Proceedings of the 15th International Conference
on Computational and Mathematical Methods
in Science and Engineering, CMMSE 2015
6–10 July, 2015.

RBF approximation of large datasets by partition of unity and local stabilization

R. Cavoretto[1], S. De Marchi[2], A. De Rossi[1], E. Perracchione[1] and G. Santin[2]

[1] *Department of Mathematics "G. Peano", University of Torino*

[2] *Department of Mathematics, University of Padova*

emails: roberto.cavoretto@unito.it, demarchi@math.unipd.it, alessandra.derossi@unito.it, emma.perracchione@unito.it, gsantin@math.unipd.it

Abstract

We present an algorithm to approximate large dataset by Radial Basis Function (RBF) techniques. The method couples a fast domain decomposition procedure with a localized stabilization method. The resulting algorithm can efficiently deal with large problems and it is robust with respect to the typical instability of kernel methods.

Key words: Approximation, Radial Basis Functions, partition of unity, stabilization

1 Introduction

Radial Basis Function (RBF) methods provide a flexible way to recover multivariate functions from scattered data. In the recent years, such techniques have been successfully applied in various fields of applied mathematics, computer science and engineering to solve several problems, such as approximation of functions and numerical solutions of PDEs.

Although effective, these methods require special attention when applied to large problems. Namely, difficulties often arise on the side of the computational cost, due to the use of usually dense matrices, and on the side of stability of the algorithms, and these limitations are sometimes an obstacle to their application on concrete and real-world problems.

 ISBN: 978-84-617-2230-3

As regards stability, it is known that the cause of ill-conditioning is the particular algorithm used to construct the solution. In particular, it has been proven that the approximation operator itself is stable in the function space associated with the RBF kernel (see [6]), while the instability comes from the use of the unstable *standard basis* of translates. To this end, in recent years a lot of efforts have been played to study and construct better conditioned bases. A general theoretical framework for change of basis in kernel-based spaces is developed in [12]. For specific kernels, namely the Gaussian and inverse multiquadric kernels, almost perfectly conditioned bases have been constructed ([8, 9, 10, 7]), and some method to produce better bases for a generic kernel have been developed too (see [4, 5, 11]).

To overcome the aforementioned limitations, we propose a new method based on the coupling of a fast domain decomposition technique with a local stabilization method relying on the change of basis presented in [4]. This stabilization method have been chosen thanks to its applicability, without restriction, to any kernel. The algorithm presented here is intended as an exploratory attempt to couple such methods, and we plan to improve it by replacing the domain decomposition/points search technique used in this paper with the recently developed *block search* algorithms (see [2, 3]).

2 Kernel based approximation

Assuming to have a finite set of n distinct data points $X_n = \{x_i, 1 \leq i \leq n\}$ in a bounded domain $\Omega \subset \mathbb{R}^d$ and a set of data values $F_n = \{f_i, 1 \leq i \leq n\}$, our goal is to recover a continuous function $R_n : \Omega \to \mathbb{R}$ that interpolates the data. To construct R_n one considers a positive definite and symmetric kernel $K : \Omega \times \Omega \to \mathbb{R}$ and looks for a solution in the form

$$R_n = \sum_{i=1}^{n} c_i K(\cdot, x_i). \tag{1}$$

The coefficient vector $c = [c_1, \ldots, c_n]^T$ is determined by imposing the interpolation condition $R_n(x_i) = f_i$, $1 \leq i \leq n$, i.e., by solving the linear system $Ac = f$, where $A = [K(x_i, x_j)]_{i,j=1}^n$ is the *kernel matrix* and $f = [f_1, \ldots, f_n]^T$. Since the kernel is positive definite, so it is the matrix A, hence the solution c exists and it is unique for any given data. The kernels we will deal with are always radial, meaning that there exist a univariate function $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}$ and a *shape parameter* $\varepsilon > 0$ such that $K(x, y) = \phi(\varepsilon \|x - y\|_2)$ for all $x, y \in \Omega$. The shape parameter has the effect of localize or flatten the kernel, and it has a crucial role in the behavior of the approximant.

We recall that associated to a positive definite kernel K on Ω there is a uniquely defined reproducing kernel Hilbert space $\mathcal{N}_K(\Omega)$, called the *native space* of K on Ω. When the data comes from the sampling of a function $f \in \mathcal{N}_K(\Omega)$, i.e., $f_i = f(x_i)$, the operator $f \mapsto R_n$ is the projector from $\mathcal{N}_K(\Omega)$ into the subspace $V_n = \text{span}\{K(\cdot, x_i) : 1 \leq i \leq n\}$. For $f \in \mathcal{N}_K(\Omega)$ we have $R_n \to f$ as X_n gets dense into Ω, and explicit convergence estimates

 ISBN: 978-84-617-2230-3

are known (see e.g. [16]), depending on the smoothness of K and Ω and on the distribution of X_n in Ω.

Nevertheless, the actual computation of R_n is often numerically intractable using the standard method described above. In particular, the kernel matrix is generally dense and of the same size n of the data, hence the construction and the solution of the linear system is computationally expensive. Moreover, it is in general required to tune the parameter ε to fit the data, but reducing it is a source of instability, especially when K is smooth.

We will see in Section 3 how to deal with large data sets by splitting the problem in smaller subproblems, and in Section 4 how to locally employ the change of basis to control the instability.

3 Domain decomposition

We want to avoid to solve a global problem involving a large and possibly dense matrix. To do so, we focus on the partition of unity (PU) method (see [15]).

The idea is to decompose the problem, solve (many) small local approximation subproblems, and then blend together the results in a global approximant. We give here a general review of this method, while further implementation details are provided in Section 5.

We consider an open and finite covering $\{\Omega^{(k)}\}_{k=1}^N$ of the bounded set Ω, that is

$$\Omega \subset \bigcup_{k=1}^{N} \Omega^{(k)},$$

where we require mild overlapping of the patches. Associated to the covering one considers a partition of unity, i.e., a set of nonnegative and continuous functions $\{W^{(k)}\}_{k=1}^N$ with

$$\text{supp}(W^{(k)}) \subset \Omega^{(k)}, \ \sum_{k=1}^{N} W^{(k)}(x) = 1 \text{ for all } x \in \Omega.$$

With these tools in hand, if we are able to produce a local approximant $R^{(k)}$ for any subset $\Omega^{(k)}$, it is immediate to recover a global approximant R_n on Ω as

$$R_n(x) = \sum_{k=1}^{N} W^{(k)}(x) R^{(k)}(x), \ x \in \Omega.$$

But the kernel K is still positive definite when restricted to each $\Omega^{(k)}$, so it is possible to uniquely solve the approximation problem restricted to the local domains. Namely, for each $\Omega^{(k)}$, $1 \leq k \leq N$, we consider the restricted set of $n^{(k)}$ data locations defined as

 ISBN: 978-84-617-2230-3

$X_n^{(k)} = X_n \cap \Omega^{(k)}$, $n^{(k)} = \text{card}(X_n^{(k)})$, and the corresponding values $F_n^{(k)} = \{f(x) : x \in X_n^{(k)}\}$. The local approximant $R^{(k)}$ is then computed as in (1), i.e,

$$R^{(k)} = \sum_{i=1}^{n^{(k)}} c_i^{(k)} K(\cdot, x_i^{(k)}),$$

where now it suffices to solve a much smaller linear system to determine the coefficients in the above expansion.

It is important to recall that, under further assumptions on the covering and on the partition of unity, it is possible to prove that the global approximant keeps the same approximation order of the local ones (see [15]), but we will not go into the details here.

Instead we emphasize that, besides the theoretical framework, the success of the PU method relies on a good point searching technique. Although we will employ in Section 5 a standard algorithm, we plan to improve the method by updating the current domain decomposition method with the recently developed *block-based algorithms* (see [2, 3]). We stress here that this update will presumably lead to a reduction of the computational cost, while nothing will change in terms of accuracy. In this view, the numerical experiments presented in Section 5 should be indicative, as regards the error, of the behavior of the enhanced method.

4 Local stabilization

The partition of unity method allows to deal with large datasets in an efficient way, and partially reduces the ill-conditioning found in many cases when dealing with RBF problems, since it reduces the size of the problem. Nevertheless, especially in the *flat limit* $\varepsilon \to 0$, using a domain decomposition technique can be not enough. To this end, we want to employ locally a stabilization method that permits to face in an even more stable way large approximation problems. The coupling of the PU algorithm with the local stabilization will lead to the full algorithm of the next Section.

For any subdomain $\Omega^{(k)}$ there is an associated space of functions $\mathcal{N}_K(\Omega^{(k)})$ given by the restriction of $\mathcal{N}_K(\Omega)$ to $\Omega^{(k)}$. As in the global setting, the interpolation operator is the projection into the finite subspace $V_n^{(k)} = \text{span}\{K(\cdot, x_j^{(k)}) : x_j^{(k)} \in X_n^{(k)}\}$ generated by the restricted set of data $X_n^{(k)}$. To this subspace we apply the change of basis described in [4, 5]. Namely, we substitute the standard basis $\{K(\cdot, x_i^{(k)}) : x_i^{(k)} \in X_n^{(k)}\}$ with the Weighted Singular Value Decomposition (WSVD) basis $\{u_j^{(k)}\}_{j=1}^{n^{(k)}}$,

$$u_j^{(k)} = \sum_{i=1}^{n^{(k)}} d_{ij}^{(k)} K(\cdot, x_i^{(k)}).$$

 ISBN: 978-84-617-2230-3

The invertible coefficient matrix $D^{(k)} = [d_{ij}^{(k)}]_{i,j=1}^{n^{(k)}}$ is computed through a singular value decomposition of the local kernel matrix $A^{(k)} = [K(x_i^{(k)}, x_j^{(k)})]_{i,j=1}^{n^{(k)}}$, i.e.,

$$A^{(k)} = U^{(k)}\Sigma^{(k)}(U^{(k)})^T, \;\; D^{(k)} = U^{(k)}(\Sigma^{(k)})^{-1/2}, \;\; \Sigma^{(k)} = \mathrm{diag}(\sigma_1^{(k)}, \ldots, \sigma_{n^{(k)}}^{(k)}).$$

This basis has been constructed to mimic in a discrete way the *eigenbasis* defined through the Mercer's Theorem (see e.g. [13]), where the $L_2(\Omega)$ inner product is replaced by its discrete pointwise version $\ell_2(X_n)$. In particular, as proven in [4], the basis enjoys the following properties.

Proposition 4.1 *The WSVD basis* $\{u_j^{(k)}\}_{j=1}^{n^{(k)}}$ *has the following properties:*

(i) it is $\mathcal{N}_K(\Omega^{(k)})$*-orthonormal,*

(ii) it is $\ell_2(X_n^{(k)})$*-orthogonal with norms* $\{\sigma_j^{(k)}\}_j$*,*

(iii) $(u_j^{(k)}, f)_{\ell_2(X_n^{(k)})} = \sigma_j^{(k)}(u_j^{(k)}, f)_{\mathcal{N}_K(\Omega^{(k)})}$, $\forall f \in \mathcal{N}_K(\Omega^{(k)})$, $1 \le j \le n^{(k)}$*,*

(iv) $\sigma_{n^{(k)}}^{(k)} \ge \ldots \ge \sigma_1^{(k)} > 0$*.*

(v) $\sum_{j=1}^{n^{(k)}} \sigma_j^{(k)} = n^{(k)} \; \phi(0)$*.*

We want now to approximate the restriction of $f \in \mathcal{N}_K(\Omega)$ to $\Omega^{(k)}$, say $f^{(k)}$, using this basis. Thanks to the orthogonality property (i), the local interpolant of $f^{(k)}$ can be represented as

$$R^{(k)} = \sum_{j=1}^{n^{(k)}} (f^{(k)}, u_j^{(k)})_{\mathcal{N}_K(\Omega^{(k)})} u_j^{(k)}.$$

Moreover, each element of this new basis is associated to a singular value of the matrix A. Since we expect the singular values to decay very fast to zero, we can drop the last elements of the basis to remove from the interpolant the terms responsible for the ill-conditioning. To be more precise, for any $m^{(k)} \le n^{(k)}$, the truncated interpolant

$$R_{m^{(k)}}^{(k)} = \sum_{j=1}^{m^{(k)}} (f^{(k)}, u_j^{(k)})_{\mathcal{N}_K(\Omega^{(k)})} u_k^{(k)}$$

is the solution of the least-squares problem

$$\min_{g \in \mathrm{span}\{u_1^{(k)}, \ldots, u_{m^{(k)}}^{(k)}\}} \|f^{(k)} - g\|_{\ell_2(X_n^{(k)})}.$$

 ISBN: 978-84-617-2230-3

Indeed, thanks to property (iii), we have for any $m^{(k)} \leq n^{(k)}$

$$R^{(k)}_{m^{(k)}} = \sum_{j=1}^{m^{(k)}} (f^{(k)}, u^{(k)}_j)_{\mathcal{N}_K(\Omega^{(k)})} u^{(k)}_j = \sum_{j=1}^{m^{(k)}} (\sigma^{(k)}_j)^{-1} (f^{(k)}, u^{(k)}_j)_{\ell_2(X^{(k)}_n)} u^{(k)}_j, \tag{2}$$

and the rightmost term in the above expression is the $\ell_2(X^{(k)}_n)$ projection of $f^{(k)}$ into the subspace $\text{span}\{u^{(k)}_1, \ldots, u^{(k)}_{m^{(k)}}\}$. Obviously, for $m^{(k)} = n^{(k)}$ the latter approximant coincides with the interpolant, while we can look for a suitable truncation index $m^{(k)} < n^{(k)}$ to remove the elements of the basis corresponding to the singular values below some prescribed tolerance.

The use of this basis allows to reduce the instability, but it has some drawbacks. Namely, in case of severe ill-conditioning the truncation approach may be over regularizing. Moreover, the computation of the SVD is expensive, especially since we want to drop part of its factors. To overcome the second problem we will approximate the singular value decomposition using the Lanczos method (see e.g [14]), as done in [5]. This approximation introduces a perturbation in the basis that is thoroughly analyzed in the cited paper, nevertheless, the form (2) of the approximant is still valid for the approximated basis. As regards the regularizing effect of the truncation, it has been observed that the method, when applied globally, is capable to deal with only a limited instability. In fact, when $\varepsilon \to 0$ or when the kernel is too smooth, the original algorithm needs to leave out too much elements of the basis, and the resulting approximant may be meaningless. Instead, when repeatedly applied to the small subdomains $\Omega^{(k)}$, the regularizing effect simply improves the approximation, as shown in the next Section.

5 Numerical experiments

We present here two experiments to test our algorithm on two small problems in dimension $d = 2$. For both examples we compare the behavior of two different RBF kernels $K(x, y) = \phi(\varepsilon\|x - y\|_2)$ with different smoothness, namely the Gaussian $\phi(r) = \exp(-\varepsilon^2 r^2)$ (which is a $\mathcal{C}^\infty$ kernel) and the $\mathcal{C}^2$ Matérn kernel $\phi(r) = \exp(-\varepsilon r)(1 + \varepsilon r)$. In both examples the test function is the bivariate Franke's test function

$$\begin{aligned} f(x,y) = \quad & \tfrac{3}{4} e^{-\frac{1}{4}((9x-2)^2+(9y-2)^2)} \quad && + \tfrac{3}{4} e^{-(\frac{1}{49}(9x+1)^2+\frac{1}{10}(9y+1))} \\ & + \tfrac{1}{2} e^{-\frac{1}{4}((9x-7)^2+(9y-3)^2)} \quad && - \tfrac{1}{5} e^{-((9x-4)^2+(9y-7)^2)}, \end{aligned}$$

approximated on the unit square $\Omega = [0, 1]^2$. The shape parameter ε varies on a log-spaced grid in $[10^{-3}, 10^2]$. The error is measured by means of the root mean squared error (RMSE) on a equally spaced grid of 40^d points in Ω.

 ISBN: 978-84-617-2230-3

In both examples the PU method uses a covering of Ω based on balls $\{\Omega^{(k)}\}_{k=1}^{N}$, where the centers $\{y^{(k)}\}_{k=1}^{N}$ of the balls lie on an equally spaced 2D grid in Ω and the radii are the same for any $k = 1, \dots, N$. The points are organized through a *kd-tree*, as in [1]. We use n Halton points, where $n = 65^d, 129^d, 257^d$, which correspond respectively to $N = 16^d, 32^d, 64^d$. The common radius of the balls is chosen to be $1/\sqrt{N}$, so that the boundary of every patch intersects the centers of the nearby patches. As a partition of unity we use a superposition of $\mathcal{C}^2$ Wendland's functions $W^{(k)}(x) = \psi(\nu\|x - y^{(k)}\|)$, $\psi(r) = (1-\nu r)_+^4(4\nu r+1)$, where the shape parameter $\nu > 0$ is scaled to control the support of the partition. We remark that these choices of $\Omega^{(k)}$ and $W^{(k)}$ fulfills the requirement on the covering and the partition in order to preserve globally the order of the local approximation.

As a stopping rule for the locally employed Lanczos algorithm we use an extension of the rule proposed in [5]. Namely, since we know that the trace of the local kernel matrix $A^{(k)}$ is $n^{(k)}\ \phi(0)$ (see property (v) in Proposition 4.1), we stop the iteration when the distance between the trace of the Lanczos matrix and $n^{(k)}\ \phi(0)$ is below a fixed tolerance τ, here $\tau = 10^{-14}$.

The first example compares the approximation obtained with the new method with the one obtained with a global application of the change of basis described in Section 4. Since we use here a global approach, the experiment is limited to $n = 65^d$. The results are shown in Figure 1. For a not too small shape parameter, for both kernels, the global method and the new algorithm behaves in the same way. In the case of the Matérn kernel there is an intermediate range of the shape parameter where the global method is slightly superior, since it uses all the data simultaneously while no instability is yet present. As $\varepsilon \to 0$ the global method loses accuracy, while the local one is still able to compute an accurate and stable approximant.

In the second example we compare the behavior of the PU method without local regularization with the new algorithm. Figure 2 presents the results for this test. As expected, there is a first phase where no instability is present and the two methods behaves exactly in the same way. Instead, in the flat limit case the standard PU method starts to become unstable, while the stabilized method retain its accuracy and it is able to effectively compute the approximants. The effect is much more evident for the smoother Gaussian (which is smoother) than for the Matérn kernel.

Other tests have been performed for $d > 2$, with similar results.

Acknowledgements

The first, third and fourth authors have been partially supported by the funds of the University of Torino, project *ex 60%* "Metodi numerici nelle scienze applicate". The second and fifth authors have been partially supported by the funds of the University of Padova, project *CPDA124755* "Multivariate approximation with application to image reconstruction".

 ISBN: 978-84-617-2230-3

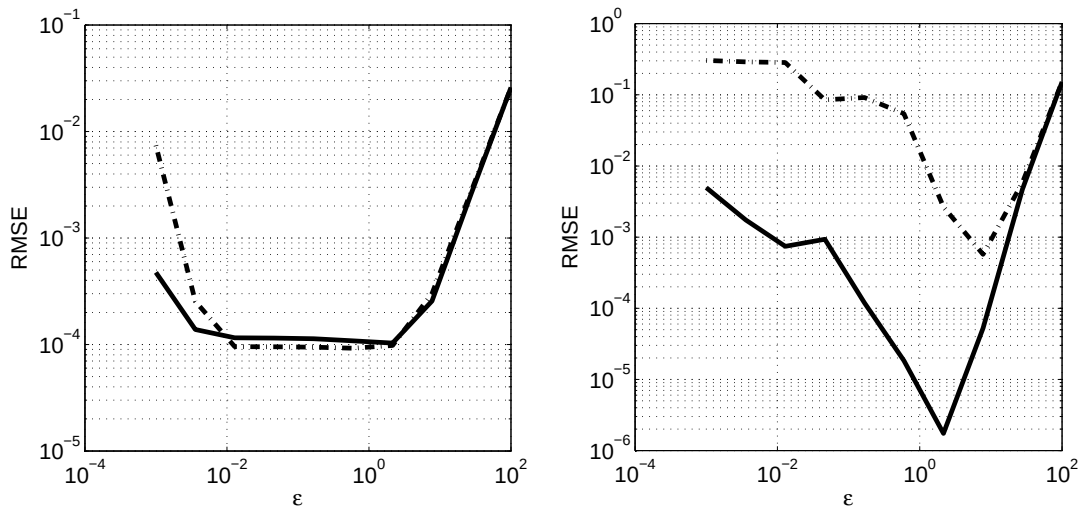

Figure 1: Comparison between the global approximant computed with the WSVD basis (dotted lines) and the new method (solid lines), for the $\mathcal{C}^2$ Matérn kernel (left) and the Gaussian kernel (right), for $10^{-3} \leq \varepsilon \leq 10^2$ and $n = 65^2$ Halton points in the unit square.

References

[1] R. Cavoretto, *A numerical algorithm for multidimensional modeling of scattered data points*, Comput. Appl. Math. **34** (2015) 65–80.

[2] R. Cavoretto, A. De Rossi, *A meshless interpolation algorithm using a cell-based searching procedure*, Comput. Math. Appl. **67** (2014) 1024–1038.

[3] R. Cavoretto and A. De Rossi, *A trivariate interpolation algorithm using a cube-partition searching procedure*, submitted (2015)

[4] S. De Marchi, G. Santin, *A new stable basis for radial basis function interpolation*, J. Comput. Appl. Math. **253** (2014) 1–13.

[5] S. De Marchi, G. Santin, *Fast computation of orthonormal basis for RBF spaces through Krylov space methods*, to appear in BIT DOI: 10.1007/s10543-014-0537-6 (2015).

[6] S. De Marchi, R. Schaback, *Stability of kernel-based interpolation*, Adv. Comput. Math. **32** (2010) 155–161.

[7] G. E. Fasshauer, M. J. McCourt, *Stable evaluation of Gaussian radial basis function interpolants*, SIAM J. Sci. Comput. **34** (2012) A737–A762.

 ISBN: 978-84-617-2230-3

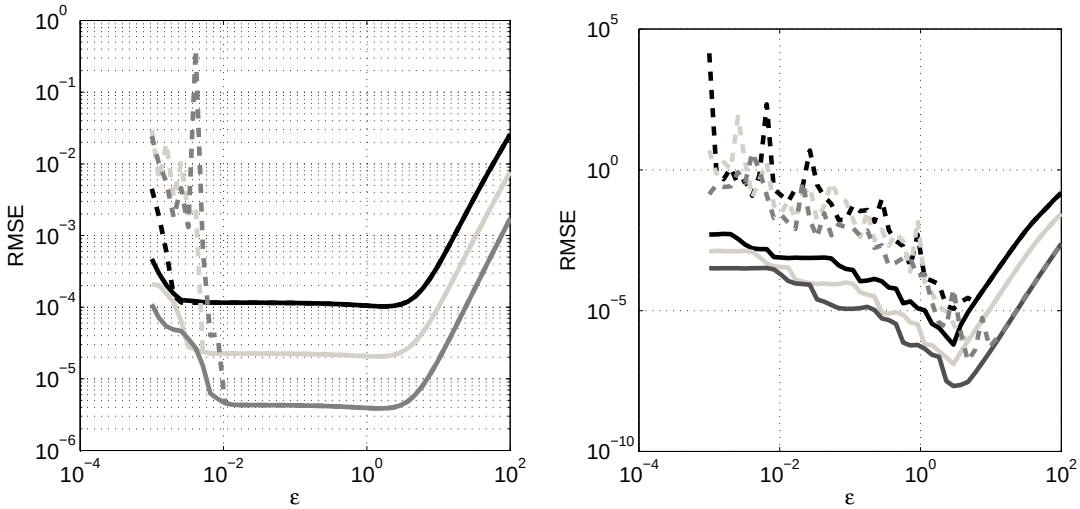

Figure 2: Comparison between the standard PU approximant (dotted lines) and the new method (solid lines), for the $\mathcal{C}^2$ Matérn kernel (left) and the Gaussian kernel (right), for $10^{-3} \leq \varepsilon \leq 10^2$ and $n = 65^2$ (black line), $n = 129^2$ (light gray), $n = 257^2$ (dark gray) Halton points in the unit square.

[8] B. Fornberg, E. Larsson, N. Flyer, *Stable computations with Gaussian radial basis functions*, SIAM J. Sci. Comput. **33** (2011) 869–892.

[9] B. Fornberg, C. Piret, *A stable algorithm for at radial basis functions on a sphere*, SIAM J. Sci. Comput. **30** (2007/08) 60–80.

[10] B. Fornberg, G. Wright, *Stable computation of multiquadrics interpolants for all values of the shape parameter*, Comput. Math. Appl. **47** (2004) 497–523.

[11] S. Müller and R. Schaback, *A Newton basis for kernel spaces*, J. Approx. Theory, **161** (2009) 645–655.

[12] M. Pazouki, R. Schaback, *Bases for kernel-based spaces*, J. Comput. Appl. Math. **236** (2011) 575–588.

[13] W. Pogorzelski, *Integral Equations and Their Applications. Vol. I*, International Series of Monographs in Pure and Applied Mathematics, vol. 88, Pergamon Press, Oxford, 1966.

[14] H. D. Simon and H. Zha, *Low-rank matrix approximation using the Lanczos bidiagonalization process with applications*, SIAM J. Sci. Comput. **21** (2000) 2257–2274.

 ISBN: 978-84-617-2230-3

[15] H. Wendland, *Fast evaluation of radial basis functions: methods based on partition of unity*, Approximation theory, X (St. Louis, MO, 2001), Vanderbilt Univ. Press, Nashville, TN, (2002) 473–483.

[16] H. Wendland, *Scattered Data Approximation*, Cambridge Monogr. Appl. Comput. Math., vol. 17, Cambridge Univ. Press, Cambridge, 2005.

 ISBN: 978-84-617-2230-3