On a new orthonormal basis for RBF native spaces

Stefano De Marchi and Gabriele Santin
San Diego: July 8, 2013

Università degli Studi di Padova

1. **Data:** $\Omega \subset \mathbb{R}^n$, $X \subset \Omega$, test function f
   - $X = \{x_1, \ldots, x_N\} \subset \Omega$
   - $f_1, \ldots, f_N$, where $f_i = f(x_i)$
2. **Approximation setting:** kernel K_ε, $\mathcal{N}_K(\Omega)$, $\mathcal{N}_K(X) \subset \mathcal{N}_K(\Omega)$
   - kernel $K = K_\varepsilon$, positive definite and radial
   - native space $\mathcal{N}_K(\Omega)$ (where K is the reproducing kernel)
   - finite subspace $\mathcal{N}_K(X) = \mathrm{span}\{K(\cdot, x) : x \in X\} \subset \mathcal{N}_K(\Omega)$

Aim

Find $s_f \in \mathcal{N}_K(X)$ s.t. $s_f \approx f$

Problem: the standard (data-dependent) basis of $\mathcal{N}_K(X)$ is unstable and not flexible

Question 1

Is it possible to find a "better" basis $\mathcal{U}$ of $\mathcal{N}_K(X)$?

Question 2

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

Question 3

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

Change of basis

Università degli Studi di Padova

Q1: It is possible to find a "better" basis?

Change of basis ([Pazouki-Schaback 2011]):

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

  The best bases in terms of stability are the $\mathcal{N}_K(\Omega)$-o.n. ones.

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

Let K be a continuous, positive definite kernel on a bounded $\Omega \subset \mathbb{R}^n$. Then K has an eigenfunction expansion

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

Moreover,

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

$$\begin{array}{lcl} \{\varphi_j\}_{j>0} & & \textit{orthonormal} \in \mathcal{N}_K(\Omega) \\ \{\varphi_j\}_{j>0} & & \textit{orthogonal} \in L_2(\Omega),\ \|\varphi_j\|^2_{L_2(\Omega)} = \lambda_j \overset{\infty}{\longrightarrow} 0, \\ \sum_{j>0} \lambda_j & = & K(0,0)\,|\Omega| \end{array}$$

WSVD Basis

Definition

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

Idea: Approximate the integral equation $\lambda_j \varphi_j(x) = \mathcal{T}[\varphi_j](x)$ with the *symmetric Nyström method*, with a cubature formula $(X, \mathcal{W})$: $\{\lambda_j, \varphi_j\}_{j>0}$ are approximated by eigenvalues/eigenvectors of $A_W := \sqrt{W} \cdot A \cdot \sqrt{W}$, with $W = \mathtt{diag}(w_j)$, i.e. the solution of the scaled eigenvalue problem $\boxed{\lambda(\sqrt{W} \cdot v) = A_W(\sqrt{W} \cdot v)}$.

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 $A_W = Q \cdot \Sigma^2 \cdot Q^T$ is the SVD (and unitary diagonalization).

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

Properties (cf. [De Marchi-Santin 2013])

- $\sigma_j^2 \, u_j(x) = \mathcal{T}_{K_N}[u_j](x)$
  where $T_{K_N}[u_j](\cdot) := (u_j, K(\cdot, x))_{\ell_2^w(X)}$ $\forall j,\ \forall x \in \Omega$
- $\mathcal{N}_K(\Omega)$-orthonormal
- $\ell_2^w(X)$-orthogonal, $\|u_j\|^2_{\ell_2^w(X)} = \sigma_j^2 \quad \forall j$
- $\sum_{j=1}^{N} \sigma_j^2 = K(0,0)\,|\Omega|$

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

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

WDLS as truncation:

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

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

Q3: Can we extract $\mathcal{U}' \subset \mathcal{U}$ s.t. s_f' is as good as s_f? We can take $\mathcal{U}' = \{u_1, \dots, u_M\}$.

UNIVERSITÀ DEGLI STUDI DI PADOVA

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

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

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

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

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

Q3: Can we extract $\mathcal{U}' \subset \mathcal{U}$ s.t. s_f' is as good as s_f? We can take $\mathcal{U}' = \{u_1, \dots, u_M\}$

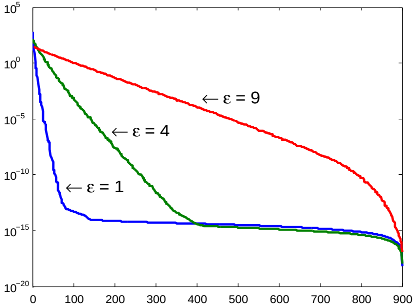

Figure: Gaussian kernel, σ_j^2 on equispaced points of the square

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

WSVD Basis

An Example: I

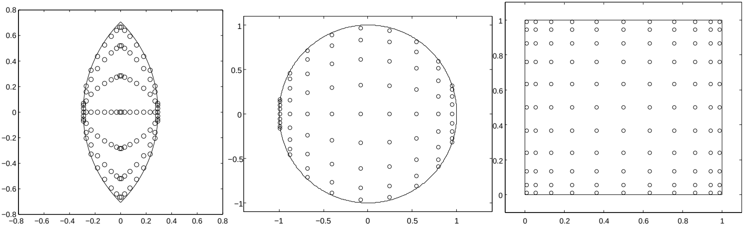

Figure: The domains used in the numerical experiments with an example of the corresponding sample points. From left to right: the lens Ω_1 (trigonometric-gaussian points), the disk Ω_2 (trigonometric-gaussian points) and the square Ω_3 (product Gauss-Legendre points).

	$\varepsilon = 1$	$\varepsilon = 4$	$\varepsilon = 9$
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

An Example: III

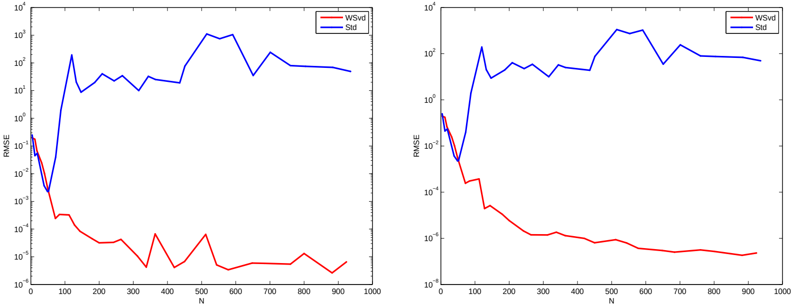

Figure: Franke's function, *lens*, IMQ Kernel, $\varepsilon = 1$ and RMSE. Left: complete basis. Right: $\sigma_{M+1}^2 < 10^{-17}$.

Problem: We have to compute the whole basis before truncation!

Solution: Krylov methods.

Consider $A_{ij} = K(x_i, x_j)$, $b_i = f(x_i)$, $1 \leqslant i, j \leqslant N$

- define the Krylov subspace $\mathcal{K}_M(A, b) = \mathrm{span}\{b, Ab, \ldots, A^{M-1}b\}$, $M \ll N$.
- compute an o.n. basis $\{\phi_1, \ldots, \phi_M\}$ of $\mathcal{K}_M(A, b)$
- define the (tridiagonal) matrix H_M which represents the projection of A into $\mathcal{K}_M(A, b)$ (since $\Phi_M^T A \Phi_M = H_M$, with $\Phi_M = [\phi_1, \ldots, \phi_M],\ N \times M$)

Arnoldi's algorithm gives $A\Phi_M = \Phi_{M+1}\overline{H}_M$ where $\overline{H}_M = \begin{bmatrix} H_M \\ h_{M+1,M} e_M^T \end{bmatrix}$ is $(M+1) \times M$. In practise $h_{M+1,M} \approx 0$ so that $\mathcal{K}_{M+1}(A, b) = \mathcal{K}_M(A, b)$.

Consider a SVD $\overline{H}_M = U_M \Sigma_M^2 V_M^T$, where $U_M \in \mathbb{R}^{(M+1)\times(M+1)}$, $V_M \in \mathbb{R}^{M\times M}$, $\Sigma_M^2 = \left[\tilde{\Sigma}_M^2, 0\right]^T$ and $\tilde{\Sigma}_M^2 = \mathrm{diag}(\sigma_{M,1}^2, \ldots, \sigma_{M,M}^2)$.

Approximate SVD (Novati-Russo 2013:)

Let $\overline{U}_M = \Phi_{M+1} U_M$, $\overline{V}_M = \Phi_M V_M$, then

- $A\overline{V}_M = \overline{U}_M \Sigma_M^2$, $A\overline{U}_M = \overline{V}_M (\Sigma_M^2)^T$
- the first M singular values of A are well approx. by $\sigma_{M,j}^2$
- If $M = N$, in exact arithmetic the triplet $\left(\Phi_{M+1}\tilde{U}_M, \tilde{\Sigma}_M, \Phi_M V_M\right)$ is a SVD of A, where $\tilde{U}_M$ is U_M without the last column.

The new basis

Definition

Recall:

- $A\Phi_M = \Phi_{M+1}\overline{H}_M$
- $\overline{H}_M = U_M \Sigma_M^2 V_M^T$
- $\Sigma_M^2 = \left[\tilde{\Sigma}_M^2, 0\right]^T$
- $\tilde{U}_M$ is U_M without the last column.

Definition:

The sub-basis $\mathcal{U}_M$ is a set $\{u_1, \dots, u_M\} \subset \mathcal{N}_\kappa(X)$ defined by

$$V_{\mathcal{U}} = \Phi_{M+1}\tilde{U}_M\tilde{\Sigma}_M, \quad C_{\mathcal{U}} = \Phi_M V_M \tilde{\Sigma}_M^{-1}.$$

Università degli Studi di Padova

Properties:

The sub-basis $\mathcal{U}_M$ has the following properties for each $1 \leqslant M \leqslant N$:

- it is $\mathcal{N}_\kappa(\Omega)$-orthonormal
- it is $\ell_2(X)$-orthogonal with $\|u_j\|_{\ell_2(X)} = \sigma^2_{M,j}$ $1 \leqslant j \leqslant M$
- if $M = N$ it is the SVD basis $\mathcal{U}$ $(\Phi_M = I)$

Using this basis we get $\forall f \in \mathcal{N}_\kappa(\Omega)$

$$s'_f(x) = \sum_{j=1}^{M} \frac{(f, u_j)_{\ell_2(X)}}{\sigma^2_{M,j}} u_j(x) = \sum_{j=1}^{M} (f, u_j)_K u_j(x) \quad \forall x \in \Omega$$

(and $P^{(M)}_{K,X}(x)$, $\tilde{\Lambda}_X$ as before)

Numerical Results

Stopping rule

$$\left(\overline{H}_M\right)_{M+1,M} \approx \sigma^2_{M,j} < 10^{-13}$$

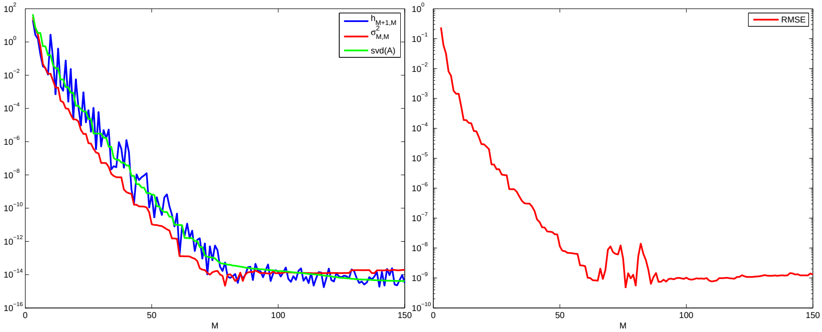

Figure: Gaussian kernel, $\varepsilon = 1$, square $[-1,1]^2$, $N = 200$ e.s. points, $f \in \mathcal{N}_K(\Omega)$, with $f(x) = K(x,y1) + 2K(x,y2) - 2K(x,y3) + 3K(x,y4)$, $y1 = (0,-1.2)$, $y2 = (-0.4,0.5)$, $y3 = (-0.4,1.1)$, $y4 = (1.2,1.3)$.

Numerical Results

Comparison with the SVD basis

Università degli Studi di Padova

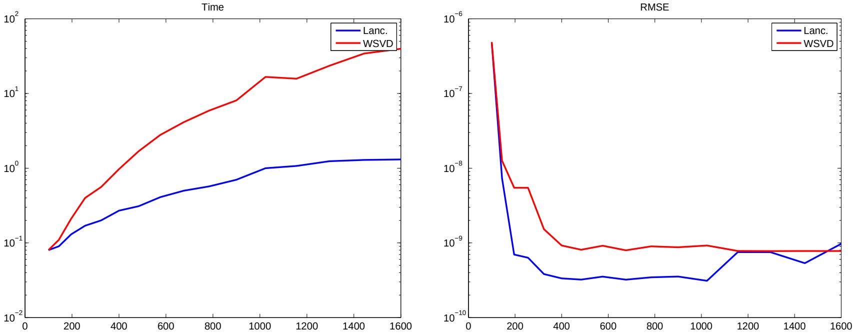

Figure: Gaussian kernel, $\varepsilon = 1$, disk, *trig.-gauss. points*, $f \in \mathcal{N}_K(\Omega)$

Numerical Results

An example

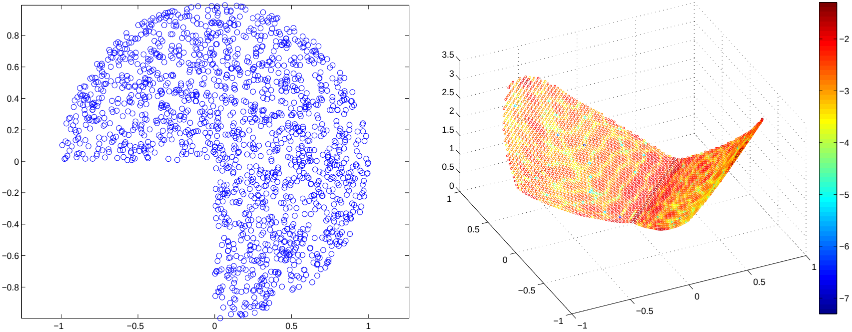

Figure: IMQ kernel, $\varepsilon = 1$, cutted-disk, $N = 1600$ random points, $M = 260$, $f(x, y) = exp(|x - y|) - 1$

Numerical Results

Lebesgue Constant and Power Function

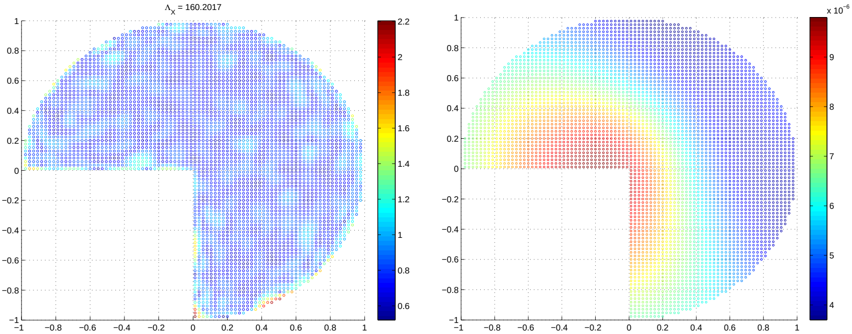

Figure: IMQ kernel, $\varepsilon = 1$, cutted-disk, $N = 1600$ random points, $M = 260$, $f(x, y) = exp(|x - y|) - 1$. Left: Lebesgue function. Right: power function

Further work

Further investigation is needed:

- a better stopping rule
- understand the decay rate of $P_{K,X}^{(M)}$
- understand the growing rate of $\tilde{\Lambda}_X$
- understand how X, ε influence s_f'

Thank you for your attention

References

R. K. Beatson, J. B. Cherrie, C. T. Mouat .
Fast fitting of radial basis functions: methods based on preconditioned GMRES iteration.
Adv. Comp. Math., **11** : 253–270, 1999.

S. De Marchi, G. Santin.
A new stable basis for radial basis function interpolation.
J. Comput. Appl. Math., **253** : 1–13, 2013.

G. Fasshauer, M. J. McCourt.
Stable evaluation of Gaussian radial basis functions interpolants.
SIAM J. Sci. Comput., **34**: A737–A762, 2012.

A. C. Faul, M. J. D. Powell.
Krylov subspace methods for radial basis function interpolation.
Chapman & Hall/CRC Res. Notes Math., **420** : 115–141, 1999.

P. Novati, M. R. Russo.
A GCV based Arnoldi-Tikhonov regularization method.
preprint, 2013.

M. Pazouki, R. Schaback.
Bases for kernel-based spaces.
J. Comput. Appl. Math., **236** : 575–588, 2011.

M. Vianello and G. Da Fies.
Algebraic cubature on planar lenses and bubbles.
Dolomites Res. Notes Approx. DRNA **15**, 7–12, 2012.