} \|g\|_F < \infty$. Then for all $h \in \mathcal{X}$, $\|h\|_G \leq c_{G,F} \|h\|_F$.*

Combining Theorems 2, 3, and Proposition 2, we obtain the next theorem on rates of approximation of functions which can be expressed as $h = T_K(w)$ by networks with units from a dictionary G .

Theorem 4. *Let $X \subseteq \mathbb{R}^d$, $K : X \times Y \rightarrow \mathbb{R}$ be a bounded kernel, and $h \in \mathcal{L}^2(X)$ such that $h = T_K(w) = \int_Y w(y) K(\cdot, y) dy$ for some $w \in \mathcal{L}^1(Y)$, where $G_K(X, Y)$ is a bounded subset of $\mathcal{L}^2(X)$. Let G be a bounded subset of $\mathcal{L}^2(X)$ with $s_G = \sup_{g \in G} \|g\|_{\mathcal{L}^2}$ such that $c_{G,K} = \sup_{g \in G} \|g\|_{G_K}$ is finite. Then for all $n > 0$,*

$$\|h - \text{span}_n G\|_{\mathcal{L}^2} \leq \frac{s_G c_{G,K} \|w\|_{\mathcal{L}^1(Y)}}{\sqrt{n}}.$$

A critical factor in the upper bound from Theorem 4 is the \mathcal{L}^1 -norm of the “output-weight function” w in the representation

$$h(x) = T_K(w)(x) = \int_Y w(y) K(x, y) dy$$

of h as an “infinite network” with units computing $K(\cdot, y)$.

The solution ϕ of the Fredholm equation minus the function f representing the data, $\phi - f$, is the image of $\lambda \phi$ mapped by the integral operator T_K , i.e.,

$$\phi - f = \lambda \int_X f(y) R_K^\lambda(x, y) dy = \lambda \int_X \phi(y) K(x, y) dy = T_K(\lambda \phi).$$

Thus to apply Theorem 4 to approximation of $\phi - f$, we need to estimate the \mathcal{L}^1 -norm of the solution ϕ itself as $\lambda \phi$ plays the role of the “output-weight” function in the infinite network $\lambda \int_X \phi(y) K(x, y) dy$.

Theorem 5. *Let $X \subset \mathbb{R}^d$ be compact, $K : X \times X \rightarrow \mathbb{R}$ be a bounded kernel such that $K \in \mathcal{L}^2(X \times X)$, $s_K := \sup_{y \in Y} \|K(\cdot, y)\|_{\mathcal{L}^2}$, $\rho_K := \int_X \sup_{y \in X} |K(x, y)| dx$ be finite, G be a bounded subset of $\mathcal{L}^2(X)$ with $s_G = \sup_{g \in G} \|g\|_{\mathcal{L}^2}$ such that $c_{G,K} = \sup_{g \in G} \|g\|_{G_K}$ is finite, and $\lambda \neq 0$ be such that $\frac{1}{\lambda}$ is not an eigenvalue of T_K and $|\lambda| \rho_K < 1$. Then the solution ϕ of the equation (1) satisfies for all $n > 0$,*

$$\|\phi - f - \text{span}_n G\|_{\mathcal{L}^2} \leq \frac{s_G c_{G,K} |\lambda| \|f\|_{\mathcal{L}^1}}{(1 - |\lambda| \rho_K) \sqrt{n}}.$$

Proof. By (1), we have for every $x \in X$,

$$|\phi(x)| \leq |\lambda| \|\phi\|_{\mathcal{L}^1} \sup_{y \in X} |K(x, y)| + |f(x)|.$$

Hence $\|\phi\|_{\mathcal{L}^1} \leq |\lambda| \rho_K \|\phi\|_{\mathcal{L}^1} + \|f\|_{\mathcal{L}^1}$ and so $\|\phi\|_{\mathcal{L}^1} (1 - |\lambda| \rho_K) \leq \|f\|_{\mathcal{L}^1}$. This inequality is non trivial only when $|\lambda| < \frac{1}{\rho_K}$. Thus we get $\|w\|_{\mathcal{L}^1} = |\lambda| \|\phi\|_{\mathcal{L}^1} \leq \frac{|\lambda| \|f\|_{\mathcal{L}^1}}{1 - |\lambda| \rho_K}$. The statement then follows from Theorem 4. \square

Theorem 5 estimates rates of approximation of the function $\phi - f$ by functions computable by networks with units from a dictionary G . The values of the function $\lambda \int_X f(y) R_K^\lambda(x, y) dy$ are difficult to be compute numerically. However, $\phi - f$ can be replaces with its approximation computable by a neural network. With increasing number n of units, input-output functions of such networks converge to the function $\phi - f$. When for a reasonable size of the network measured by the number n of units, the upper bound from Theorem 5 is sufficiently small, the network can serve as a good surrogate model of the solution of Fredholm equation.

The next corollary illustrates our results by an example of approximation of solutions of Fredholm equations with Gaussian kernels by Gaussian kernel and perceptron networks. Fredholm equations with Gaussian kernels arise, e.g., in image restoration problems [5]. By μ is denoted the Lebesgue measure on \mathbb{R}^d and by $P_d^\sigma(X)$ the dictionary of functions on $X \subseteq \mathbb{R}^d$ computable by sigmoidal perceptrons.

Corollary 1. *Let $X \subset \mathbb{R}^d$ be compact, $b > 0$, $K_b(x, y) = e^{-b\|x-y\|^2}$, $\lambda \neq 0$ be such that $\frac{1}{\lambda}$ is not an eigenvalue of T_{K_b} and $|\lambda| < 1$. Then the solution ϕ of the equation $\phi - f = \lambda \int_X \phi(y)K(x, y) dy$ with f continuous satisfies for all $n > 0$*

$$\|\phi - f - \text{span}_n G_{K_b}(X)\|_{\mathcal{L}^2} \leq \frac{\mu(X) |\lambda| \|f\|_{\mathcal{L}^1}}{(1 - |\lambda| \mu(X)) \sqrt{n}}$$

and

$$\|\phi - f - \text{span}_n P_d^\sigma(X)\|_{\mathcal{L}^2} \leq \frac{\mu(X) 2d |\lambda| \|f\|_{\mathcal{L}^1}}{(1 - |\lambda| \mu(X)) \sqrt{n}}.$$

Proof. It was shown in [25] that P_d^σ -variation (i.e., variation with respect to the dictionary formed by sigmoidal perceptrons) of the d -dimensional Gaussian is bounded from above by $2d$. The statement then follows by Theorem 5, Proposition 2, equalities $\tau_{K_b} = 1$, $\rho_{K_b} = \mu(X)$, and the estimate $s_{K_b} \leq \mu(X)$. \square

4 Discussion

Taking advantage of results from mathematical theory of neurocomputing holding for functions representable as “infinite neural networks” we derived estimates of rates of convergence of surrogate solutions of Fredholm equations to exact solutions guaranteed by Fredholm theory. The bounds are in the form $\frac{c}{\sqrt{n}}$ where n is the number of network units and c depends on \mathcal{L}^1 -norm of the function f corresponding to data of the inverse problem given by the operator $I - \lambda T_K$, the parameter λ , the Lebesgue measure $\mu(X)$ of the domain $X \subset \mathbb{R}^d$, and a bound on $\|K(\cdot, y)\|_{\mathcal{L}^2}$. The results hold under certain restrictions on the size of λ similarly as in the case of traditional linear approximation of solutions of Fredholm equations.

Acknowledgments. This work was partially supported by MŠMT grant COST INTELLI OC10047 and RVO 67985807.

References

1. Forrester, A., Sobester, A., Keane, A.: Engineering Design via Surrogate Modelling: A Practical Guide. Wiley (2008)
2. Baerns, M., Holeňa, M.: Combinatorial Development of Solid Catalytic Materials. Imperial College Press, London (2009)
3. Lovitt, W.V.: Linear Integral Equations. Dover, New York (1950)
4. Lonseth, A.T.: Sources and applications of integral equations. SIAM Review **19** (1977) 241–278
5. Lu, Y., Shen, L., Xu, Y.: Integral equation models for image restoration: high accuracy methods and fast algorithms. Inverse Problems **26** (2010) doi: 10.1088/0266-5611/26/4/045006.

6. Atkinson, K., Han, W.: *Theoretical Numerical Analysis: A Functional Analysis Framework*. Springer (2005)
7. S. Effati, R.B.: A neural network approach for solving Fredholm integral equations of the second kind. *Neural Computing and Applications* (2010) doi:10.1007/s00521-010-0489-y
8. Golbabai, A., Seifollahi, S.: Numerical solution of the second kind integral equations using radial basis function networks. *Applied Mathematics and Computation* **174** (2006) 877–883
9. Gnecco, G., Kůrková, V., Sanguineti, M.: Some comparisons of complexity in dictionary-based and linear computational models. *Neural Networks* **24** (2011) 171–182
10. Gnecco, G., Kůrková, V., Sanguineti, M.: Can dictionary-based computational models outperform the best linear ones? *Neural Networks* **24** (2011) 881–887
11. Gnecco, G., Kůrková, V., Sanguineti, M.: Accuracy of approximations of solutions to Fredholm equations by kernel methods. *Applied Mathematics and Computation* **218** (2012) 74817497
12. Steinwart, I., Christmann, A.: *Support Vector Machines*. Springer, New York (2008)
13. Rudin, W.: *Functional Analysis*. McGraw-Hill, Boston (1991)
14. Atkinson, K.: *The Numerical Solution of Integral Equations of the Second Kind*. Cambridge University Press (1997)
15. Courant, R., Hilbert, D.: *Methods of Mathematical Physics. Volume I*. Wiley, New York (1989)
16. Gribonval, R., Vandergheynst, P.: On the exponential convergence of matching pursuits in quasi-incoherent dictionaries. *IEEE Transactions on Information Theory* **52** (2006) 255–261
17. Barron, A.R.: Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory* **39** (1993) 930–945
18. Girosi, F.: Approximation error bounds that use VC-bounds. In: *Proceedings of ICANN 1995, Paris (1995)* 295–302
19. Barron, A.R.: Neural net approximation. In Narendra, K., ed.: *Proc. 7th Yale Workshop on Adaptive and Learning Systems*, Yale University Press (1992)
20. Kůrková, V.: Dimension-independent rates of approximation by neural networks. In Warwick, K., Kárný, M., eds.: *Computer-Intensive Methods in Control and Signal Processing. The Curse of Dimensionality*. Birkhäuser, Boston, MA (1997) 261–270
21. Kůrková, V.: High-dimensional approximation and optimization by neural networks. In Suykens, J., Horváth, G., Basu, S., Micchelli, C., Vandewalle, J., eds.: *Advances in Learning Theory: Methods, Models and Applications*. IOS Press, Amsterdam (2003) 69–88 (Chapter 4)
22. Kainen, P.C., Kůrková, V., Sanguineti, M.: Complexity of Gaussian radial-basis networks approximating smooth functions. *Journal of Complexity* **25** (2009) 63–74
23. Pisier, G.: Remarques sur un résultat non publié de B. Maurey. In: *Séminaire d'Analyse Fonctionnelle 1980-81, vol. I, no. 12*, École Polytechnique, Centre de Mathématiques, Palaiseau, France (1981)
24. 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** (1992) 608–613
25. Kainen, P.C., Kůrková, V., Vogt, A.: A Sobolev-type upper bound for rates of approximation by linear combinations of Heaviside plane waves. *Journal of Approximation Theory* **147** (2007) 1–10