Radial basis functions topics in 40 +1 slides

Stefano De Marchi
Department of Mathematics “Tullio Levi-Civita”
University of Padova
Napoli, 22nd March 2018

DIPARTIMENTO MATEMATICA

Outline

From cubic splines to RBF

Cubic splines

- Let $\mathcal{S}_3(X) = \{s \in C^2[a,b] : s_{|[x_i,x_{i+1}]} \in \mathbb{P}_3,\ 1 \le i \le N-1,\ s_{[a,x_1]}, s_{[x_N,b]} \in \mathbb{P}_1\}$ be the space of natural cubic splines.
- $\mathcal{S}_3(X)$ has the basis of truncated powers $(\cdot - x_j)^3_+,\ 1 \le j \le N$ plus an arbitrary basis for $\mathbb{P}_3(\mathbb{R})$

$$s(x) = \sum_{j=1}^{N} a_j (x - x_j)^3_+ + \sum_{j=0}^{3} b_j x^j, \quad x \in [a,b]\,. \tag{1}$$

- Using the identity $x^3_+ = \frac{(|x|^3 + x^3)}{2}$ and the fact that $s_{[a,x_1]}, s_{[x_N,b]} \in \mathbb{P}_1$

Cubic splines

- Let $\mathcal{S}_3(X) = \{s \in C^2[a,b] : s_{|[x_i,x_{i+1}]} \in \mathbb{P}_3,\ 1 \le i \le N-1,\ s_{[a,x_1]}, s_{[x_N,b]} \in \mathbb{P}_1\}$ be the space of natural cubic splines.
- $\mathcal{S}_3(X)$ has the basis of truncated powers $(\cdot - x_j)_+^3,\ 1 \le j \le N$ plus an arbitrary basis for $\mathbb{P}_3(\mathbb{R})$

$$s(x) = \sum_{j=1}^{N} a_j (x - x_j)_+^3 + \sum_{j=0}^{3} b_j x^j, \quad x \in [a,b]\,. \tag{1}$$

- Using the identity $x_+^3 = \frac{(|x|^3 + x^3)}{2}$ and the fact that $s_{[a,x_1]}, s_{[x_N,b]} \in \mathbb{P}_1$
- Every natural cubic spline s has the representation

$$s(x) = \sum_{j=1}^{N} a_j \phi(|x - x_j|) + p(x), \quad x \in \mathbb{R}$$

where $\phi(r) = r^3,\ r \ge 0$ and $p \in \mathbb{P}_1(\mathbb{R})$ s.t.

$$\sum_{j=1}^{N} a_j = \sum_{j=1}^{N} a_j x_j = 0\,. \tag{2}$$

Extension to $\mathbb{R}^d$

Going to $\mathbb{R}^d$ is straightforward: "radiality" becomes more evident

$$s(x) = \sum_{i=1}^{N} a_j \phi(\|x - x_j\|_2) + p(x), \quad x \in \mathbb{R}^d\,, \tag{3}$$

where $\phi : [0, \infty) \to \mathbb{R}$ is a univariate fixed function and $p \in \mathbb{P}_{l-1}(\mathbb{R}^d)$ is a low degree d-variate polynomial. The additional conditions in (**??**) become

$$\sum_{i=1}^{N} a_j q(x_j) = 0, \quad \forall\, q \in \mathbb{P}_{l-1}(\mathbb{R}^d)\,. \tag{4}$$

Extension to $\mathbb{R}^d$

Going to $\mathbb{R}^d$ is straightforward: "radiality" becomes more evident

$$s(x) = \sum_{i=1}^{N} a_j \phi(\|x - x_j\|_2) + p(x), \quad x \in \mathbb{R}^d \,, \tag{3}$$

where $\phi : [0, \infty) \to \mathbb{R}$ is a univariate fixed function and $p \in \mathbb{P}_{l-1}(\mathbb{R}^d)$ is a low degree d-variate polynomial. The additional conditions in (**??**) become

$$\sum_{i=1}^{N} a_j q(x_j) = 0, \quad \forall \; q \in \mathbb{P}_{l-1}(\mathbb{R}^d) \,. \tag{4}$$

Obs:

*We can omit the side conditions on the coefficients (**??**) (like using "B-splines") and consider approximants of the form*

$$s(x) = \sum_{i=1}^{N} a_j \phi(\|x - x_j\|_2)$$

and also omit the 2-norm symbol

Scattered data fitting:I

Setting

Given data $X = \{\mathbf{x}_j,\ j = 1, \ldots, N\}$, with $\mathbf{x}_j \in \Omega \subset \mathbb{R}^d$ and values $Y = \{\mathbf{y}_j \in \mathbb{R},\ j = 1, ..., N\}$ ($\mathbf{y}_j = f(\mathbf{x}_j)$), find a (continuous) function $P_f \in \mathrm{span}\{\phi(\cdot - \mathbf{x}_i),\quad \mathbf{x}_i \in X\}$ s.t.

$$P_f = \sum_{j=1}^{N} \alpha_j \phi(\|\cdot - \mathbf{x}_i\|), \qquad s.t.. \quad (P_f)_{|X} = Y \quad (interpolation)$$

Scattered data fitting:I

Setting

Given data $X = \{\mathbf{x}_j,\ j = 1, \ldots, N\}$, with $\mathbf{x}_j \in \Omega \subset \mathbb{R}^d$ and values $Y = \{\mathbf{y}_j \in \mathbb{R},\ j = 1, ..., N\}$ ($\mathbf{y}_j = f(\mathbf{x}_j)$), find a (continuous) function $P_f \in \mathrm{span}\{\phi(\cdot - \mathbf{x}_i),\quad \mathbf{x}_i \in X\}$ s.t.

$$P_f = \sum_{j=1}^{N} \alpha_j \phi(\| \cdot - \mathbf{x}_i \|), \qquad s.t.. \quad (P_f)_{|X} = Y \quad (interpolation)$$

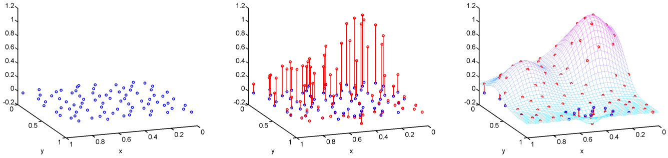


Figure: Data points, data values and data function

Obs: $\mathrm{P}_f = \sum_{i=j}^{n} u_j\ f_j$, with $u_j(\mathbf{x}_i) = \delta_{ji}$.

Scattered data fitting:II

Problem

For which functions $\phi : [0,\infty) \to \mathbb{R}$ *such that for all* $d, N \in \mathbb{N}$ *and all pairwise distrinct* $x_1, \ldots, x_n \in \mathbb{R}^d$ *the matrix* $\mathtt{det}(A_{\phi,X}) \neq 0$*? When the matrix* $A_{\phi,X} := (\phi(\|x_i - x_j\|_2)_{1\le i,j\le N}$ *, is invertible?*

Scattered data fitting:II

Problem

For which functions $\phi : [0, \infty) \to \mathbb{R}$ *such that for all* $d, N \in \mathbb{N}$ *and all pairwise distrinct* $x_1, \ldots, x_n \in \mathbb{R}^d$ *the matrix* $\mathtt{det}(A_{\phi,X}) \neq 0$*? When the matrix* $A_{\phi,X} := (\phi(\|x_i - x_j\|_2)_{1 \le i,j \le N}$ *, is invertible?*

Answer

The functions ϕ should be positive definite or conditionally positive definite of some order.

General kernel-based approximation

- ϕ, Conditionally Positive Definite (CPD) of order ℓ or Strictly Positive Definite (SPD) and radial
  - globally supported:

name	ϕ	ℓ
Gaussian C^∞ (GA)	$e^{-\varepsilon^2 r^2}$	0
Generalized Multiquadrics C^∞ (GM)	$(1 + r^2/\varepsilon^2)^{3/2}$	2

  - locally supported:

name	ϕ	ℓ
Wendland C^2 (W2)	$(1 - \varepsilon r)_+^4\ (4\varepsilon r + 1)$	0
Buhmann C^2 (B2)	$2r^4 \log r - 7/2r^4 + 16/3r^3 - 2r^2 + 1/6$	0

- we often consider $\phi(\varepsilon\cdot)$, with ε called shape parameter

General kernel-based approximation

- ϕ, Conditionally Positive Definite (CPD) of order ℓ or Strictly Positive Definite (SPD) and radial
  - globally supported:

name	ϕ	ℓ
Gaussian C^∞ (GA)	$e^{-\varepsilon^2 r^2}$	0
Generalized Multiquadrics C^∞ (GM)	$(1+r^2/\varepsilon^2)^{3/2}$	2

  - locally supported:

name	ϕ	ℓ
Wendland C^2 (W2)	$(1-\varepsilon r)_+^4\,(4\varepsilon r+1)$	0
Buhmann C^2 (B2)	$2r^4\log r - 7/2r^4 + 16/3r^3 - 2r^2 + 1/6$	0

- we often consider $\phi(\varepsilon\cdot)$, with ε called shape parameter
- kernel notation $K(\mathbf{x},\mathbf{y})(=K_\varepsilon(\mathbf{x},\mathbf{y})) = \Phi_\varepsilon(\mathbf{x}-\mathbf{y}) = \phi(\varepsilon\|\mathbf{x}-\mathbf{y}\|_2)$

General kernel-based approximation

DIPARTIMENTO MATEMATICA

- ϕ, Conditionally Positive Definite (CPD) of order ℓ or Strictly Positive Definite (SPD) and radial
  - globally supported:

name	ϕ	ℓ
Gaussian C^∞ (GA)	$e^{-\varepsilon^2 r^2}$	0
Generalized Multiquadrics C^∞ (GM)	$(1 + r^2/\varepsilon^2)^{3/2}$	2

  - locally supported:

name	ϕ	ℓ
Wendland C^2 (W2)	$(1 - \varepsilon r)_+^4\,(4\varepsilon r + 1)$	0
Buhmann C^2 (B2)	$2r^4 \log r - 7/2r^4 + 16/3r^3 - 2r^2 + 1/6$	0

- we often consider $\phi(\varepsilon\cdot)$, with ε called shape parameter
- kernel notation $K(\mathbf{x},\mathbf{y})(= K_\varepsilon(\mathbf{x},\mathbf{y})) = \Phi_\varepsilon(\mathbf{x}-\mathbf{y}) = \phi(\varepsilon\|\mathbf{x}-\mathbf{y}\|_2)$
- native space $\mathcal{N}_K(\Omega)$ (where K is the reproducing kernel)
- finite subspace $\mathcal{N}_K(X) = \mathrm{span}\{K(\cdot,x) : x \in X\} \subset \mathcal{N}_K(\Omega)$.

Error estimates, conditioning and stability

Separation distance, fill-distance and power function

$q_X := \frac{1}{2} \min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|_2\,,$ (*separation distance*) $\qquad h_{X,\Omega} := \sup_{\mathbf{x} \in \Omega} \min_{\mathbf{x}_j \in X} \|\mathbf{x} - \mathbf{x}_j\|_2\,,$ (*fill – distance*)

$P_{\Phi_\varepsilon, X}(\mathbf{x}) := \sqrt{\Phi_\varepsilon(\mathbf{0}) - (\mathbf{u}^*(\mathbf{x}))^T \mathbf{A} \mathbf{u}^*(\mathbf{x})}\,,$ (*power function*) $\qquad$ $\mathbf{u}^*$ **vector of cardinal functions**

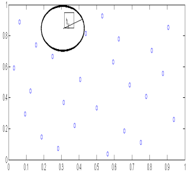


Figure: The fill-distance of 25 Halton points $h \approx 0.2667$

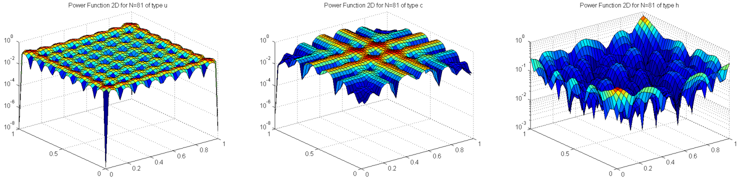


Figure: Power function for the Gaussian kernel with $\varepsilon = 6$ on a grid of 81 uniform, Chebyshev and Halton points, respectively.

Pointwise error estimates

Theorem

Let $\Omega \subset \mathbb{R}^d$ and $K \in \mathcal{C}(\Omega \times \Omega)$ be PD on $\mathbb{R}^d$. Let $X = \{\mathbf{x}_1, \ldots, , \mathbf{x}_n\}$ be a set of distinct points. Take a function $f \in \mathcal{N}_\Phi(\Omega)$ and denote with P_f its interpolant on X. Then, for every $\mathbf{x} \in \Omega$

$$|f(\mathbf{x}) - P_f(\mathbf{x})| \le P_{\Phi_\varepsilon, X}(\mathbf{x}) \|f\|_{\mathcal{N}_K(\Omega)} . \tag{5}$$

Theorem

Let $\Omega \subset \mathbb{R}^d$ and $K \in \mathcal{C}^{2\kappa}(\Omega \times \Omega)$ be symmetric and positive definite, $X = \{\mathbf{x}_1, \ldots, , \mathbf{x}_N\}$ a set of distinct points. Consider $f \in \mathcal{N}_K(\Omega)$ and its interpolant P_f on X. Then, there exist positive constants h_0 and C (independent of $\mathbf{x}$, f and Φ), with $h_{X,\Omega} \le h_0$, such that

$$|f(\mathbf{x}) - P_f(\mathbf{x})| \le C\, h_{X,\Omega}^{\kappa} \sqrt{C_K(\mathbf{x})} \|f\|_{\mathcal{N}_K(\Omega)} . \tag{6}$$

and $C_K(\mathbf{x}) = \max_{|\beta|=2\kappa} \ \max_{\mathbf{w},\mathbf{z} \in \Omega \cup B(\mathbf{x}, c_2 h_{X,\Omega})} |D_2^{\beta} \Phi(\mathbf{w}, \mathbf{z})| .$

Reducing the interpolation error

Obs:

The choice of the shape parameter ε in order to get the smallest (possible) interpolation error is crucial.

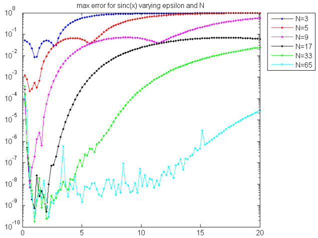
- Trial and Error
- Power function minimization
- Leave One Out Cross Validation (LOOCV)

Reducing the interpolation error

Obs:

The choice of the shape parameter ε in order to get the smallest (possible) interpolation error is crucial.

- Trial and Error
- Power function minimization
- Leave One Out Cross Validation (LOOCV)



Trial and error strategy: interpolation of the 1-d sinc function with Gaussian for $\varepsilon \in [0, 20]$, taking 100 values of ε and different equispaced data points.

Trade-off principle

Consider the errors (RMSE and MAXERR) and the condition number μ_2 of $A = (\Phi_\varepsilon(\|\mathbf{x}_i - \mathbf{x}_j\|))_{i,j=1}^N$

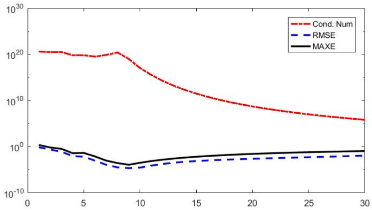


Figure: RMSE, MAXERR and Condition Number, μ_2, with 30 values of $\varepsilon \in [0.1, 20]$, for interpolation of the Franke function on a grid of 40×40 Chebyshev points

Trade-off or uncertainty principle [Schaback 1995]

- Accuracy vs Stability
- Accuracy vs Efficiency
- Accuracy and stability vs Problem size

Accuracy vs stability

- The error bounds for non-stationary interpolation using infinitely smooth basis functions show that the error decreases (exponentially) as the fill distance decreases.
- For well-distributed data a decrease in the fill distance also implies a decrease of the separation distance
- But now the condition estimates above imply that the condition number of A grows exponentially
- This leads to numerical instabilities which make it virtually impossible to obtain the highly accurate results promised by the theoretical error bounds.

Stable basis approaches

Problem setting and questions

Obs:

the standard basis of translates (data-dependent) of $\mathcal{N}_\kappa(X)$ is unstable and not flexible

Problem setting and questions

Obs:

the standard basis of translates (data-dependent) of $\mathcal{N}_\kappa(X)$ is unstable and not flexible

Question 1

Is it possible to find a “better” basis $\mathcal{U}$ of $\mathcal{N}_\kappa(X)$?

Question 2

How to embed information about K and Ω in the basis $\mathcal{U}$?

Question 3

Can we extract $\mathcal{U}' \subset \mathcal{U}$ s.t. s_f' is as good as s_f?

The "natural" basis

The "natural" (data-independent) basis for Hilbert spaces (Mercer's theorem,1909)

Let K be a continuous, positive definite kernel on a bounded $\Omega \subset$. Then K has an eigenfunction expansion with non-negative coefficients, the eigenvalues, s.t.

$$K(x,y) = \sum_{j=0}^{\infty} \lambda_j \varphi_j(x) \varphi_j(y), \quad \forall\, x, y \in \Omega\,.$$

Moreover,

$$\lambda_j \varphi_j(x) = \int_{\Omega} K(x,y)\varphi_j(y)dy := \mathcal{T}[\varphi_j](x)\,, \ \forall x \in \Omega,\ j \geq 0$$

$$\begin{array}{lcl} \{\varphi_j\}_{j>0} & & \textit{orthonormal} \in \mathcal{N}_K(\Omega) \\ \{\varphi_j\}_{j>0} & & \textit{orthogonal} \in L_2(\Omega),\ \|\varphi_j\|^2_{L_2(\Omega)} = \lambda_j \xrightarrow{\infty} 0, \\ \sum_{j>0} \lambda_j & = & K(0,0)\,|\Omega|, \quad \text{(the operator is of trace-class)} \end{array}$$

Notice: the functions φ_i are explicitely in very few cases [Fasshauer,McCourt 2012, for GRBF].

Notation

- $\mathcal{T}_X = \{K(\cdot, x_i),\ x_i \in X\}$: the standard basis of translates;
- $\mathcal{U} = \{u_i \in \mathcal{N}_K(\Omega),\ i = 1, \ldots, N\}$: another basis s.t.

$$\mathrm{span}(\mathcal{U}) = \mathrm{span}(\mathcal{T}_X)\,.$$

Notation

- $\mathcal{T}_X = \{K(\cdot, x_i),\ x_i \in X\}$: the standard basis of translates;
- $\mathcal{U} = \{u_i \in \mathcal{N}_K(\Omega),\ i = 1, \ldots, N\}$: another basis s.t.

$$\mathrm{span}(\mathcal{U}) = \mathrm{span}(\mathcal{T}_X)\,.$$

Change of basis [Pazouki,Schaback 2011]

Let $A_{ij} = K(x_i, x_j) \in \mathbb{R}^{N\times N}$. Any other basis $\mathcal{U}$ arises from a factorization $A \cdot C_{\mathcal{U}} = V_{\mathcal{U}}$ or $A = V_{\mathcal{U}} \cdot C_{\mathcal{U}}{}^{-1}$, where $V_{\mathcal{U}} = (u_j(x_i))_{1\leqslant i,j\leqslant N}$ and $C_{\mathcal{U}}$ is the matrix of change of basis.

- Each $\mathcal{N}_\kappa(\Omega)$-orthonormal basis $\mathcal{U}$ arises from an orthonormal decomposition $A = B^T \cdot B$ with $B = C_{\mathcal{U}}{}^{-1},\ V_{\mathcal{U}} = B^T = (C_{\mathcal{U}}{}^{-1})^T$.
- Each $\ell_2(X)$-orthonormal basis $\mathcal{U}$ arises from a decomposition $A = Q \cdot B$ with $Q = V_{\mathcal{U}}$, $Q^TQ = I$, $B = C_{\mathcal{U}}{}^{-1} = Q^TA$. Notice: the best bases in terms of stability are the $\mathcal{N}_\kappa(\Omega)$-orthonormal ones!

Notation

- $\mathcal{T}_X = \{K(\cdot, x_i),\ x_i \in X\}$: the standard basis of translates;
- $\mathcal{U} = \{u_i \in \mathcal{N}_K(\Omega),\ i = 1, \ldots, N\}$: another basis s.t.

$$\mathrm{span}(\mathcal{U}) = \mathrm{span}(\mathcal{T}_X)\,.$$

Change of basis [Pazouki,Schaback 2011]

Let $A_{ij} = K(x_i, x_j) \in \mathbb{R}^{N\times N}$. Any other basis $\mathcal{U}$ arises from a factorization $A \cdot C_{\mathcal{U}} = V_{\mathcal{U}}$ or $A = V_{\mathcal{U}} \cdot C_{\mathcal{U}}{}^{-1}$, where $V_{\mathcal{U}} = (u_j(x_i))_{1\leqslant i,j\leqslant N}$ and $C_{\mathcal{U}}$ is the matrix of change of basis.

- Each $\mathcal{N}_\kappa(\Omega)$-orthonormal basis $\mathcal{U}$ arises from an orthonormal decomposition $A = B^T \cdot B$ with $B = C_{\mathcal{U}}{}^{-1},\ V_{\mathcal{U}} = B^T = (C_{\mathcal{U}}{}^{-1})^T$.
- Each $\ell_2(X)$-orthonormal basis $\mathcal{U}$ arises from a decomposition $A = Q \cdot B$ with $Q = V_{\mathcal{U}}$, $Q^T Q = I$, $B = C_{\mathcal{U}}{}^{-1} = Q^T A$. Notice: the best bases in terms of stability are the $\mathcal{N}_\kappa(\Omega)$-orthonormal ones!

Q1: Yes, we can!

WSVD Basis

construction

Q2: How to embed information on K and Ω in $\mathcal{U}$?

Symmetric Nyström method [Atkinson,Han 2001]

Idea: discretize the “natural” basis (Mercer’s theorem) by a convergent cubature rule $(X, \mathcal{W})$, with $X = \{x_j\}_{j=1}^N \subset \Omega$ and positive weights $\mathcal{W} = \{w_j\}_{j=1}^N$

WSVD Basis

construction

Q2: How to embed information on K and Ω in $\mathcal{U}$?

Symmetric Nyström method [Atkinson,Han 2001]

Idea: discretize the "natural" basis (Mercer's theorem) by a convergent cubature rule $(X, \mathcal{W})$, with $X = \{x_j\}_{j=1}^N \subset \Omega$ and positive weights $\mathcal{W} = \{w_j\}_{j=1}^N$

$$\lambda_j \varphi_j(x_i) = \int_\Omega K(x_i, y)\varphi_j(y)dy \quad i = 1, \ldots, N, \ \forall j > 0,$$

applying the cubature rule

$$\boxed{\lambda_j \varphi_j(x_i) \approx \sum_{h=1}^N K(x_i, x_h)\varphi_j(x_h)w_h \quad i, j = 1, \ldots, N.} \tag{7}$$

Letting $W = \mathrm{diag}(w_j)$, we reduce to solve the eigen-problem

$$\boxed{\lambda v = (A \cdot W)v} \tag{8}$$

WSVD basis

Cont'

Rewrite (**??**) (using the fact that the weights are positive) as

$$\lambda_j(\sqrt{w_i}\varphi_j(x_i)) = \sum_{h=1}^{N}(\sqrt{w_i}\,K(x_i,x_h)\,\sqrt{w_h})(\sqrt{w_h}\varphi_j(x_h)) \quad \forall i,j=1,\ldots,N, \tag{9}$$

and then to consider the corresponding scaled eigenvalue problem

$$\lambda\left(\sqrt{W}\cdot v\right) = \left(\sqrt{W}\cdot A\cdot\sqrt{W}\right)\left(\sqrt{W}\cdot v\right)$$

which is equivalent to the previous one, now involving the symmetric and positive definite matrix

$$\boxed{A_W := \sqrt{W}\cdot A\cdot\sqrt{W}} \tag{10}$$

WSVD Basis

Definition

$\{\lambda_j, \varphi_j\}_{j>0}$ are then approximated by eigenvalues/eigenvectors of A_W. This matrix is normal, then a *singular value decomposition* of A_W is a unitary diagonalization.

Definition

A *weighted SVD basis* $\mathcal{U}$ is a basis for $\mathcal{N}_\kappa(X)$ s.t.

$$V_{\mathcal{U}} = \sqrt{W^{-1}} \cdot Q \cdot \Sigma, \quad C_{\mathcal{U}} = \sqrt{W} \cdot Q \cdot \Sigma^{-1}$$

since $A = V_{\mathcal{U}} C_{\mathcal{U}}{}^{-1}$, then $Q \cdot \Sigma^2 \cdot Q^T$ is the SVD of A_W.

Here $\Sigma_{jj} = \sigma_j,\ j = 1, \ldots, N$ and $\sigma_1^2 \geq \cdots \geq \sigma_N^2 > 0$ are the singular values of A_W.

WSVD Basis

Properties

This basis is in fact an approximation of the “natural” one (provided $w_i > 0$, $\sum_{i=1}^{N} w_i = |\Omega|$)

WSVD Basis

Properties

This basis is in fact an approximation of the "natural" one (provided $w_i > 0$, $\sum_{i=1}^{N} w_i = |\Omega|$)

Properties of the new basis $\mathcal{U}$ (cf. [De Marchi-Santin 2013])

- $\mathcal{T}_N[u_j](x) = \sigma_j u_j(x), \quad \forall\ 1 \leqslant j \leqslant N,\ \forall x \in \Omega;$
- $\mathcal{N}_\kappa(\Omega)$-orthonormal;
- $\ell_2^w(X)$-orthogonal, $\|u_j\|^2_{\ell_2^w(X)} = \sigma_j^2, \quad \forall u_j \in \mathcal{U};$
- $\displaystyle\sum_{j=1}^{N} \sigma_j^2 = K(0,0)\ |\Omega|.$

WSVD Basis

Approximation

Interpolant: $s_f(x) = \sum_{j=1}^{N} (f, u_j)_K u_j(x) \quad \forall x \in \Omega$

WDLS: $s_f^M := \operatorname{argmin} \left\{ \|f - g\|_{\ell_2^w(X)} : g \in \operatorname{span}\{u_1, \ldots, u_M\} \right\}$

Weighted Discrete Least Squares as truncation

Let $f \in \mathcal{N}_K(\Omega)$, $1 \leqslant M \leqslant N$. Then $\forall x \in \Omega$

$$s_f^M(x) = \sum_{j=1}^{M} \frac{(f, u_j)_{\ell_2^w(X)}}{(u_j, u_j)_{\ell_2^w(X)}} u_j(x) = \sum_{j=1}^{M} \frac{(f, u_j)_{\ell_2^w(X)}}{\sigma_j^2} u_j(x) = \sum_{j=1}^{M} (f, u_j)_K u_j(x)$$

WSVD Basis

Approximation

Interpolant: $s_f(x) = \sum_{j=1}^{N} (f, u_j)_K u_j(x) \quad \forall x \in \Omega$

WDLS: $s_f^M := \operatorname{argmin} \left\{ \|f - g\|_{\ell_2^w(X)} : g \in \operatorname{span}\{u_1, \dots, u_M\} \right\}$

Weighted Discrete Least Squares as truncation

Let $f \in \mathcal{N}_K(\Omega)$, $1 \leqslant M \leqslant N$. Then $\forall x \in \Omega$

$$s_f^M(x) = \sum_{j=1}^{M} \frac{(f, u_j)_{\ell_2^w(X)}}{(u_j, u_j)_{\ell_2^w(X)}} u_j(x) = \sum_{j=1}^{M} \frac{(f, u_j)_{\ell_2^w(X)}}{\sigma_j^2} u_j(x) = \sum_{j=1}^{M} (f, u_j)_K u_j(x)$$

Q3: Can we extract $\mathcal{U}' \subset \mathcal{U}$ s.t. s_f' is as good as s_f?

Yes we can! Take $\mathcal{U}' = \{u_1, \dots, u_M\}$.

WSVD Basis

Approximation II

If we define the pseudo-cardinal functions as $\tilde{\ell}_i = s^M_{\ell_i}$, we get

$$s^M_f(x) = \sum_{i=1}^{N} f(x_i)\tilde{\ell}_i(x), \quad \tilde{\ell}_i(x) = \sum_{j=1}^{M} \frac{u_j(x_i)}{\sigma_j^2} u_j(x).$$

Generalized Power Function and Lebesgue constant

If $f \in \mathcal{N}_K(\Omega)$, $\left|f(x) - s^M_f(x)\right| \leqslant P^{(M)}_{K,X}(x)\|f\|_{\mathcal{N}_K(\Omega)}$ $\forall x \in \Omega$, where

$$\left[P^{(M)}_{K,X}(x)\right]^2 = K(0,0) - \sum_{j=1}^{M}[u_j(x)]^2, \quad (\textit{generalized PF})$$

$$\|s^M_f\|_\infty \leqslant \tilde{\Lambda}_X \|f\|_X \ ,$$

where $\tilde{\Lambda}_X = \max_{x\in\Omega} \sum_{i=1}^{N} |\tilde{\ell}_i(x)|$ is the “pseudo-Lebesgue constant”.

WSVD Basis

Sub-basis

How can we extract $\mathcal{U}' \subset \mathcal{U}$ s.t. s_f' is as good as s_f?

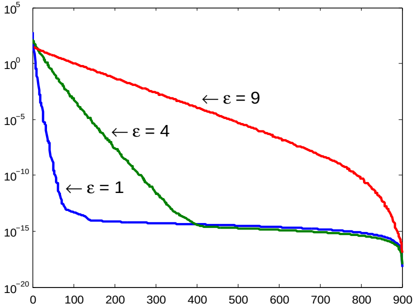


Figure: Gaussian kernel, σ_j^2 on equispaced points of the square

- recall that $\|u_j\|_{\ell_2^w(X)} = \sigma_j^2 \to 0$
- we can choose M s.t. $\boxed{\sigma_{M+1}^2 < \tau}$
- we don't need u_j, $j > M$

WSVD Basis

Example

	$\varepsilon = 1$	$\varepsilon = 4$	$\varepsilon = 9$
Gaussian	100	340	500
IMQ	180	580	580
Matern3	460	560	580

Table: Optimal M for different kernels and shape parameter that correspond to the indexes such that the weighted least-squares approximant s_f^M provides the best approximation of the function $f(x, y) = \cos(20(x + y))$ on the disk with center $C = (1/2, 1/2)$ and radius $R = 1/2$

WSVD Basis

Example, cont'

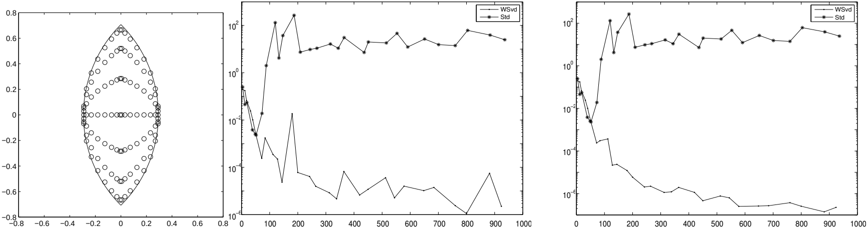


Figure: Franke's test function, *lens*, IMQ Kernel, $\varepsilon = 1$ and RMSE.
Left: complete basis. Right: $\sigma^2_{M+1} < 10^{-17}$.

WSVD Basis

Example, cont'

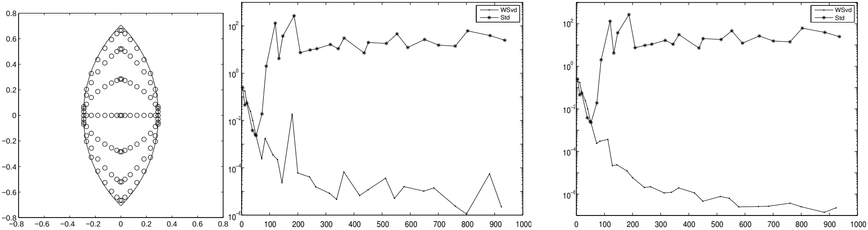


Figure: Franke's test function, *lens*, IMQ Kernel, $\varepsilon = 1$ and RMSE.
Left: complete basis. Right: $\sigma_{M+1}^2 < 10^{-17}$.

Problem: We have to compute the whole basis before truncation!
Solution: Krylov methods, approximate SVD [Novati, Russo 2013]

New fast basis [De Marchi-Santin, BIT15]: a comparison

N	225	529	961	1521
M	110	114	115	116
RMSE new	$3.4 \cdot 10^{-10}$	$6.7 \cdot 10^{-11}$	$5.5 \cdot 10^{-11}$	$3.4 \cdot 10^{-11}$
RMSE WSVD	$3.3 \cdot 10^{-9}$	$1.1 \cdot 10^{-9}$	$8.3 \cdot 10^{-10}$	$7.9 \cdot 10^{-10}$
Time new	$3.4 \cdot 10^{-1}$	$1.0 \cdot 10^{0}$	$2.6 \cdot 10^{0}$	$6.5 \cdot 10^{0}$
Time WSVD	$7.2 \cdot 10^{-1}$	$4.2 \cdot 10^{0}$	$2.5 \cdot 10^{1}$	$1.1 \cdot 10^{2}$

Table: Comparison of the WSVD basis and the new basis. Computational time in seconds and corresponding RMSE for the example consisting of interpolation of a function sum of gaussians by GRBF, restricted to $N = 15^2, 23^2, 31^2, 39^2$ equally spaced points on $[-1, 1]^2$.

Notice: for each N we computed the optimal M using the new algorithm with tolerance $\tau = 10^{-4}$

Rescaled, Rational RBF and some applications

Rescaled RBF Interpolation

- Classical interpolant: $P_f(\mathbf{x}) = \sum_{k=1}^{N} \alpha_k K(\mathbf{x}, \mathbf{x}_k), \quad \mathbf{x} \in \Omega, \ \mathbf{x}_k \in X.$

  [Hardy and Gofert 1975] used multiquadrics
  $K(\mathbf{x}, \mathbf{y}) = \sqrt{1 + \epsilon^2 \|\mathbf{x} - \mathbf{y}\|^2}$.

Rescaled RBF Interpolation

- Classical interpolant: $P_f(\mathbf{x}) = \sum_{k=1}^{N} \alpha_k K(\mathbf{x}, \mathbf{x}_k), \quad \mathbf{x} \in \Omega, \; \mathbf{x}_k \in X.$
  [Hardy and Gofert 1975] used multiquadrics
  $K(\mathbf{x}, \mathbf{y}) = \sqrt{1 + \epsilon^2 \|\mathbf{x} - \mathbf{y}\|^2}.$
- Rescaled interpolant: $\hat{P}_f(\mathbf{x}) = \dfrac{P_f(\mathbf{x})}{P_g(\mathbf{x})} = \dfrac{\sum_{k=1}^{N} \alpha_k K(\mathbf{x}, \mathbf{x}_k)}{\sum_{k=1}^{N} \beta_k K(\mathbf{x}, \mathbf{x}_k)}$ where P_g is the kernel interpolant of $g(\mathbf{x}) = 1, \; \forall \mathbf{x} \in \Omega.$

Obs:

• *Rescaled Localized RBF (RL-RBF) based on CSRBF introduced in [Deparis et al, SISC 2014]: they are smoother even for small radii of the support*

• *In [DeM et al 2017] it is shown that it is a Shepard's PU method: exacteness on constants.*

• *Linear convergence of localized rescaled interpolants is still an open problem [DeM and Wendland, draft 2017].*

Example

Take, $f(x) = x$ on $[0, 1]$ by using W2 at the points set $X = \{0, 1/6, 1/3, 1/2, 2/3, 5/6, 1\}$, $\varepsilon = 5$ ($\epsilon = 1/r$).

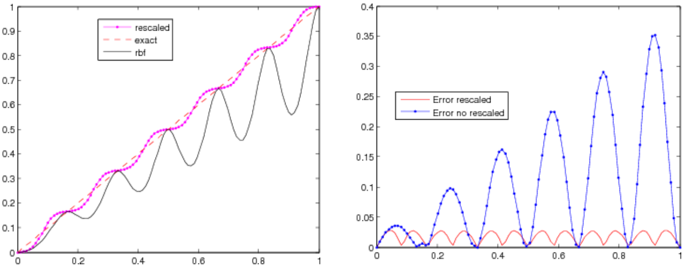


Figure: (left) interpolants and (right) the abs error

For more results see [Deparis et al 14, Idda's master thesis 2015].

RBF- PU interpolation

- $\Omega = \cup_{j=1}^{s} \Omega_i$ (can overlap): that is, each $\mathbf{x} \in \Omega$ belongs to a limited number of subdomains, say $s_0 < s$.
- Then we consider, W_j non-negative functions on Ω_j, s.t. $\sum_{j=1}^{s} W_j = 1$ is a partition of unity . A possibile choice is

$$W_j(\mathbf{x}) = \frac{\tilde{W}_j(\mathbf{x})}{\sum_{k=1}^{s} \tilde{W}_k(\mathbf{x})}, \quad j = 1, \ldots, s \tag{11}$$

where $\tilde{W}_j$ are compactly supported functions on Ω_j.

RBF-PU interpolant

$$I(\mathbf{x}) = \sum_{j=1}^{s} R_j(\mathbf{x}) W_j(\mathbf{x}), \;\; \forall\, x \in \Omega \tag{12}$$

where R_j is a RBF interpolant on Ω_j i.e. $R_j(\mathbf{x}) = \sum_{i=1}^{N_j} \alpha_i^j K(\mathbf{x}, \mathbf{x}_i^j)\,, \;\; N_j = |\Omega_j|$

$\mathbf{x}_k^j \in X_{N_j},\; k = 1, \ldots, N_j$.

Rescaling gives a Shepard method

We start by writing the interpolant of a function $f \in \mathcal{N}_K$ using the cardinals $u_j(\mathbf{x}_i) = \delta_{i,j}$, $\boxed{P_f = \sum_{j=1}^{N} f(\mathbf{x}_j) u_j}$, so that for $g \equiv 1$ we get

$$\boxed{P_g = \sum_{j=1}^{N} u_j}.$$

Rescaling gives a Shepard method

We start by writing the interpolant of a function $f \in \mathcal{N}_K$ using the cardinals $u_j(\mathbf{x}_i) = \delta_{i,j}$, $\boxed{P_f = \sum_{j=1}^{N} f(\mathbf{x}_j) u_j}$, so that for $g \equiv 1$ we get $\boxed{P_g = \sum_{j=1}^{N} u_j}$. The rescaled interpolant is then

$$\boxed{\hat{P}_f = \frac{\sum_{j=1}^{N} f(\mathbf{x}_j) u_j}{\sum_{k=1}^{N} u_k} = \sum_{j=1}^{N} f(\mathbf{x}_j) \frac{u_j}{\sum_{k=1}^{N} u_k} =: \sum_{j=1}^{N} f(\mathbf{x}_j) \hat{u}_j,}$$

where we introduced the (new) cardinal functions $\hat{u}_j := \dfrac{u_j}{\sum_{k=1}^{N} u_k}$.

Rescaling gives a Shepard method

We start by writing the interpolant of a function $f \in \mathcal{N}_K$ using the cardinals $u_j(\mathbf{x}_i) = \delta_{i,j}$, $\boxed{P_f = \sum_{j=1}^{N} f(\mathbf{x}_j) u_j}$, so that for $g \equiv 1$ we get $\boxed{P_g = \sum_{j=1}^{N} u_j}$. The rescaled interpolant is then

$$\boxed{\hat{P}_f = \frac{\sum_{j=1}^{N} f(\mathbf{x}_j) u_j}{\sum_{k=1}^{N} u_k} = \sum_{j=1}^{N} f(\mathbf{x}_j) \frac{u_j}{\sum_{k=1}^{N} u_k} =: \sum_{j=1}^{N} f(\mathbf{x}_j) \hat{u}_j,}$$

where we introduced the (new) cardinal functions $\hat{u}_j := \dfrac{u_j}{\sum_{k=1}^{N} u_k}$.

Corollary

The rescaled interpolation method is a Shepard-PU method, where the weight functions are defined as $\hat{u}_j = u_j / \left(\sum_{k=1}^{N} u_k\right)$, $\{u_j\}_j$ being the cardinal basis of $\mathrm{span}\{K(\cdot, \mathbf{x}), \mathbf{x} \in X\}$.

Variably Scaled Kernels (VSK)

Definition

Let $\psi : \mathbb{R}^d \to (0, \infty)$ be a given scale function. A Variably Scaled Kernel (VSK) K_ψ on $\mathbb{R}^d$ is

$$K_\psi(\mathbf{x}, \mathbf{y}) := \mathcal{K}((\mathbf{x}, \psi(\mathbf{x})), (\mathbf{y}, \psi(\mathbf{y}))), \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

where $\mathcal{K}$ is a kernel on $\mathbb{R}^{d+1}$.

Variably Scaled Kernels (VSK)

Definition

Let $\psi : \mathbb{R}^d \to (0, \infty)$ be a given scale function. A Variably Scaled Kernel (VSK) K_ψ on $\mathbb{R}^d$ is

$$K_\psi(\mathbf{x}, \mathbf{y}) := \mathcal{K}((\mathbf{x}, \psi(\mathbf{x})), (\mathbf{y}, \psi(\mathbf{y}))), \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

where $\mathcal{K}$ is a kernel on $\mathbb{R}^{d+1}$.

Given X and the scale functions $\psi_j : \mathbb{R}^d \to (0, \infty)$, $j = 1, \ldots, s$, the RVSK-PU is

$$\mathcal{I}_\psi(\mathbf{x}) = \sum_{j=1}^{s} R_{\psi_j}(\mathbf{x})\, W_j(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{13}$$

with $R_{\psi_j}(\mathbf{x})$ the RVSK-PU on Ω_j.

Variably Scaled Kernels (VSK)

Definition

Let $\psi : \mathbb{R}^d \to (0, \infty)$ be a given scale function. A Variably Scaled Kernel (VSK) K_ψ on $\mathbb{R}^d$ is

$$K_\psi(\mathbf{x}, \mathbf{y}) := \mathcal{K}((\mathbf{x}, \psi(\mathbf{x})), (\mathbf{y}, \psi(\mathbf{y}))), \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^d.$$

where $\mathcal{K}$ is a kernel on $\mathbb{R}^{d+1}$.

Given X and the scale functions $\psi_j : \mathbb{R}^d \to (0, \infty)$, $j = 1, \ldots, s$, the RVSK-PU is

$$\mathcal{I}_\psi(\mathbf{x}) = \sum_{j=1}^{s} R_{\psi_j}(\mathbf{x})\, W_j(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{13}$$

with $R_{\psi_j}(\mathbf{x})$ the RVSK-PU on Ω_j.

Obs: (if Φ is radial)

$$(A_{\psi_j})_{ik} = \Phi(\|\mathbf{x}_i^j - \mathbf{x}_k^j\|^2 + (\psi_j(\mathbf{x}_i^j) - \psi_j(\mathbf{x}_k^j))^2), \quad i, k = 1, \ldots, N_j.$$

Rational RBF (RRBF)

definition

$$R(\mathbf{x}) = \frac{R^{(1)}(\mathbf{x})}{R^{(2)}(\mathbf{x})} = \frac{\sum_{k=1}^{N} \alpha_k K(\mathbf{x}, \mathbf{x}_k)}{\sum_{k=1}^{N} \beta_k K(\mathbf{x}, \mathbf{x}_k)}$$

[Jackbsson et al. 2009, Sarra and Bai 2017]

$\Longrightarrow$ RRBFs well approximate data with steep gradients or discontinuites [rational with PU+VSK in DeM et al. 2017].

Rational RBF

Learning from rational functions, $d = 1$

- polynomial case.

$$r(x) = \frac{p_1(x)}{p_2(x)} = \frac{a_m x^m + \cdots + a_0 x^0}{x^n + b_{n-1} x^{n-1} \cdots + b_0}.$$

$M = m + n + 1$ unknowns (Padé approximation). If $M < N$ to get the coefficients we may solve the LS problem

$$\min_{p_1 \in \Pi_m^1, p_2 \in \Pi_n^1} \left(\sum_{k=1}^{N} \left| f(x_k) - r(x_k) \right|^2 \right).$$

Rational RBF

Learning from rational functions, $d = 1$

- polynomial case.

$$r(x) = \frac{p_1(x)}{p_2(x)} = \frac{a_m x^m + \cdots + a_0 x^0}{x^n + b_{n-1} x^{n-1} \cdots + b_0}.$$

$M = m + n + 1$ unknowns (Padé approximation). If $M < N$ to get the coefficients we may solve the LS problem

$$\min_{p_1 \in \Pi_m^1, p_2 \in \Pi_n^1} \left(\sum_{k=1}^{N} \left| f(x_k) - r(x_k) \right|^2 \right).$$

- RBF case. Let $X_m = \{\mathbf{x}_k, \ldots, \mathbf{x}_{k+m-1}\}, X_n = \{\mathbf{x}_j, \ldots, \mathbf{x}_{j+n-1}\} \subset X$ be non empty, such that $m + n \leq N$

$$R(\mathbf{x}) = \frac{R^{(1)}(\mathbf{x})}{R^{(2)}(\mathbf{x})} = \frac{\sum_{i_1=k}^{k+m-1} \alpha_{i_1} K(\mathbf{x}, \mathbf{x}_{i_1})}{\sum_{i_2=j}^{j+n-1} \beta_{i_2} K(\mathbf{x}, \mathbf{x}_{i_2})}, \qquad (14)$$

provided $R^{(2)}(\mathbf{x}) \neq 0$, for all $\mathbf{x} \in \Omega$.

Rational RBF

Find the coefficients

[Jackbsson 2009] show that this is equivalent to solve the following generalized eigenvalue problem

Problem 3

$$\Sigma \boldsymbol{q} = \lambda \Theta \boldsymbol{q},$$

with

$$\Sigma = \frac{1}{||\boldsymbol{f}||_2^2} D^T A^{-1} D + A^{-1}, \quad \text{and} \quad \Theta = \frac{1}{||\boldsymbol{f}||_2^2} D^T D + I_N,$$

where I_N is the identity matrix.

Rational RBF

Find the coefficients

[Jackbsson 2009] show that this is equivalent to solve the following generalized eigenvalue problem

Problem 3

$$\Sigma \boldsymbol{q} = \lambda \Theta \boldsymbol{q},$$

with

$$\Sigma = \frac{1}{||\boldsymbol{f}||_2^2} D^T A^{-1} D + A^{-1}, \quad \text{and} \quad \Theta = \frac{1}{||\boldsymbol{f}||_2^2} D^T D + I_N,$$

where I_N is the identity matrix.

$\hookrightarrow$ **q** is the eigenvector associated to the smallest eigenvalue! $\hookleftarrow$

New class of rational RBF [Buhmann, DeM, Perracchione 2018]

$$\hat{P}_f(\boldsymbol{x}) = \frac{\sum_{i=1}^{N} \alpha_i K(\boldsymbol{x}, \boldsymbol{x_i}) + \sum_{m=1}^{L} \gamma_m p_m(\boldsymbol{x})}{\sum_{k=1}^{N} \beta_k \bar{K}(\boldsymbol{x}, \boldsymbol{x_k})} := \frac{P_g(\boldsymbol{x})}{P_h(\boldsymbol{x})} \tag{15}$$

Ratio of a CPD K of order ℓ and an associate PD $\bar{K}$ .... $\Longrightarrow$ two kernel matrices, Φ_K and $\Phi_{\bar{K}}$.

New class of rational RBF [Buhmann, DeM, Perracchione 2018]

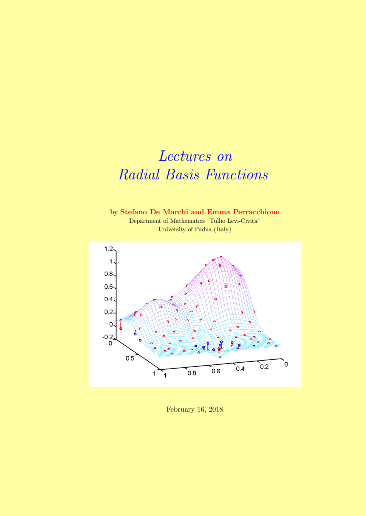
$$\hat{P}_f(\boldsymbol{x}) = \frac{\sum_{i=1}^{N} \alpha_i K(\boldsymbol{x}, \boldsymbol{x_i}) + \sum_{m=1}^{L} \gamma_m p_m(\boldsymbol{x})}{\sum_{k=1}^{N} \beta_k \bar{K}(\boldsymbol{x}, \boldsymbol{x_k})} := \frac{P_g(\boldsymbol{x})}{P_h(\boldsymbol{x})} \tag{15}$$

Ratio of a CPD K of order ℓ and an associate PD $\bar{K}$ .... $\Longrightarrow$ two kernel matrices, Φ_K and $\Phi_{\bar{K}}$.

Obs:

1. *Once we know the function values $P_h(\boldsymbol{x}_i) = h_i$, $i = 1, \ldots, N$, we can construct P_g, i.e. it interpolates $\boldsymbol{g} = (f_1 h_1, \ldots, f_N h_N)^T$. Hence $\hat{P}_f$ interpolates the given function values* **f** *at the nodes* $\mathbf{X}_N$.
2. *If K is PD, we fix $\bar{K} = K$ so that we deal with the same kernel matrix for both numerator and denominator.*
3. *We can prove well-posedness, find cardinal functions and give stability analysis.*

Lectures Notes on RBF


Lectures on
Radial Basis Functions

by Stefano De Marchi and Emma Perracchione
Department of Mathematics "Tullio Levi-Civita"
University of Padua (Italy)

February 16, 2018

1. Learning from splines
2. Positive definite functions
3. Conditionally positive definite functions
4. Error estimates
5. Stable bases for RBFs
6. Rational RBFs
7. The Partition of unity method
8. Collocation method via RBF
9. Financial applications
10. Exercises

Other references

R. Cavoretto, S. De Marchi, A. De Rossi, E. Perracchione, G. Santin, *Partition of unity interpolation using stable kernel-based techniques*, Appl. Numer. Math. **116** (2017), pp. 95–107

R. Cavoretto, A. De Rossi, E. Perracchione, *Optimal selection of local approximants in RBF-PU interpolation*, to appear on J. Sci. Comput. (2017).

S. De Marchi, G. Santin, *Fast computation of orthonormal basis for RBF spaces through Krylov space methods*, BIT **55** (2015), pp. 949–966.

Stefano De Marchi, Andrea Idda and Gabriele Santin: A rescaled method for RBF approximation. Springer Proceedings on Mathematics and Statistics, Vol. 201, pp.39–59. (2017).

S. De Marchi, E. Perracchione, A. Martinez Calomardo and M. Rossini: RBF-based partition of unity method for elliptic PDEs: adaptivity and stability issues via VSKs, submitted to J. Sci. Comput. 2017

S. De Marchi, E. Perracchione, A. Martinez Calomardo : Rational Radial Basis Functions interpolation , submitted to J. Comput. Appl. Math. 2017

M. Buhmann, S. De Marchi, E. Perracchione : Analysis of a new class of rational RBF expansions, submitted to IMA Num. Analysis, 2018.

E. Perracchione: Rational RBF-based partition of unity method for efficiently and accurately approximating 3D objects arXiv preprint arXiv:1802.01842, 2018. To appear in Comput. Appl. Math.

grazie per la vostra attenzione!

thanks for your attention!

and grazie a R.IT.A.