Journal of Complexity 25 (2009) 63–74

Contents lists available at ScienceDirect

Journal of Complexity

journal homepage: www.elsevier.com/locate/jco

Complexity of Gaussian-radial-basis networks approximating smooth functions

Paul C. Kainen [a], Věra Kůrková [b], Marcello Sanguineti [c,*]

[a] Department of Mathematics, Georgetown University, Washington, DC, 20057-1233, USA

[b] Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 182 07, Prague 8, Czech Republic

[c] Department of Communications, Computer, and System Sciences (DIST), University of Genoa, Via Opera Pia 13, 16145 Genova, Italy

ARTICLE INFO

Article history:
Received 4 November 2007
Accepted 20 August 2008
Available online 7 September 2008

Keywords:
Gaussian-radial-basis-function networks
Rates of approximation
Model complexity
Variation norms
Bessel and Sobolev norms
Tractability of approximation

ABSTRACT

Complexity of Gaussian-radial-basis-function networks, with varying widths, is investigated. Upper bounds on rates of decrease of approximation errors with increasing number of hidden units are derived. Bounds are in terms of norms measuring smoothness (Bessel and Sobolev norms) multiplied by explicitly given functions $a(r, d)$ of the number of variables d and degree of smoothness r. Estimates are proven using suitable integral representations in the form of networks with continua of hidden units computing scaled Gaussians and translated Bessel potentials. Consequences on tractability of approximation by Gaussian-radial-basis function networks are discussed.

1. Introduction

Radial-basis function (RBF) networks with Gaussian computational units are known to be able to approximate with an arbitrary accuracy all continuous and all $\mathcal{L}^2$-functions on compact subsets of $\mathbb{R}^d$ [10,23,27,30,31]. In such approximations, the number n of RBF units plays the role of a measure of model complexity and its size determines the feasibility of network implementation.

Several authors have investigated rates of approximation by RBF networks with n Gaussians units of *fixed width*. Girosi and Anzellotti [9] derived an asymptotic upper bound of order $n^{-1/2}$ on approximation error measured by the supremum norm for band-limited functions with continuous derivatives up to order r with $r > d/2$, where d is the number of variables [9, p. 106]. Using results

[*] Corresponding author.
E-mail addresses: kainen@georgetown.edu (P.C. Kainen), vera@cs.cas.cz (V. Kůrková), marcello@dist.unige.it (M. Sanguineti).

0885-064X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jco.2008.08.001

from statistical learning theory, Girosi [8] extended these bounds to more general classes of kernels. For Gaussians of *varying* widths, Kon, Raphael, and Williams [14, Corollary 3] obtained bounds on a weighted $\mathcal{L}^\infty$-distance from the target function to a linear combination of Gaussians.

Some bounds improve on the exponent of $-1/2$. Mhaskar [24,25] and Narcovich et al. [29] obtained bounds of order $n^{-r/2d}$, and in one special case, Maiorov [21] found $n^{-r/(d-1)}$. Although order with respect to n improves, the remaining multiplicative factors in such bounds involve constants that are unknown and these upper bounds increase as d increases. Also, they apply to different classes of target and approximating functions. Moreover, dependence on parameters may differ, and approximation error is computed with respect to different metrics. Thus, it is not easy to compare these bounds.

In this paper, we approximate smooth functions by Gaussian RBF networks with units of varying widths, using $\mathcal{L}^2$-distance with respect to the Lebesgue measure. We derive upper bounds on rates of approximation in terms of the Bessel and Sobolev norms of the functions to be approximated. Bessel norms are defined in terms of convolutions with Bessel-potential kernels, while Sobolev norms use integrals of partial derivatives. The Bessel norm $\|\cdot\|_{L^{2,r}}$ and the Sobolev norm $\|\cdot\|_{W^{2,r}}$ are equivalent but the ratio between them depends on the number d of variables.

Our estimates hold for all numbers n of hidden units and all degrees $r > d/2$ of Bessel potentials. The estimates are of the form $n^{-1/2}$ times the Bessel norm $\|f\|_{L^{1,r}}$ of the function f to be approximated times a factor $k(r, d)$. For a fixed $c > 0$ and the degree $r_d = d/2 + c$, the factor $k(r_d, d)$ decreases to zero exponentially fast. We also derive estimates in terms of $\mathcal{L}^2$ Bessel and Sobolev norms. Our results show that reasonably smooth functions can be approximated quite efficiently by Gaussian-radial-basis networks. A preliminary version of the results appeared in [13].

The paper is organized as follows. Section 2 presents some concepts, notations, and auxiliary results for studying approximation by Gaussian RBF networks. Section 3 derives upper bounds on rates of approximation of Bessel potentials by linear combinations of scaled Gaussians in terms of variation norms obtained from integral representations of Bessel potentials and their Fourier transforms. In Section 4, for functions representable as convolutions with Bessel potentials, upper bounds are derived in terms of Bessel-potential norms. These bounds are then combined with estimates of variational norms from the previous section to obtain bounds for approximation by Gaussian RBFs in terms of Bessel norms. Section 5 uses the relationship between Sobolev and Bessel norms to obtain bounds in terms of Sobolev norms. In Section 6, we discuss consequences for tractability of multivariate approximation by Gaussian-radial-basis networks.

2. Approximation by Gaussian RBF networks

For $\Omega \subseteq \mathbb{R}^d$, $\mathcal{L}^2(\Omega)$ denotes the space of real-valued functions on Ω with norm $\|f\|_{\mathcal{L}^2(\Omega)} = \left(\int |f(x)|^2 \mathrm{d}x\right)^{1/2}$. Two functions are identified if they differ only on a set of Lebesgue-measure zero. When $\Omega = \mathbb{R}^d$, we omit it in the notation.

For nonzero $f \in \mathcal{L}^2$, $f^o = f/\|f\|_{\mathcal{L}^2}$ denotes the *normalization* of f; for convenience, we put $0^o = 0$. For $F \subset \mathcal{L}^2$, $F|_\Omega$ denotes the set of functions from F restricted to Ω, $\hat{F}$ the set of Fourier transforms of functions in F, and F^o the set of their normalizations. For $n \geq 1$, define

$$\mathrm{span}_n F := \left\{ \sum_{i=1}^{n} w_i f_i \mid f_i \in F,\ w_i \in \mathbb{R} \right\}.$$

In this paper, we investigate accuracy measured by $\mathcal{L}^2$-norm with respect to the Lebesgue measure λ in approximation by Gaussian-radial-basis-function networks.

A *Gaussian-radial-basis-function unit* with d inputs computes all *scaled and translated Gaussian functions* on $\mathbb{R}^d$. For $b > 0$, let $\gamma_b : \mathbb{R}^d \to \mathbb{R}$ denote the Gaussian function of *width b* defined by

$$\gamma_b(x) = \mathrm{e}^{-b\|x\|^2}.$$

A simple calculation shows that

$$\|\gamma_b\|_{\mathcal{L}^2} = (\pi/2b)^{d/4}. \tag{1}$$

Let

$$G_0 = \{\gamma_b \mid b > 0\}$$

denote the *set of the Gaussians centered at* 0 *with varying widths*. For τ_y the *translation operator* defined for any $y \in \mathbb{R}^d$ and any f on $\mathbb{R}^d$ as $(\tau_y f)(x) = f(x-y)$, let

$$G = \left\{\tau_y \gamma_b \mid y \in \mathbb{R}^d, b > 0\right\}$$

denote the *set of all translations of the Gaussians with varying widths*.

We investigate rates of approximation by *networks with n Gaussian RBF units and one linear output unit*, which compute functions from the set $\text{span}_n G$.

We exploit properties of the Fourier transform of the Gaussian function. The *d-dimensional Fourier transform* is the operator $\mathcal{F}$ on $\mathcal{L}^2 \cap \mathcal{L}^1$ given by

$$\mathcal{F}(f)(s) = \hat{f}(s) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{ix\cdot s} f(x)\, dx, \tag{2}$$

where $\cdot$ denotes the Euclidean inner product on $\mathbb{R}^d$.

For every $b > 0$,

$$\widehat{\gamma_b}(x) = (2b)^{-d/2} \gamma_{\frac{1}{4b}}(x) \tag{3}$$

(cf. [34, p. 43]). Thus

$$\text{span}_n G_0 = \text{span}_n \widehat{G_0}. \tag{4}$$

Plancherel's identity [34, p. 31] asserts that the Fourier transform is an isometry on $\mathcal{L}^2$, i.e., for all $f \in \mathcal{L}^2$

$$\|f\|_{\mathcal{L}^2} = \|\hat{f}\|_{\mathcal{L}^2}, \tag{5}$$

and directly by (1) we have

$$\|\gamma_b\|_{\mathcal{L}^2} = \left(\frac{\pi}{2b}\right)^{d/4} = \|\widehat{\gamma_b}\|_{\mathcal{L}^2}. \tag{6}$$

In a normed linear space $(\mathcal{X}, \|.\|_{\mathcal{X}})$, for $f \in \mathcal{X}$ and $A \subset \mathcal{X}$,

$$\|f - A\|_{\mathcal{X}} = \inf_{g \in A} \|f - g\|_{\mathcal{X}}$$

denotes the distance from f to A. The following proposition shows that in estimating rates of approximation by linear combinations of scaled Gaussians centered at 0, one can switch between a function and its Fourier transform.

Proposition 2.1. *For all positive integers* d, n *and all* $f \in \mathcal{L}^2$, $\|f - \text{span}_n G_0\|_{\mathcal{L}^2} = \|f - \text{span}_n \widehat{G_0}\|_{\mathcal{L}^2} = \|\hat{f} - \text{span}_n \widehat{G_0}\|_{\mathcal{L}^2} = \|\hat{f} - \text{span}_n G_0\|_{\mathcal{L}^2}$.

Proof. Using (5) and (4), respectively, we get $\|f - \text{span}_n G_0\|_{\mathcal{L}^2} = \|\hat{f} - \text{span}_n \hat{G_0}\|_{\mathcal{L}^2} = \|f - \text{span}_n \hat{G_0}\|_{\mathcal{L}^2} = \|\hat{f} - \text{span}_n G_0\|_{\mathcal{L}^2}$. □

To derive our estimates, we use a result on approximation by convex combinations of n elements of a bounded subset of a Hilbert space derived by Maurey [32], Jones [11] and Barron [2,3]. Let F be a bounded subset of a Hilbert space $(\mathcal{H}, \|.\|_{\mathcal{H}})$, and

$$\text{uconv}_n F = \left\{\frac{1}{n}\sum_{i=1}^{n} f_i \mid f_i \in F\right\}$$

denote the set of n-fold convex combinations of elements of F with all coefficients equal. By the result of Maurey–Jones–Barron [3, p. 934], for every function h in cl conv $(F \cup -F)$, i.e., in the closure of the symmetric convex hull of F, we have

$$\|h - \mathrm{uconv}_n F\|_{\mathcal{H}} \leq n^{-1/2} \left(s_F^2 - \|h\|_{\mathcal{H}}^2\right)^{1/2}, \tag{7}$$

where $s_F = \sup_{f \in F} \|f\|_{\mathcal{H}}$. The bound (7) implies an estimate of the distance from $\mathrm{span}_n F$ holding for any function from $\mathcal{H}$. The estimate is formulated in terms of a norm tailored to F, called F-*variation*, which was introduced in [15] as an extension of "variation with respect to half-spaces" defined in [2].

For any bounded subset F of any normed linear space $(\mathcal{X}, \|.\|_{\mathcal{X}})$, F-variation is defined as the *Minkowski functional of the closed convex symmetric hull of* F (where closure is taken with respect to the norm $\|.\|_{\mathcal{X}}$). The variational norm with respect to F in $\mathcal{X}$ is denoted by $\|.\|_{F,\mathcal{X}}$, i.e.,

$$\|h\|_{F,\mathcal{X}} = \inf\left\{c > 0 \mid c^{-1}h \in \mathrm{cl\,conv}(F \cup -F)\right\}. \tag{8}$$

Note that F-variation can be infinite (when the set on the right-hand side is empty) and that it depends on the ambient space norm. When we consider variation with respect to the $\mathcal{L}^2$-norm, we omit $\mathcal{L}^2$ in the notation of variational norm.

The Maurey–Jones–Barron estimate (7) implies that for any bounded subset F of a Hilbert space $(\mathcal{H}, \|.\|_{\mathcal{H}})$ and all positive integers n

$$\|h - \mathrm{span}_n F\|_{\mathcal{H}} \leq n^{-1/2} \left(\|h\|_{F^o,\mathcal{H}}^2 - \|h\|_{\mathcal{H}}^2\right)^{1/2} \tag{9}$$

(see [16]). To apply the upper bound (9) to approximation by Gaussian RBFs we take advantage of properties of variational norms given in the remainder of this section.

From the definitions, if ψ is any linear isometry of $(\mathcal{X}, \|.\|_{\mathcal{X}})$, then for any $f \in \mathcal{X}$, $\|f\|_{F,\mathcal{X}} = \|\psi(f)\|_{\psi(F),\mathcal{X}}$. In particular,

$$\|f\|_{G_0^o,\mathcal{X}} = \|\hat{f}\|_{G_0^o,\mathcal{X}}. \tag{10}$$

Variations with respect to two subsets satisfy the following inequality [19, Proposition 3(iii)].

Lemma 2.2. *Let F, H be nonempty, nonzero subsets of a normed linear space $(\mathcal{X}, \|.\|_{\mathcal{X}})$ and $s_{H,F} := \sup_{h \in H} \|h\|_{F,\mathcal{X}}$. Then for every $f \in \mathcal{X}$,*

$$\|f\|_{F,\mathcal{X}} \leq s_{H,F} \|f\|_{H,\mathcal{X}}.$$

The next lemma states that the variation of the limit of a sequence of functions is bounded by the limit of variations (see [18, Lemma 7.2]) and [12, Lemma 3.5]).

Lemma 2.3. *Let F be a nonempty bounded subset of a normed linear space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$, $h \in \mathcal{X}$, and $\{h_i\} \subset \mathcal{X}$ such that $\lim_{i\to\infty} \|h_i - h\|_{\mathcal{X}} = 0$. For all i, let $b_i = \|h_i\|_{F,\mathcal{X}} < \infty$ and suppose that there exists $\lim_{i\to\infty} b_i = b$. Then $\|h\|_{F,\mathcal{X}} \leq b$.*

Variation with respect to a parameterized family of functions can be estimated for functions representable by a suitable integral formula, where integration is with respect to the parameter. Let $\Omega \subseteq \mathbb{R}^d$, $\phi : \Omega \times Y \to \mathbb{R}$. If for all $y \in Y$, $\phi(., y) \in \mathcal{L}^2(\Omega)$, then we denote by $\Phi : Y \to \mathcal{L}^2$ the mapping defined for every $y \in Y$ as $\Phi(y) = \phi(., y)$ and

$$\Phi(Y) := \{\phi(., y) : \Omega \to \mathbb{R} \mid y \in Y\}.$$

The following theorem was proven in [12, Corollary 5.1] using properties of Bochner integral of the mapping Φ together with the limit property of variational norms given in Lemma 2.3. We denote by $w\Phi : Y \to \mathcal{L}^2(\Omega)$ the mapping defined for all $y \in Y$ as $w\Phi(y) = w(y)\Phi(y)$.

Theorem 2.4. *Let $\Omega \subseteq \mathbb{R}^d$ be Lebesgue measurable, $f \in \mathcal{L}^2(\Omega)$ be such that for a.e. $x \in \Omega$,*

$$f(x) = \int_Y w(y)\phi(x, y)\mathrm{d}y,$$

where Y, w, and ϕ satisfy the following three conditions:

(i) *$Y \subseteq \mathbb{R}^p$ is Lebesgue measurable, p is a positive integer, $Y \setminus Y_0 = \cup_{m=1}^{\infty} Y_m$, where $\lambda(Y_0) = 0$ and for all positive integers m, Y_m is compact and $Y_m \subseteq Y_{m+1}$,*
(ii) *$\Phi(Y)$ is a bounded subset of $\mathcal{L}^2(\Omega)$, $w \in \mathcal{L}^1(Y)$, and $w\Phi : Y \setminus Y_0 \to \mathcal{X}$ is continuous,*
(iii) *$\phi : \Omega \times Y \to \mathbb{R}$ is Lebesgue measurable.*

Then $\|f\|_{\Phi(Y)} \leq \|w\|_{\mathcal{L}^1(Y)}$ and for all positive integers n,

$$\|f - \mathrm{span}_n \Phi(Y)\|^2_{\mathcal{L}^2(\Omega)} \leq \frac{s_\Phi^2 \|w\|^2_{\mathcal{L}^1(Y)} - \|f\|^2_{\mathcal{L}^2(\Omega)}}{n}.$$

Theorem 2.4 guarantees that if f can be represented as a neural network with a continuum of hidden units computing functions from $\Phi(Y)$, then the $\Phi(Y)$-variational norm of f is bounded by the $\mathcal{L}^1$-norm of the weight function.

3. Approximation of Bessel potentials by Gaussian RBFs

In this section, we estimate rates of approximation by $\mathrm{span}_n G$ for certain special functions, called Bessel potentials, which are defined by means of their Fourier transforms. For $r > 0$, the *Bessel potential* of order r, denoted by β_r, is the function on $\mathbb{R}^d$ with Fourier transform

$$\hat{\beta}_r(s) = (1 + \|s\|^2)^{-r/2}.$$

The $\mathcal{L}^2$-norm of β_r can be calculated by switching to $\hat{\beta}_r$ and using Plancherel's equality (5). For every $r > d/2$

$$\|\beta_r\|_{\mathcal{L}^2} = \|\hat{\beta}_r\|_{\mathcal{L}^2} = \lambda(r, d) := \pi^{d/4} \left(\frac{\Gamma(r - d/2)}{\Gamma(r)} \right)^{1/2}. \tag{11}$$

Indeed, using radial symmetry $\|\hat{\beta}_r\|^2_{\mathcal{L}^2} = \int_{\mathbb{R}^d} (1 + \|x\|^2)^{-r} \mathrm{d}x = \omega_d I$, where $\omega_d := 2\pi^{d/2}/\Gamma(d/2)$ is the area of the unit sphere in $\mathbb{R}^d$ [7, p. 303] and $I = \int_0^\infty (1 + \rho^2)^{-r} \rho^{d-1} \mathrm{d}\rho$. Substituting $\sigma = \rho^2$, one gets $\mathrm{d}\rho = (1/2)\sigma^{-1/2}\mathrm{d}\sigma$; hence,

$$I = (1/2) \int_0^\infty \frac{\sigma^{d/2-1}}{(1+\sigma)^r} \mathrm{d}\sigma = \frac{\Gamma(d/2)\Gamma(r - d/2)}{2\Gamma(r)}$$

(see [6, p. 60] for the last equality).

To estimate G_0^o-variations of β_r and $\hat{\beta}_r$, we use Theorem 2.4 with representations of these two functions as integrals of scaled Gaussians.

For $r > 0$, it is known [33, p. 132] that β_r is non-negative, radial, exponentially decreasing at infinity, analytic except at the origin, and belongs to $\mathcal{L}^1$. It can be expressed by the integral formula (see [22, p. 296] or [33])

$$\beta_r(x) = c_1(r, d) \int_0^\infty \mathrm{e}^{-t/(4\pi)} \, t^{-d/2+r/2-1} \mathrm{e}^{-(\pi/t)\|x\|^2} \mathrm{d}t, \tag{12}$$

where

$$c_1(r, d) = (2\pi)^{d/2} (4\pi)^{-r/2} / \Gamma(r/2)$$

and $\Gamma(z) = \int_0^\infty t^{z-1}\mathrm{e}^{-t} \, \mathrm{d}t$ is the Gamma function. The factor $(2\pi)^{d/2}$ occurs since our choice of Fourier transform (2) includes the factor $(2\pi)^{-d/2}$. Combining (12) with (1), we get a representation of the Bessel potential as an integral of normalized scaled Gaussians.

Proposition 3.1. *For every $r > 0$, d a positive integer, and $x \in \mathbb{R}^d$*

$$\beta_r(x) = \int_0^\infty v_r(t) \gamma^o_{\pi/t}(x) \, \mathrm{d}t,$$

where $v_r(t) = c_1(r, d) \, 2^{-d/4} \, \mathrm{e}^{-t/4\pi} \, t^{-d/4+r/2-1}$.

The next proposition estimates G_0^o-variation of β_r.

Proposition 3.2. *For d a positive integer and $r > d/2$,*

$$\|\beta_r\|_{G^o} \le \|\beta_r\|_{G_0^o} \le \int_0^\infty v_r(t)\,\mathrm{d}t = k(r,d),$$

where $k(r,d) = \frac{(\pi/2)^{d/4}\Gamma(r/2-d/4)}{\Gamma(r/2)}$.

Proof. As $G_0^o \subset G^o$, we get $\|\beta_r\|_{G^o} \le \|\beta_r\|_{G_0^o}$. To estimate $\|\beta_r\|_{G_0^o}$, we apply Theorem 2.4 with $w = v_r$, $\phi(x,y) = \phi(x,t) = \gamma^o_{\pi/t}(x)$, $Y = (0,\infty)$, and $\Omega = \mathbb{R}^d$ to the integral representation from Proposition 3.1, getting

$$\begin{aligned}\|\beta_r\|_{G_0^o} &\le \int_0^\infty v_r(t)\,\mathrm{d}t = c_1(r,d)\,2^{-d/4}\int_0^\infty \mathrm{e}^{-t/(4\pi)}\,t^{-d/4+r/2-1}\,\mathrm{d}t\\ &= (4\pi)^{-d/4+r/2}\,c_1(r,d)\,2^{-d/4}\int_0^\infty u^{-d/4+r/2-1}\,\mathrm{e}^{-u}\,\mathrm{d}u.\end{aligned}$$

Hence, by the definition of the Gamma function, one has

$$\begin{aligned}\|\beta_r\|_{G_0^o} &\le c_1(r,d)\,2^{-d/4}(4\pi)^{-d/4+r/2}\Gamma(r/2-d/4)\\ &= \frac{(\pi/2)^{d/4}\Gamma(r/2-d/4)}{\Gamma(r/2)} = k(r,d). \quad \square\end{aligned}$$

The Fourier transform of the Bessel potential can also be expressed as an integral of normalized scaled Gaussians.

Proposition 3.3. *For every $r > 0$, d a positive integer, and $s \in \mathbb{R}^d$*

$$\hat{\beta}_r(s) = \int_0^\infty w_r(t)\gamma_t^o(s)\,\mathrm{d}t,$$

where $w_r(t) = (\pi/2t)^{d/4}\,t^{r/2-1}\,\mathrm{e}^{-t}/\Gamma(r/2)$.

Proof. First we show that $\hat{\beta}_r(s) = I/\Gamma(r/2)$, where

$$I = \int_0^\infty t^{r/2-1}\,\mathrm{e}^{-t}\,\mathrm{e}^{-t\|s\|^2}\,\mathrm{d}t.$$

Indeed, putting $u = t(1+\|s\|^2)$, so $\mathrm{d}t = \mathrm{d}u(1+\|s\|^2)^{-1}$, we get

$$I = (1+\|s\|^2)^{-r/2}\int_0^\infty u^{r/2-1}\,\mathrm{e}^{-u}\,\mathrm{d}u = \hat{\beta}_r(s)\Gamma(r/2).$$

By (1), $\|\gamma_t\|_{\mathcal{L}^2} = (\pi/2t)^{d/4}$, so $\hat{\beta}_r(s) = \int_0^\infty (\pi/2t)^{d/4}\,t^{r/2-1}\,\mathrm{e}^{-t}/\Gamma(r/2)\,\gamma_t^o(s)\mathrm{d}t$. □

The next proposition gives an upper bound on G_0^o-variation of $\hat{\beta}_r$.

Proposition 3.4. *For d a positive integer and $r > d/2$,*

$$\|\hat{\beta}_r\|_{G^o} \le \|\hat{\beta}_r\|_{G_0^o} \le \int_0^\infty w_r(t)\mathrm{d}t = k(r,d),$$

where $k(r,d) = \frac{(\pi/2)^{d/4}\Gamma(r/2-d/4)}{\Gamma(r/2)}$.

Proof. A straightforward calculation shows that the $\mathcal{L}^1$-norm of the weighting function w_r is the same as the $\mathcal{L}^1$-norm of the weighting function v_r and the upper bound follows from Theorem 2.4 as in Proposition 3.2 but with $\phi(x,y) = \phi(x,t) = \gamma_t^o(x)$. □

Since the Fourier transform is an isometry on $\mathcal{L}^2$, by (10) the functions β_r and $\hat{\beta}_r$ have the same variation with respect to G_0^o. Propositions 3.2 and 3.4 give the same upper bound $k(r, d)$ on this number. If for some fixed $c > 0$, $r_d = d/2 + c$, then $k(r_d, d) \to 0$ exponentially fast as $d \to \infty$.

As all elements of G_{β_r} have the same $\mathcal{L}^2$-norm equal to $\lambda(r, d)$,

$$\|.\|_{G^o_{\beta_r}} = \lambda(r, d)\|.\|_{G_{\beta_r}}. \tag{13}$$

An application of (9) and (11) with Propositions 3.2 or 3.4 shows the following result:

Theorem 3.5. *For d, n positive integers and $r > d/2$*

$$\|\beta_r - \operatorname{span}_n G_0\|_{\mathcal{L}^2} = \|\hat{\beta}_r - \operatorname{span}_n G_0\|_{\mathcal{L}^2} \le n^{-1/2} \left(k(r, d)^2 - \lambda(r, d)^2\right)^{1/2}.$$

As above, for $c > 0$ and d large enough, the theorem shows that the Bessel potential of order $r_d = d/2 + c$ can be well-approximated by a network with just one Gaussian unit; hence, β_{r_d} is close in $\mathcal{L}^2$-norm to a multiple of some d-dimensional Gaussian centered at the origin.

4. Approximation of smooth functions by Gaussian RBFs

In this section we estimate rates of approximation by Gaussian RBF for functions in the Bessel-potential spaces. To obtain the estimates we first derive upper bounds on variation with respect to the set of translated Bessel potentials and then combine them with the estimates of G_0-variation of Bessel potentials from the previous section.

Let $h * g$ denote the convolution of two functions h and g,

$$(h * g)(x) = \int_{\mathbb{R}^d} h(y)g(x - y)\mathrm{d}y.$$

For d a positive integer, $r > d/2$, and $q \in [1, \infty]$, the *Bessel-potential space* (with respect to $\mathbb{R}^d$) [33, pp. 134–136] denoted by $(L^{q,r}, \|.\|_{L^{q,r}})$ is defined as

$$L^{q,r} := \{f \mid f = w * \beta_r,\ w \in \mathcal{L}^q\}$$

and

$$\|f\|_{L^{q,r}} := \|w\|_{\mathcal{L}^q} \quad \text{for } f = w * \beta_r.$$

Since the Fourier transform (2) of a convolution is $(2\pi)^{d/2}$ times the product of the transforms, we have $\hat{w} = (2\pi)^{-d/2}\hat{f}/\hat{\beta}_r$. Thus $w = (2\pi)^{-d/2}(\hat{f}/\hat{\beta}_r)\check{}$ is uniquely determined by f and so the Bessel-potential norm is well-defined.

For τ_y the translation operator given by $(\tau_y f)(x) = f(x - y)$ let

$$G_{\beta_r} = \{\tau_y \beta_r \mid y \in \mathbb{R}^d\}$$

denote the *set of translates of the Bessel potential of order* r. For $r > d/2$, β_r belongs to $\mathcal{L}^2$; since translation does not change the $\mathcal{L}^2$-norm, $G_{\beta_r} \subset \mathcal{L}^2$.

Functions in the Bessel-potential space are convolutions with β_r which are integral formulas. Thus we get the following upper bound:

Proposition 4.1. *Let d be a positive integer, $r > d/2$, $w : \mathbb{R}^d \to \mathbb{R}$ continuous except on a set Z_0 of measure zero, $w \in \mathcal{L}^1$, and $f = w * \beta_r$. Then $\|f\|_{G_{\beta_r}} \le \|w\|_{\mathcal{L}^1} = \|f\|_{L^{1,r}}$.*

Proof. The bounds follow from Theorem 2.4, applied to the integral formula $f(x) = \int w(y)\beta_r(x - y)\mathrm{d}y = \int w(y)\lambda(r, d)\beta_r^o(x - y)\mathrm{d}y$ combined with (13). Take $Y = \mathbb{R}^d$, $Y_0 = Z_0$, let $\phi(x, y) = \beta_r^o(x - y)$, and let $w(y)\lambda(r, d)$ be the weight function. The condition $r > d/2$ is needed to ensure that $G_{\beta_r} \subset \mathcal{L}^2$. □

For $h : U \to \mathbb{R}$, U a topological space, let

$$\operatorname{supp} h = \operatorname{cl}\{u \in U \mid h(u) \neq 0\}.$$

Proposition 4.2. *Let d be a positive integer, $r > d/2$, $w \in \mathcal{L}^2$ continuous except on a set of measure zero, $\lambda(\operatorname{supp} w) = \nu < \infty$, and $f = w * \beta_r$. Then*

$$\|f\|_{G_{\beta_r}} \leq \nu^{1/2}\|f\|_{L^{2,r}}.$$

Proof. By the Cauchy–Schwartz inequality, $\|w\|_{\mathcal{L}^1} \leq \nu^{1/2}\|w\|_{\mathcal{L}^2} = \nu^{1/2}\|f\|_{L^{2,r}}$. But by Proposition 4.1, $\|f\|_{G_{\beta_r}} \leq \|w\|_{\mathcal{L}^1}$. □

These estimates of variations give an upper bound on rates of approximation by linear combinations of n translates of the Bessel potential β_r.

Theorem 4.3. *Let d, n be positive integers, $r > d/2$, w continuous except on a set of measure zero, $f = w * \beta_r$, and $\lambda(r, d) = \pi^{d/4}\left(\frac{\Gamma(r-d/2)}{\Gamma(r)}\right)^{1/2}$.*

(i) *For $w \in \mathcal{L}^1$,*

$$\|f - \operatorname{span}_n G_{\beta_r}\|_{\mathcal{L}^2} \leq n^{-1/2}\left(\lambda(r, d)^2\|f\|^2_{L^{1,r}} - \|f\|^2_{\mathcal{L}^2}\right)^{1/2}.$$

(ii) *For $w \in \mathcal{L}^2$ with $\nu = \lambda(\operatorname{supp} w) < \infty$,*

$$\|f - \operatorname{span}_n G_{\beta_r}\|_{\mathcal{L}^2} \leq n^{-1/2}\left(\nu\, \lambda(r, d)^2\|f\|^2_{L^{2,r}} - \|f\|^2_{\mathcal{L}^2}\right)^{1/2}.$$

Proof. (i) By Proposition 4.1, (13) and (9).
(ii) As in Proposition 4.2, $w \in \mathcal{L}^2$ and $\operatorname{supp}(w) = \nu < \infty$ imply $w \in \mathcal{L}^1$; the rest follows from Proposition 4.2. □

Composing estimates of variations with respect to sets of translated Bessel potentials and Gaussians, we get an upper bound on rates of approximation by networks with n Gaussian RBF units for functions from Bessel spaces.

Theorem 4.4. *Let d, n be positive integers, $r > d/2$, w continuous except on a set of measure zero, $f = w * \beta_r$, and $k(r, d) = \frac{(\pi/2)^{d/4}\Gamma(r/2-d/4)}{\Gamma(r/2)}$.*

(i) *For $w \in \mathcal{L}^1$,*

$$\|f - \operatorname{span}_n G\|_{\mathcal{L}^2} \leq n^{-1/2}\left(k(r, d)^2\,\|f\|^2_{L^{1,r}} - \|f\|^2_{\mathcal{L}^2}\right)^{1/2}.$$

(ii) *For $w \in \mathcal{L}^2$ and $\lambda(\operatorname{supp} w) = \nu < \infty$,*

$$\|f - \operatorname{span}_n G\|_{\mathcal{L}^2} \leq n^{-1/2}\left(k(r, d)^2\,\nu\,\|f\|^2_{L^{2,r}} - \|f\|^2_{\mathcal{L}^2}\right)^{1/2}.$$

Proof. By (9), $\|f - \operatorname{span}_n G\|_{\mathcal{L}^2} \leq \left(\|f\|^2_{G^o} - \|f\|^2_{\mathcal{L}^2}\right)^{1/2} n^{-1/2}$. By Lemma 2.2 with $\mathcal{X} = \mathcal{L}^2, F = G^o, H = G_{\beta_r}$, using Proposition 3.2 and the fact that G^o is closed under translations, we have

$$\|f\|_{G^o} \leq \sup\{\|\tau_y(\beta_r)\|_{G^o} \mid y \in \mathbb{R}^d\}\|f\|_{G_{\beta_r}} \leq k(r, d)\|f\|_{G_{\beta_r}}.$$

Thus $\|f - \operatorname{span}_n G\|_{\mathcal{L}^2} \leq n^{-1/2}\left(k(r, d)^2\|f\|^2_{G_{\beta_r}} - \|f\|^2_{\mathcal{L}^2}\right)^{1/2}$.

Then the statements follow from the upper bounds on $\|f\|_{G_{\beta_r}}$ given in Propositions 4.1 and 4.2 respectively. □

5. Upper bounds in terms of Sobolev norms

In this section, bounds on approximation by Gaussian RBFs are given in terms of Sobolev norms.

Let r be a positive integer and let $W^{2,r}$ denote the *Sobolev space* of functions with tth-order partials in $\mathcal{L}^2$ for $t \in \{0, 1, \ldots, r\}$ and norm

$$\|f\|_{W^{2,r}} = \left(\sum_{|\alpha| \le r} \|D^\alpha f\|^2_{\mathcal{L}^2} \right)^{1/2},$$

where α denotes a multi-index (i.e., a vector of non-negative integers), D^α denotes the corresponding partial derivative operator, and $|\alpha| = \alpha_1 + \cdots + \alpha_d$.

Two norms are *equivalent* if each is bounded by a multiple of the other. For integer smoothness, equivalence of the Sobolev norm $\|\cdot\|_{W^{2,r}}$ and the Bessel-potential norm $\|\cdot\|_{L^{2,r}}$ is well-known (e.g., [34] or [1, p. 252]). As constants of equivalence were not readily available, we derive one of them here.

Let d and r be positive integers, $r > d/2$, $w \in \mathcal{L}^2$, and $f = w * \beta_r$. Then

$$\|f\|_{L^{2,r}} \le (2\pi)^{-d/2} (r!)^{1/2} \|f\|_{W^{2,r}}. \tag{14}$$

Indeed, since $f = w * \beta_r$, $\hat{f} = (2\pi)^{d/2} \hat{w} \hat{\beta}_r$ and so

$$\|f\|_{L^{2,r}} = (2\pi)^{-d/2} \|\hat{f}/\hat{\beta}_r\|_{\mathcal{L}^2} = (2\pi)^{-d/2} \left(\int_{\mathbb{R}^d} |\hat{f}(s)|^2 (1 + |s|^2)^r \mathrm{d}s \right)^{1/2}.$$

Let $\binom{r}{\sigma}$ denote the multinomial coefficient $r!/\sigma_1! \ldots \sigma_t!$. Note that $(1+|s|^2)^r = \sum_{|\sigma|=r} \binom{r}{\sigma} |u^{2\sigma}|$, for $u \in \mathbb{R}^{d+1}$ defined by $u_j = s_j, j = 1, \ldots, d$, $u_{d+1} = 1$, for $\sigma = (\sigma_1, \ldots \sigma_{d+1}) \in \mathbb{N}^{d+1}$ a multi-index of length $d+1$, and $|u^{2\sigma}| = |u_1^{2\sigma_1} \cdots u_{d+1}^{2\sigma_{d+1}}|$. Hence, we have

$$\int_{\mathbb{R}^d} |\hat{f}(s)|^2 (1+|s|^2)^r \mathrm{d}s \le C(r,d) \int_{\mathbb{R}^d} |\hat{f}(s)|^2 \sum_{|\alpha| \le r} |s^{2\alpha}| \mathrm{d}s,$$

where $C(r,d) = \max \left\{ \binom{r}{\sigma} \mid |\sigma| = r \right\}$. It follows from basic properties of the Fourier transform that the integral on the right-hand side is the square of the Sobolev norm of f; see, e.g., [34, p. 162]. Clearly, $C(r,d) \le r!$, and equality holds if and only if $r \le d$. This establishes (14).

Thus the larger the dimension, the more the magnitudes of these two equivalent norms differ. We can now estimate the rate of approximation by scaled and translated Gaussians in terms of the Sobolev norm of the function to be approximated.

Theorem 5.1. *Let d, n, r be positive integers, $r > d/2$, w continuous except on a set of measure zero, and $f = w * \beta_r$. For $w \in \mathcal{L}^2$ and $\lambda(\operatorname{supp} w) = \nu < \infty$,*

$$\|f - \operatorname{span}_n G\|_{\mathcal{L}^2} \le n^{-1/2} \left(\left(\frac{\pi}{8}\right)^{d/2} \left(\frac{\Gamma(r/2 - d/4)}{\Gamma(r/2)} \right)^2 \nu\, r! \|f\|^2_{W^{2,r}} - \|f\|^2_{\mathcal{L}^2} \right)^{1/2}.$$

Proof. Using Theorem 4.4(ii) and (14), the $\mathcal{L}^2$-distance from f to $\operatorname{span}_n G$ is at most $(k(r,d)^2\, \nu\, (2\pi)^{-d} (r!)^1 \|f\|^2_{W^{2,r}} - \|f\|^2_{\mathcal{L}^2})^{1/2} n^{-1/2}$, and the result follows. □

6. Tractability of approximation by RBFs

Our results can be interpreted in terms of tractability (see below for the definition) of multivariable approximation by Gaussian-radial-basis networks.

For $\mathcal{A}_d \subseteq \mathcal{L}^2(\mathbb{R}^d)$ and n a positive integer, let $e_n(\mathcal{A}_d)$ denote the worst-case $\mathcal{L}^2$-error in approximating the elements of $\mathcal{A}_d$ by elements from $\operatorname{span}_n G$; i.e.,

$$e_n(\mathcal{A}_d) = \sup_{f \in \mathcal{A}_d} \inf_{g \in \operatorname{span}_n G} \|g - f\|_{\mathcal{L}^2},$$

where G is the family of all scaled and shifted Gaussians. Let d be a positive integer, $r > d/2$, $\nu > 0$, and let w be continuous except on a set of measure zero. For $\mathcal{A}_d$ equal to any of two subsets defined below, we have shown that

$$e_n(\mathcal{A}_d) \le n^{-1/2} a(r,d)$$

and $a(r,d)$ tends to zero exponentially fast as $d \to \infty$, as explained below.

Define the following two subsets of $\mathcal{L}^2$:

$$\mathcal{A}_d^{(1)} = \{w * \beta_r \mid \|w\|_{L^1} \le 1\},$$

and

$$\mathcal{A}_d^{(2)} = \{w * \beta_r \mid \|w\|_{L^2} \le 1,\ \lambda(\operatorname{supp} w) \le \nu\}.$$

In Theorem 4.4(i) and (ii), we have shown that

$$e_n(\mathcal{A}_d^{(1)}) \le n^{-1/2}\,\frac{(\pi/2)^{d/4}\Gamma(r/2-d/4)}{\Gamma(r/2)} \tag{15}$$

and

$$e_n(\mathcal{A}_d^{(2)}) \le n^{-1/2}\,\frac{(\pi/2)^{d/4}\Gamma(r/2-d/4)}{\Gamma(r/2)}\,\nu^{1/2}, \tag{16}$$

respectively.

Considering only the dependence on n, better rates have been obtained under different hypotheses than ours on the functions to be approximated by Gaussians without scaling, providing constructive algorithms; see, e.g., [24–26]. But many estimates available in the literature for $e_n(\mathcal{A}_d)$, where $\mathcal{A}_d$ is a suitable set of functions of d variables (see, e.g., [3,5,9,28] for sigmoidal-neural networks and radial-basis-function networks), are of the form

$$e_n(\mathcal{A}_d) \le n^{-\delta}\,\kappa(d), \tag{17}$$

where $\delta > 0$ and $\kappa(d)$ is an increasing function of the number d of variables. In some literature, the term $\kappa(d)$ in (17) is referred to as "a constant"; however, this means constant only with respect to n but not with respect to the number d of variables.

For example, rates of approximation of functions on $\mathbb{R}^d$ with all l-order partial derivatives uniformly bounded for some positive integer l were investigated in [36], but it was not specified how the multiplicative factors in the estimates depend on d.

The dependence of $\kappa(d)$ on d may not be crucial for small values of d; however, for large values of d, approximation error $e_n(\mathcal{A}_d)$ can even grow exponentially with d as a consequence of exponential growth in $\kappa(d)$ (see, e.g., [3, item 9, p. 940]), when the "curse of dimensionality" [4] strikes. However, if $\kappa(d) \le C\,d^\alpha$ for some $C, \alpha > 0$ independent of n and d, then

$$e_n(\mathcal{A}_d) \le C\,n^{-\delta}\,d^\alpha \tag{18}$$

and the approximation problem is said to be *tractable* in the number d of variables [35–37,26].

As remarked in [36], in general estimating the dependence of $e_n(\mathcal{A}_d)$ on d is much harder than estimating its dependence on n and only a few results are available. For neural-network approximation, upper bounds that depend polynomially on d, hence ensuring tractability, were derived in [3,17,19,20]. In [26, Theorem 4.2], a tractability result for approximation by RBF networks was obtained in the supremum norm on $\mathbb{R}^d$ which also gives an upper bound of the form (18).

The bounds (15) and (16) are of the form $n^{-1/2}\,a_j(r,d)$, where $a_j(r,d)$ is given by

$$a_1(r,d) = \frac{(\pi/2)^{d/4}\Gamma(r/2-d/4)}{\Gamma(r/2)} \tag{19}$$

and

$$a_2(r,d) = \frac{(\pi/2)^{d/4}\Gamma(r/2-d/4)}{\Gamma(r/2)}\,\nu^{1/2}, \tag{20}$$

respectively. For $j = 1, 2$, $a_j(c + d/2, d) \to 0$ exponentially fast for $d \to \infty$. Also, $r' > r > d/2$ implies $a_j(r', d) < a_j(r, d)$.

This behavior implies the tractability of multivariable approximation for $\mathcal{A}_d^{(1)}$ and $\mathcal{A}_d^{(2)}$ by Gaussian-radial-basis networks. In fact, for these two classes, for d sufficiently large Eq. (18) holds for all negative $\alpha > -\infty$.

Acknowledgments

The authors are grateful to one reviewer, who suggested they include the discussion of Section 6, and pointed out various references. Collaboration of P.C.K. with V.K. and M.S. was partially supported by Georgetown University. V.K. was partially supported by GA ČR grant 201/08/1744 and by the Institutional Research Plan AV0Z10300504. Collaboration between V.K. and M.S. was partially supported by 2007–2009 Scientific Agreement among University of Genoa, National Research Council of Italy, and Academy of Sciences of the Czech Republic. M.S. was partially supported by a PRIN Grant from the Italian Ministry for University and Research, project "Models and Algorithms for Robust Network Optimization".

References

[1] R.A. Adams, J.J.F. Fournier, Sobolev Spaces, Academic Press, Amsterdam, 2003.
[2] A.R. Barron, Neural net approximation, in: K. Narendra (Ed.), Proc. 7th Yale Workshop on Adaptive and Learning Systems, Yale University Press, 1992, pp. 69–72.
[3] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Transactions on Information Theory 39 (1993) 930–945.
[4] R. Bellman, Dynamic Programming, Princeton University Press, Princeton, NJ, 1957.
[5] M. Burger, A. Neubauer, Error bounds for approximation with neural networks, Journal of Approximation Theory 112 (2001) 235–250.
[6] B.C. Carlson, Special Functions of Applied Mathematics, Academic Press, New York, 1977.
[7] R. Courant, Differential and Integral Calculus, 1964 ed., vol. 2, Wiley, New York, 1936 (transl. E. J. McShane).
[8] F. Girosi, Approximation error bounds that use VC-bounds, in: Proceedings of the International Conference on Neural Networks, Paris, 1995, pp. 295–302.
[9] F. Girosi, G. Anzellotti, Rates of convergence for radial basis functions and neural networks, in: R.J. Mammone (Ed.), Artificial Neural Networks for Speech and Vision, Chapman & Hall, 1993, pp. 97–113.
[10] E.J. Hartman, J.D. Keeler, J.M. Kowalski, Layered neural networks with Gaussian hidden units as universal approximations, Neural Computation 2 (1990) 210–215.
[11] 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 (1992) 608–613.
[12] P.C. Kainen, V. Kůrková, An integral upper bound for neural-network approximation, Neural Computation. Research Report V-1023, Institute of Computer Science, Prague, 2008. http://www.cs.cas.cz/research (submitted for publication).
[13] P.C. Kainen, V. Kůrková, M. Sanguineti, Estimates of approximation rates by Gaussian radial-basis functions, in: B. Beliczynski, A. Dzielinski, M. Iwanowski, B. Ribeiro (Eds.), Lecture Notes in Computer Science, vol. 4432, Springer, Berlin, Heidelberg, 2007, pp. 11–18.
[14] M.A. Kon, L.A. Raphael, D.A. Williams, Extending Girosi's approximation estimates for functions in Sobolev spaces via statistical learning theory, Journal of Analysis and Applications 3 (2005) 67–90.
[15] V. Kůrková, Dimension-independent rates of approximation by neural networks, in: K. Warwick, M. Kárný (Eds.), Computer-Intensive Methods in Control and Signal Processing: The Curse of Dimensionality, Birkhäuser, Boston, 1997, pp. 261–270.
[16] V. Kůrková, High-dimensional approximation and optimization by neural networks, in: J. Suykens, et al. (Eds.), Advances in Learning Theory: Methods, Models and Applications, IOS Press, Amsterdam, 2003, pp. 69–88 (Chapter 4).
[17] V. Kůrková, Minimization of error functionals over perceptron networks, Neural Computation 20 (2008) 252–270.
[18] V. Kůrková, P.C. Kainen, V. Kreinovich, Estimates of the number of hidden units and variation with respect to half-spaces, Neural Networks 10 (1997) 1061–1068.
[19] V. Kůrková, M. Sanguineti, Comparison of worst case errors in linear and neural network approximation, IEEE Transactions on Information Theory 48 (2002) 264–275.
[20] V. Kůrková, P. Savický, K. Hlaváčková, Representations and rates of approximation of real-valued Boolean functions by neural networks, Neural Networks 11 (1998) 651–659.
[21] V.E. Maiorov, Approximation by ridge functions, Journal of Approximation Theory 99 (1999) 68–94.
[22] C. Martínez, M. Sanz, The Theory of Fractional Powers of Operators, Elsevier, Amsterdam, 2001.
[23] H.N. Mhaskar, Versatile Gaussian networks, in: Proc. IEEE Workshop of Nonlinear Image Processing, 1995, pp. 70–73.
[24] H.N. Mhaskar, An Introduction to the Theory of Weighted Polynomial Approximation, World Scientific, Singapore, 1996.
[25] H.N. Mhaskar, When is approximation by Gaussian networks necessarily a linear process? Neural Networks 17 (2004) 989–1001.
[26] H.N. Mhaskar, On the tractability of multivariate integration and approximation by neural networks, Journal of Complexity 20 (2004) 561–590.
[27] H.N. Mhaskar, C.A. Micchelli, Approximation by superposition of a sigmoidal function and radial basis functions, Advances in Applied Mathematics 13 (1992) 350–373.
[28] H.N. Mhaskar, C.A. Micchelli, Dimension-independent bounds on approximation by neural networks, IBM Journal of Research and Development 38 (1994) 277–284.
[29] F.J. Narcowich, J.D. Ward, H. Wendland, Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions, Constructive Approximation 24 (2006) 175–186.
[30] J. Park, I.W. Sandberg, Universal approximation using radial-basis-function networks, Neural Computation 3 (1991) 246–257.
[31] J. Park, I. Sandberg, Approximation and radial basis function networks, Neural Computation 5 (1993) 305–316.

[32] G. Pisier, Remarques sur un resultat non publié de B. Maurey, in: Seminaire d'Analyse Fonctionelle, vol. I(12), École Polytechnique, Centre de Mathématiques, Palaiseau, 1980–1981.
[33] E.M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, NJ, 1970.
[34] R. Strichartz, A Guide to Distribution Theory and Fourier Transforms, World Scientific, NJ, 2003.
[35] J.F. Traub, A.G. Werschulz, Complexity and Information, Cambridge University Press, 1999.
[36] G.W. Wasilkowski, H. Woźniakowski, Complexity of weighted approximation over $\mathbb{R}^d$, Journal of Complexity 17 (2001) 722–740.
[37] H. Woźniakowski, Tractability and strong tractability of linear multivariate problems, Journal of Complexity 10 (1994) 96–128.