Surrogate solutions of Fredholm equations by feedforward networks*

Věra Kůrková

Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague
vera@cs.cas.cz, http://www.cs.cas.cz/∼vera

Abstract. *Surrogate solutions of Fredholm integral equations by feedforward neural networks are investigated theoretically. Convergence of surrogate solutions computable by networks with increasing numbers of computational units to theoretically optimal solutions is proven and upper bounds on rates of convergence are derived. The results hold for a variety of computational units, they are illustrated by examples of perceptrons and Gaussian radial units.*

1 Introduction

Surrogate modeling is one of successful applications of neural networks. Often it has been used for empirical functions, i.e., functions for which no mathematical formulas are known and thus their values can only be gained experimentally. When such experimental evaluations are too expensive or time consuming, it can be useful to perform them merely for some samples of points of the domains of the empirical functions and the obtained values use as training data for neural networks. The networks trained on such data can play roles of *surrogate models* of these empirical functions. For example, input-output functions of feedforward networks have been 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, in biology as models of empirical functions classifying structures of RNA, and in economy as models of functions assigning credit ratings to companies [7, 2]. However, it should be emphasized that results obtained by surrogate modeling of empirical functions can only be used as suggestions to be confirmed by additional experiments as no other than empirical knowledge of the functions is available. Moreover, no methodology for choice of suitable network architectures, type of computational units and their number has been developed.

In contrast to the case of empirical functions, analytically described functions, which are subjects of surrogate modeling due to their complicated and time consuming numerical calculations, provide a potential for theoretical analysis of quality of their surrogate models. Available analytic expressions can be compared with various surrogate models. One can investigate mathematical properties of these functions as well as properties of their surrogate models aiming to estimate speed of convergence of approximations computable by surrogate models with increasing model complexity to functions described by the complicated formulas. Mathematical theory of approximation of functions by neural networks offers some tools for derivation of such estimates.

A large class of functions described by mathematical formulas, numerical calculations of which are difficult, is formed by solutions of Fredholm integral equations. These 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., [23, 22, 24]). Mathematical descriptions of solutions of Fredholm equations following from classical Fredholm theorem [27, p.499] involve complicated expressions in terms of infinite Liouville-Neumann series with coefficients in the forms of integrals. Thus numerical calculations of these expressions are time consuming.

Recently, several authors [13, 6] explored experimentally possibilities of surrogate modeling of solutions of Fredholm equations by perceptron and kernel networks. Motivated by these experimental studies, in [9] we initiated a theoretical analysis of approximation of solutions of Fredholm equtions by neural networks. In [9, 12, 20], estimates of rates of approximation were derived for surrogate modeling by networks with kernel units induced by the same kernels as the kernels defining the equations and extended to certain smooth kernels.

In this paper, we investigate surrogate solutions of Fredholm integral equations by networks with general computational units. Taking advantage of results from nonlinear approximation theory and suitable integral representations of functions in the form of "infinite" networks, we estimate how well surrogate solutions computable by feedforward networks can approximate exact solutions of Fredholm equations. We apply general results to perceptron and Gaussian radial networks.

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

The paper is organized as follows. In section 2, we describe surrogate modeling of functions by feedforward neural networks and in section 3, we introduce Fredholm integral equations and theoretical approach to their solutions. In section 4, we recall some results from nonlinear approximation theory and apply them to approximation of solutions of Fredholm equations by feedorward networks. We illustrate our results by an example of approximation of Fredholm equations with the Gaussian kernel by networks with perceptrons and Gaussian radial units.

2 Surrogate modeling by neural networks

A traditional approach to surrogate modeling of functions has employed linear methods such as polynomial interpolation. For suitable points $x_1, \ldots, x_m$ from a domain $X \subset \mathbb{R}^d$, empirically or numerically obtained approximations $\bar{\phi}(x_1), \ldots, \bar{\phi}(x_m)$ of values $\phi(x_1), \ldots, \phi(x_m)$ of a function ϕ are interpolated by functions from n-dimensional function spaces. These spaces are obtained as *linear spans*

$$\mathrm{span}\{g_1, \ldots, g_n\} := \left\{ \sum_{i=1}^{n} w_i g_i \,|\, w_i \in \mathbb{R} \right\}, \qquad (1)$$

where the functions $g_1, \ldots g_n$ are first n elements from a set $G = \{g_n \,|\, n \in \mathbb{N}_+\}$ with a *fixed linear ordering* (we use the standard notation := meaning a definition). Typical examples of linear approximators are algebraic or trigonometric polynomials. They are obtained by linear combinations of powers of increasing degrees or trigonometric functions with increasing frequencies, resp.

Feedforward 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 approximation of functions of large numbers of variables, it was proven that neural networks achieve better approximation rates than linear models with much smaller model complexity [11, 10].

One-hidden-layer networks with one linear output unit compute input-output functions from sets of the form

$$\mathrm{span}_n\, G := \left\{ \sum_{i=1}^{n} w_i g_i \,|\, w_i \in \mathbb{R}, g_i \in G \right\}, \qquad (2)$$

where the set G is sometimes called a *dictionary* [14] 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 dictionary G has no fixed ordering.

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

$$G_F(X,Y) := \{F(\cdot, y) : X \to \mathbb{R} \,|\, y \in Y\}\ , \qquad (3)$$

where $F{:}X \times Y \to \mathbb{R}$ 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 $G_F(X)$. So one-hidden-layer networks with n units from a dictionary $G_F(X,Y)$ compute functions from the set

$$\mathrm{span}_n\, G_F(X,Y) := \left\{ \sum_{i=1}^{n} w_i F(x, y_i) \,|\, w_i \in \mathbb{R}, y_i \in Y \right\}.$$

In some contexts, F 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

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

and $\sigma : \mathbb{R} \to \mathbb{R}$ a sigmoidal function, the computational scheme (2) describes one-hidden-layer *perceptron networks*. *Radial (RBF) units* with an activation function $\beta : \mathbb{R} \to \mathbb{R}$ are modeled by the kernel

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

Typical choice of β is the Gaussian function. *Kernel units* used in support vector machine (SVM) have the form $F(x,y)$ where $F : X \times X \to \mathbb{R}$ is a symmetric positive semidefinite function [27].

Various learning algorithms optimize parameters $y_1, \ldots, y_n$ of computational units as well as coefficients $w_1, \ldots, w_n$ of their linear combinations so that network input-output functions

$$\sum_{i=1}^{n} w_i\, F(., y_i)$$

from the set $\mathrm{span}_n\, G_F(X,Y)$ fit well to training samples $\{(x_i, \bar{\phi}(x_i)\,|i = 1, \ldots, m\}$.

3 Solutions of Fredholm integral equations

Solving an *inhomogeneous Fredholm integral equation of the second kind* on a domain $X \subseteq \mathbb{R}^d$ for a given $\lambda \in \mathbb{R} \setminus \{0\}$, $K : X \times X \to \mathbb{R}$, and $f : X \to \mathbb{R}$ is

a task of finding a function $\phi : X \to \mathbb{R}$ such that for all $x \in X$

$$\phi(x) - \lambda \int_X \phi(y) K(x,y)\, dy = f(x). \tag{4}$$

The function ϕ is called *solution*, f *data*, K *kernel*, and λ *parameter* of the equation (4).

Fredholm equations can be described in terms of theory of inverse problems. Formally, an *inverse problem* is defined by a linear operator $A : \mathcal{X} \to \mathcal{Y}$ between two function spaces. It is a task of finding for $f \in \mathcal{Y}$ (called *data*) some $\phi \in \mathcal{X}$ (called *solution*) such that

$$A(\phi) = f.$$

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

$$T_K(\phi)(x) := \int_X \phi(y)\, K(x,y)\, dy \tag{5}$$

and $I_{\mathcal{X}}$ denotes the identity operator. Then the Fredholm equation (4) can be represented as an inverse problem defined by the linear operator $I_{\mathcal{X}} - \lambda T_K$. So it is a problem of finding for a given data f a solution ϕ such that

$$(I_{\mathcal{X}} - \lambda T_K)(\phi) = f. \tag{6}$$

The classical Fredholm alternative theorem from 1903 proved existence and uniqueness of solutions of Fredholm equations for continuous one variable functions on intervals. A modern version holding for general Banach spaces is stated in the next theorem from [27, p.499]. Recall that an operator $T : (\mathcal{X}, \|.\|_{\mathcal{X}}) \to (\mathcal{Y}, \|.\|_{\mathcal{Y}})$ between two Banach spaces is called *compact* if it maps bounded sets to precompact sets (i.e., sets whose closures are compact).

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

A straightforward corollary of Theorem 1 guarantees existence and uniqueness of solutions of the inverse problem (6) when T is a compact operator and $1/\lambda$ is not its eigenvalue (i.e., there is no $\phi \in \mathcal{X}$ for which $T(\phi) = \frac{\phi}{\lambda}$).

Corollary 1. *Let $(\mathcal{X}, \|.\|_{\mathcal{X}})$ be a Banach space, $T : (\mathcal{X}, \|.\|_{\mathcal{X}}) \to (\mathcal{X}, \|.\|_{\mathcal{X}})$ be a compact operator, $I_{\mathcal{X}}$ be the identity operator, and $\lambda \neq 0$ be such that $1/\lambda$ is not an eigenvalue of T. Then the operator $I_{\mathcal{X}} - \lambda T$ is invertible (one-to-one and onto).*

The following proposition gives conditions guaranteeing compactness of operators T_K in spaces $(\mathcal{C}(X), \|.\|_{\sup})$, where $X \subseteq \mathbb{R}^d$, of bounded continuous functions on X with the supremum norm $\|f\|_{\sup} = \sup_{x \in X} |f(x)|$ and to spaces $(\mathcal{L}^2(X), \|.\|_{\mathcal{L}^2})$ of square integrable functions with the norm $\|f\|_{\mathcal{L}^2} = \left(\int_X f(x)^2\, dx\right)^{1/2}$. The proof is well-known and easy to check (see, e.g., [26, p. 112]).

Proposition 1. *(i) If $X \subset \mathbb{R}^d$ is compact and $K : X \times X \to \mathbb{R}$ is continuous, then $T_K : (\mathcal{C}(X), \|.\|_{\sup}) \to (\mathcal{C}(X), \|.\|_{\sup})$ is a compact operator.*
(ii) If $X \subset \mathbb{R}^d$ and $K \in \mathcal{L}^2(X \times X)$, then $T_K : (\mathcal{L}^2(X), \|.\|_{\mathcal{L}^2}) \to (\mathcal{L}^2(X), \|.\|_{\mathcal{L}^2})$ is a compact operator.

So by Corollary 1, when the assumptions of the Proposition 1 (i) or (ii) are satisfied and $1/\lambda$ is not an eigenvalue of T_K, then for every f in $\mathcal{C}(X)$ or $\mathcal{L}^2(X)$, resp., there exists unique solution ϕ of the equation (4). It is known (see, e.g, [1]) that the solution ϕ can be expressed as

$$\phi(x) = f(x) - \lambda \int_X f(y)\, R_K^{\lambda}(x,y)\, dy\,, \tag{7}$$

where $R_K^{\lambda} : X \times X \to \mathbb{R}$ is called a *resolvent kernel* . However, the formula expressing the resolvent kernel 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 integrals with iterated kernels [5, p.140]. So numerical calculations of values of solutions of Fredholm equations based on (7) are quite computationally demanding. Thus various methods of finding surrogate solutions of (4) have been explored [13, 6]. Some of these methods utilized feedforward networks. Such networks were trained on samples of input-output pairs $\{(x_1, \bar{\phi}(x_1)), \ldots, (x_m, \bar{\phi}(x_m)\}$ where $\{x_1, \ldots, x_m\}$ are selected points from the domain X and $\{\bar{\phi}(x_1), \ldots, \bar{\phi}(x_m)\}$ are numerically computed approximations of values $\{\phi(x_1), \ldots, \phi(x_m)\}$ of the solution ϕ. In these experiments, one-hidden-layer networks with perceptrons and Gaussian radial units were used. However, without a theoretical analysis, it is not clear how to choose a proper number n of network units to guarantee that input-output functions approximate well the solution and the networks are not too large to make their implementation unfeasible.

4 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 decrease of errors in approximation of solutions of the equation (4) by sets $\mathrm{span}_n G$ with n increasing. Approximation properties of sets of the form $\mathrm{span}_n G$ 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., [4, 8]). Some 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 norm of the function h to be approximated.

This norm is tailored to the dictionary of computation units and can be estimated for functions satisfying suitable integral equations. The norm is defined quite generally for any bounded nonempty subset G of a normed linear space $(\mathcal{X}, \|.\|_{\mathcal{X}})$. It is called *$G$-variation*, denoted $\|.\|_G$, and defined for all $f \in \mathcal{X}$ as

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

where the closure $\mathrm{cl}_{\mathcal{X}}$ is taken with respect to the topology generated by the norm $\|.\|_{\mathcal{X}}$ 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 [3] for sets of characteristic functions. Among them, the set of characteristic functions of half-spaces forming the dictionary of functions computable by Heaviside perceptrons. Barron's concept was generalized in [18, 19] to variation with respect to an arbitrary bounded set of functions and applied to various dictionaries of computational units such as Gaussian RBF units or kernel units [16].

The following theorem on rates of approximation by sets of the form $\mathrm{span}_n G$ is a reformulation from [19] of results by Maurey [25], Jones [15], Barron [4] in terms of G-variation. For a normed space $(\mathcal{X}, \|.\|_{\mathcal{X}})$, $g \in \mathcal{X}$ 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 $(\mathcal{X}, \|.\|_{\mathcal{X}})$ 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*

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

Theorem 2 guarantees 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 G-variation of the function h to be approximated is a critical factor influencing model complexity of networks approximating h within a required accuracy. Generally, it is not easy to estimate G-variation. However, the following theorem from [21] shows that 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 $X \subseteq \mathbb{R}^d$, $Y \subseteq \mathbb{R}^s$, $w \in \mathcal{L}^1(Y)$, $K : X \times Y \to \mathbb{R}$ be such that $G_K(X,Y) = \{K(.,y) \mid y \in Y\}$ is a bounded subset of $(\mathcal{L}^2(X), \|.\|_{\mathcal{L}^2})$, 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*

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

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. The next proposition describes a relationship between variations with respect to two sets, G and F; its proof follows easily from the definition of variational norm.

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

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

A critical factor in the estimate given in Theorem 4 is the $\mathcal{L}^1$-norm of the "output-weight function" w in the representation of the function h to be approximated an "infinite network" with units computing $K(.,y)$ in the form

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

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.,

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

Thus to apply Theorem 4 to approximation of a solution of Fredholm equation, 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 $\int_X \lambda\,\phi(y)\,K(x,y)\,dy$.

Theorem 5. *Let $X \subset \mathbb{R}^d$ be compact, $K : X \times X \to \mathbb{R}$ be a bounded kernel such that $K \in \mathcal{L}^2(X \times X)$, $\rho_K := \int_X \sup_{y\in X} |K(x,y)|dx$ 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 $c_{G,K} = \sup_{g\in G}\|g\|_{G_K(X)}$ 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 (4) satisfies for all $n > 0$,*

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

Proof. As $\phi - f$ satisfies the Fredholm equation (4), we have for every $x \in X$,

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

Integrating over X we get

$$\|\phi\|_{\mathcal{L}^1} \leq |\lambda|\,\rho_K\,\|\phi\|_{\mathcal{L}^1} + \|f\|_{\mathcal{L}^1}$$

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. □

Theorem 5 estimates rates of approximation of the function $\phi - f = \lambda \int_X f(y)\,R^\lambda_K(x,y)\,dy$ by functions computable by networks with units from dictionary G formed by functions with G_K-variations bounded by $c_{G,K}$. Numerical computations of values of the function $\lambda \int_X f(y)\,R^\lambda_K(x,y)\,dy$ are time consuming. For $|\lambda| < \frac{1}{\rho_K}$ and any bounded dictionary G with finite bound $c_{G,K}$ on $G_{K(X)}$-variations on its elements, input-output functions of networks with increasing numbers of units from G 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.

To illustrate our results, consider approximation of Fredholm equations with the Gaussian kernel

$$K_b(x,y) = e^{-b\|x-y\|}$$

with the width b by surrogate solutions in the form of input-output functions of networks with two types of popular units: sigmoidal perceptrons and Gaussian radial units. Note that Fredholm equations with Gaussian kernels arise, e.g., in image restoration problems [24]. By μ is denoted the *Lebesgue measure* on $\mathbb{R}^d$ and by $P^\sigma_d(X)$ the *dictionary of functions on X computable by sigmoidal perceptrons*.

Corollary 2. *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 (4) with f continuous satisfies for all $n > 0$*

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

and

$$\|\phi - f - \mathrm{span}_n\, P^\sigma_d(X)\|_{\mathcal{L}^2} \leq \frac{\mu(X)\,2d\,|\lambda|\,\|f\|_{\mathcal{L}^1}}{(1 - |\lambda|\,\mu(X)\,)\,\sqrt{n}}\,.$$

Proof. It was shown in [17] that variation of the d-dimensional Gaussian with respect to the dictionary formed by sigmoidal perceptrons is bounded from above by $2d$ and thus by Proposition 2, $c_{P^\sigma_d,K_b} \leq 2d$. The statement then follows by Theorem 5, an estimate $s_{G_{K_b}} \leq \mu(X)$ and equalities $s_{P^\sigma_d} = \mu(X)$ and $\rho_{K_b} = \mu(X)$. □

References

1. K. Atkinson: *The numerical solution of integral equations of the second kind.* Cambridge University Press, 1997.
2. M. Baerns, M. Holeňa: *combinatorial development of solid catalytic materials.* Imperial College Press, London, 2009.
3. A. R. Barron: *Neural net approximation.* In K. Narendra, (Ed.), Proc. 7th Yale Workshop on Adaptive and Learning Systems, Yale University Press, 1992.
4. A. R. Barron: *Universal approximation bounds for superpositions of a sigmoidal function.* IEEE Transactions on Information Theory 39, 1993, 930–945.
5. R. Courant, D. Hilbert: *Methods of mathematical physic*, volume I. Wiley, New York, 1989.
6. S. Effati, R. Buzhabadi: *A neural network approach for solving Fredholm integral equations of the second kind.* Neural Computing and Applications. doi:10.1007/s00521-010-0489-y.
7. A. Forrester, A. Sobester, A. Keane: *Engineering design via surrogate modelling: A practical guide.* Wiley, 2008.
8. F. Girosi: *Approximation error bounds that use VC-bounds.* In Proceedings of ICANN 1995, Paris, 1995, 295–302.

9. G. Gnecco, V. Kůrková, M. Sanguineti: *Bounds for approximate solutions of Fredholm integral equations using kernel networks.* In T. Honkela and et al., (Eds), Lecture Notes in Computer Science (Proceedings of ICANN 2011), vol. 6791, Springer, Heidelberg, 2011, 126–133.
10. G. Gnecco, V. Kůrková, M. Sanguineti: *Can dictionary-based computational models outperform the best linear ones?* Neural Networks 24, 2011, 881–887.
11. G. Gnecco, V. Kůrková, M. Sanguineti: *Some comparisons of complexity in dictionary-based and linear computational models.* Neural Networks, 24, 2011, 171–182.
12. G. Gnecco, V. Kůrková, M. Sanguineti: *Accuracy of approximations of solutions to Fredholm equations by kernel methods.* Applied Mathematics and Computation 218, 2012, 7481–7497.
13. A. Golbabai, S. Seifollahi: *Numerical solution of the second kind integral equations using radial basis function networks.* Applied Mathematics and Computation 174, 2006, 877–883.
14. R. Gribonval, P. Vandergheynst: *On the exponential convergence of matching pursuits in quasi-incoherent dictionaries.* IEEE Transactions on Information Theory 52, 2006, 255–261.
15. L. K. Jones: *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.
16. P. C. Kainen, V. Kůrková, M. Sanguineti: *Complexity of Gaussian radial-basis networks approximating smooth functions.* Journal of Complexity 25, 2009, 63–74.
17. P. C. Kainen, V. Kůrková, A. Vogt: *A Sobolev-type upper bound for rates of approximation by linear combinations of Heaviside plane waves.* Journal of Approximation Theory 147, 2007, 1–10.
18. V. Kůrková: *Dimension-independent rates of approximation by neural networks.* In K. Warwick and M. Kárný, (Eds), Computer-Intensive Methods in Control and Signal Processing. The Curse of Dimensionality, Birkhäuser, Boston, MA, 1997, 261–270.
19. V. Kůrková: *High-dimensional approximation and optimization by neural networks.* In J. Suykens, G. Horváth, S. Basu, C. Micchelli, J. Vandewalle, (Eds), Advances in Learning Theory: Methods, Models and Applications, (Chapter 4). IOS Press, Amsterdam, 2003, 69–88.
20. V. Kůrková: *Accuracy estimates for surrogate solutions of integral equations by neural networks.* In M. Bieliková, G. Fridrich, G. Gottlob, S. Katzenbeisser, R. Špánek, G. Turán, (Eds), SOFSEM 2012: Theory and Practice of Computer Science, vol. II,. Institute of Computer Science, Prague, 2012 , 95–102.
21. V. Kůrková: *Complexity estimates based on integral transforms induced by computational units.* Neural Networks, 30:160–167, 2012.
22. A. T. Lonseth: *Sources and applications of integral equations.* SIAM Review 19, 1977, 241–278.
23. W. V. Lovitt: *Linear integral equations.* Dover, New York, 1950.
24. Y. Lu, L. Shen, Y. Xu: *Integral equation models for image restoration: high accuracy methods and fast algorithms.* Inverse Problems, 26. doi: 10.1088/0266-5611/26/4/045006.
25. G. Pisier: *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.
26. W. Rudin: *Functional analysis.* McGraw-Hill, Boston, 1991.
27. I. Steinwart, A. Christmann: *Support vector machines.* Springer, New York, 2008.