

A new stable basis for radial basis function interpolation

Chiara Todesco

Università degli studi di Padova

12 December 2017

Presentation

- 1 Problem and Basic Notions
- 2 "Natural" basis
- 3 General bases
- 4 Weighted SVD bases
- 5 Weighted discrete least-square approximation
- 6 Numerical examples (Part 1)
- 7 The new basis
- 8 Lanczos method
- 9 Construction of the new basis
- 10 Approximation
- 11 Numerical examples (Part 2)

Problem

The purpose of approximation theory is the reconstruction of given function defined on a set $\Omega \subseteq \mathbb{R}^s$ from some values sampled at a finite set $X \subset \Omega$. This process is required to be **convergent and stable**, namely the approximant should reproduce the original function in a chosen norm.

Problem : The **kernel methods** although they are built to be well-posed for every data distribution, it is also well known that the interpolation based on translates of radial basis functions (RBF) is **numerically unstable** due to the ill-conditioning of the kernel matrix.

Objective : Find a general way to build stable and orthonormal bases for the native space $\mathcal{N}_\Phi(\Omega)$ associated to a kernel $\Phi: \Omega \times \Omega \rightarrow \mathbb{R}$ with $\Omega \subseteq \mathbb{R}^s$, based on a suitable factorization of the kernel matrix

$$A := (\Phi(x_i, x_j)).$$

Furthermore, after that, find a method to compute in a fast way the basis using methods related to Krylov subspaces.

Basic Notions

Consider Φ strictly positive definite and radial kernel defined on a set $\Omega \subseteq \mathbb{R}^s$, and X is a finite set of $n \in \mathbb{N}$ distinct points in Ω . Since the kernel is radial, there exist a unique function $\varphi: [0, +\infty) \rightarrow \mathbb{R}$ and $\varepsilon \in \mathbb{R}_{>0}$ such that

$$\Phi(x, y) = \Phi_\varepsilon(x, y) = \varphi(\varepsilon \|x - y\|_2) \quad \forall x, y \in \Omega$$

Theorem

Every SPD Φ on domain Ω has a unique native Hilbert space $\mathcal{N}_\Phi(\Omega)$. It is the closure of the space

$$H_\Phi(\Omega) := \left\{ \sum_{j=1}^M \lambda_j \Phi(x_j, \cdot) : \lambda_j \in \mathbb{R}, M \in \mathbb{N}, x_j \in \Omega \right\}$$

of all function of form $s(x) = \sum_{j=1}^M \lambda_j \Phi(\|x_j - x\|_2)$ under the inner product

$$(\Phi(x, \cdot), \Phi(y, \cdot))_\Phi = \Phi(x, y) \quad \forall x \in \Omega, \forall f \in H_\Phi(\Omega).$$

The elements of the native space can be interpreted as function via the reproduction formula

$$f(x) = \delta_x(f) = (f, \Phi(x, \cdot))_\Phi \quad \forall x \in \Omega, \forall f \in H_\Phi(\Omega)$$

Theorem

Assume $\int_{\Omega} \Phi(x, x) dx < +\infty$ to hold for a SPD Φ on Ω . Then the integral operator

$$C[v](x) = \int_{\Omega} v(t)\Phi(x, t) dt$$

of generalized convolution type maps $L_2(\Omega)$ continuously into the native $\mathcal{N}_{\Phi}(\Omega)$. It satisfies

$$(f, v)_{L_2(\Omega)} = (f, C[v])_{\Phi} \quad \forall f \in \mathcal{N}_{\Phi}(\Omega), v \in L_2(\Omega)$$

We will consider the operator $T_{\Phi} : L_2(\Omega) \rightarrow \mathcal{N}_{\Phi}(\Omega) \subseteq L_2(\Omega)$ defined by

$$T_{\Phi}[f](x) = \int_{\Omega} \Phi(x, y)f(y) dy \quad \forall f \in L_2(\Omega), \forall x \in \Omega$$

that maps $L_2(\Omega)$ continuously into $\mathcal{N}_{\Phi}(\Omega)$. It is the adjoint of the embedding operator of $\mathcal{N}_{\Phi}(\Omega)$ into $L_2(\Omega)$ i.e.

$$(f, v)_{L_2(\Omega)} = (f, T_{\Phi}[v])_{\Phi} \quad \forall f \in \mathcal{N}_{\Phi}(\Omega), v \in L_2(\Omega)$$

Basic Notions

If Φ is SPD function on Ω , one can interpolate any function $f \in \mathcal{N}_\Phi(\Omega)$ on any scattered set $\{x_1, \dots, x_M\} \subset \Omega$ by an unique function

$$s_f^*(x) = \sum_{j=1}^M \lambda_j \varphi(\|x_j - x\|_2)$$

. The error functional

$$\varepsilon_x^* : f \mapsto f(x) - s_f^*(x)$$

is in the dual of the native space, and it's norm $P^*(x) = \|\varepsilon_x^*\|_\Phi$ $x \in \Omega$ is called the Power function. the standard error bound is

$$|f(x) - s_f^*(x)| \leq P^*(x) \|f\|_\Phi \quad \forall f \in \mathcal{N}_\Phi(\Omega), \forall x \in \Omega$$

"Natural" basis

A particular and in some sense "natural" basis for $\mathcal{N}_\Phi(\Omega)$ comes from
Mercer's theorem

Theorem (Mercer)

Every continuous positive definite kernel Φ on bounded domain $\Omega \subseteq \mathbb{R}^s$, defines an operator $T_\Phi : \mathcal{N}(\Omega) \rightarrow \mathcal{N}(\Omega)$

$$T_\Phi[f](x) = \int_{\Omega} \Phi(x, y) f(y) dy$$

which is bounded, compact and self-adjoint. It has an enumerable set of eigenvalues $\{\lambda_j\}_{j>0}$ s.t. $\lambda_1 \geq \lambda_2 \geq \dots > 0$ and eigenvectors $\{\varphi_j\}_{j>0}$ i.e.

$$\lambda_j \varphi_j = \int_{\Omega} \Phi(x, y) \varphi_j(y) dy \quad \forall x \in \Omega$$

*which form an **orthonormal** basis for $\mathcal{N}_\Phi(\Omega)$.*

"Natural" basis

Theorem (Merzel)

In particular

- (1) $\{\varphi_j\}_{j>0}$ is orthonormal in $\mathcal{N}_\Phi(\Omega)$;
- (2) $\{\varphi_j\}_{j>0}$ is orthogonal in $L_2(\Omega)$: $\|\varphi_j\|_{L_2(\Omega)}^2 = \lambda_j$;
- (3) $\lim_{j \rightarrow +\infty} \lambda_j = 0$;

Morover the kernel has a series expansions

$$\Phi(x, y) = \sum_{j=1}^{+\infty} \varphi_j(x)\varphi_j(y) \quad \forall x, y \in \Omega$$

where the $\{\varphi_j\}$ are eigenfunctions, which is absolutely and uniformly convergent

"Natural" basis

Remark

The operator T_Φ is a trace class operator, that is to say

$$\sum_{j>0} \lambda_j = \int_{\Omega} \Phi(x, x) dx = \varphi(0) |\Omega|$$

where $|\Omega| := \text{meas}(\Omega)$. Moreover as consequence of Mercer's theorem's property (2) $\forall j > 0$

$$\begin{aligned} (f, \varphi_j)_{L_2(\Omega)} &= (f, T_\Phi[\varphi_j])_\Phi = \lambda_j (f, \varphi_j)_\Phi = \|\varphi_j\|_{L_2(\Omega)}^2 (f, \varphi_j)_\Phi \\ &= (\varphi_j, \varphi_j)_{L_2(\Omega)} (f, \varphi_j)_\Phi. \quad \forall f \in \mathcal{N}_\Phi(\Omega) \end{aligned}$$

General bases

We show the connection between a change of basis and a decomposition of the kernel matrix A .

Let $\Omega \subseteq \mathbb{R}^s$, $X = \{x_1, \dots, x_N\} \subset \Omega$ and let be

$\mathcal{T}_X = \{\Phi(\cdot, x_i) : x_i \in X\}$ standard basis of translates

$\mathcal{U} = \{u_i \in \mathcal{N}_\Phi(\Omega), i = 1, \dots, N\}$ new basis s.t.

$$\mathcal{N}_\Phi(X) := \text{span}(\mathcal{U}) = \text{span}(\mathcal{T}_X)$$

At $x \in \Omega$, \mathcal{T}_X and \mathcal{U} can be expressed as row vectors:

$$T(x) = [\Phi(x, x_1), \dots, \Phi(x, x_N)] \in \mathbb{R}^N$$

$$U(x) = [u_1(x), \dots, u_N(x)] \in \mathbb{R}^N$$

General bases

Theorem (characterization of the basis \mathcal{U})

Any basis \mathcal{U} arises from a factorization of the kernel matrix A :

$$A = V_{\mathcal{U}} \cdot C_{\mathcal{U}}^{-1}$$

where

- ▷ $V_{\mathcal{U}} = (u_j(x_i))_{1 \leq i, j \leq N}$ matrix of the new basis valuated at X ;
- ▷ $C_{\mathcal{U}} = (c_{ij})_{1 \leq i, j \leq N}$ matrix of change of basis that is $U(x) = T(x) \cdot C_{\mathcal{U}}$ i.e.

$$u_j(x) = \sum_{i=1}^N c_{ij} \Phi(x, x_i)$$

General bases

Proposition

The interpolant $P_X[f]$ on $X \subset \Omega$ of a function $f \in \mathcal{N}_\Phi(\Omega)$ can be rewritten as

$$P_X[f](x) = \sum_{j=1}^N \Delta_j(f) u_j(x) = U(x) \cdot \Delta(f) \quad \forall x \in \Omega$$

where $\Delta(f) = [\Delta_1(f), \dots, \Delta_N(f)]^T \in \mathbb{R}^N$ is a column vector of values of linear function defined by

$$\Delta(f) = C_U^{-1} A^{-1} E_X(f) = V_U^{-1} E_X(f)$$

while $E_X(f)$ is the column vector given by the evaluation of f on X

Proposition (stability estimate)

Let $G_{\mathcal{U}} := ((u_i, u_j)_{\Phi})_{1 \leq i, j \leq N}$, and $\mathcal{K}_2(G_{\mathcal{U}})$ the corresponding 2-condition number. Then $\forall x \in \Omega$

$$|P_{\mathcal{X}}[f](x)|^2 \leq \|U(x)\|_2^2 \|\Delta_{\mathcal{U}}(f)\|_2^2 \leq \mathcal{K}_2(G_{\mathcal{U}}) \varphi(0) \|f\|_{\Phi}^2$$

Corollary

If \mathcal{U} is a Φ -orthonormal basis, the stability estimate becomes

$$|P_{\mathcal{X}}[f](x)| \leq \sqrt{\varphi(0)} \|f\|_{\Phi}$$

Theorem

Each Φ -orthonormal basis \mathcal{U} arise from a decomposition

$$A = B^T \cdot B,$$

with $B^T = V_{\mathcal{U}}$ and $B^{-1} = C_{\mathcal{U}}$.

Weighted SVD bases

The main idea for the construction of the new basis is **to discretize the "natural" basis** introduced in Mercer's theorem. To this aim, consider on Ω a cubature rule $(X, \mathcal{W})_N$, $N \in \mathbb{N}$, that is a set of distinct points $X = \{x_j\}_{j=1}^N$ such that

$$\int_{\Omega} f(y) dy \approx \sum_{j=1}^N f(x_j) w_j \quad \forall f \in \mathcal{N}_{\Phi}(\Omega)$$

Thus, the operator T_{Φ} can be approximate for each eigenvalue λ_j on X as

$$\lambda_j \varphi_j(x_i) = \int_{\Omega} \Phi(x_i, y) \varphi_j(y) dy \quad i = 1, \dots, N, \quad \forall j > 0,$$

and then discretized using the cubature rule by

$$\lambda_j \varphi_j(x_i) \approx \sum_{h=1}^N \Phi(x_i, x_h) \varphi_j(x_h) w_h \quad i, j = 1, \dots, N. \quad (1)$$

Weighted SVD bases

Setting $W = \text{diag}(w_j)$, it suffices to solve the following discrete eigenvalue problem in order to find the approximation of the eigenvalues and eigenfunctions (evaluated on X) of $T_\Phi[f]$:

$$\lambda v = (A \cdot W)v.$$

This approach involves a scaled version of the kernel matrix, that is $A \cdot W$, which is no longer symmetric and that cannot be described as a factorization of $A = B^T \cdot B$, in fact it doesn't lead directly to the connection between the discretization version of the "natural" basis of Mercer's theorem and a basis of the subspace $\mathcal{N}_\Phi(\Omega)$.

Weighted SVD bases

A solution is to rewrite (1), using the fact that the weights are positive, as

$$\lambda_j(\sqrt{w_i}\varphi_j(x_i)) = \sum_{h=1}^N (\sqrt{w_i}\Phi(x_i, x_h)\sqrt{w_h})(\sqrt{w_h}\varphi_j(x_h)) \quad i, j = 1, \dots, N.$$

that correspond to consider the **scaled eigenvalue problem**

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

which is equivalent to the previous one, now involving the symmetric and positive definite matrix $A_W := \sqrt{W} \cdot A \cdot \sqrt{W}$. *This matrix is normal, then a singular value decomposition of A_W is a unitary diagonalization.*

Weighted SVD bases

Definition (WSVD basis)

A weighted SVD basis \mathcal{U} is a basis for $\mathcal{N}_\Phi(X)$ characterized by the following matrices:

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

where

$$\sqrt{W} \cdot A \cdot \sqrt{W} = Q \cdot \Sigma^2 \cdot Q^T$$

is a singular value decomposition (and a unitary diagonalization) of the scaled kernel matrix A_W . Σ is a diagonal matrix with $\Sigma_{jj} = \sigma_j$, $j = 1, \dots, N$ and $\sigma_1^2 \geq \dots \geq \sigma_N^2 > 0$ are the singular values of A_W , and W is a diagonal matrix where $W_{jj} = w_j$, $j = 1, \dots, N$ are the weights of the cubature rule $(X, \mathcal{W})_N$.

Remark

It is essential to require that $\sum_{j=1}^N w_j = |\Omega|$, which is equivalent to ask that for all $N \in \mathbb{N}$ the cubature rule $(X, \mathcal{W})_N$ is exact at least for the constant functions. It is also possible to use a set of weights which does not provide a cubature rule, but in this way we lose the connection between \mathcal{U} and the eigenbasis $\{\varphi_j\}_{j>0}$, while remains unchanged the stability properties

Theorem

Every weighted SVD basis \mathcal{U} satisfies:

- ① $u_j(x) = \frac{1}{\sigma_j^2} \sum_{i=1}^N w_i u_j(x_i) \Phi(x, x_i) \approx \frac{1}{\sigma_j^2} T_\Phi[u_j](x), \quad \forall j = 1, \dots, N,$
 $\forall x \in \Omega;$
- ② \mathcal{U} is $\mathcal{N}_\Phi(\Omega)$ -orthonormal;
- ③ \mathcal{U} is $\ell_{2,w}(X)$ -orthogonal s.t. $\|u_j\|_{\ell_{2,w}(X)}^2 = \sigma_j^2 \quad \forall u_j \in \mathcal{U};$
- ④ $\sum_{j=1}^N \sigma_j^2 = \varphi(0) |\Omega|;$

Weighted SVD bases

Remark

Given a cubature rule $(X, W)_N$ we can define the $\ell_{2,w}(X)$ inner product as

$$(f, g)_{\ell_{2,w}(X)} = \sum_{i=1}^N w_i f(x_i) g(x_i) \quad f, g \in \mathcal{N}_\Phi(\Omega)$$

which is a discrete version of $(\cdot, \cdot)_{L_2(\Omega)}$, in fact

$$(f, g)_{L_2(\Omega)} = \int_{\Omega} f(x) g(x) dx \approx \sum_{j=1}^N w_j f(x_j) g(x_j) = (f, g)_{\ell_{2,w}(X)}.$$

Remark

Hence the operator $T_N : \mathcal{N}_\Phi(\Omega) \rightarrow \mathcal{N}_\Phi(X)$

$$T_N[f](x) = (f, \Phi(\cdot, x))_{\ell_{2,w}(X)}$$

which is the discrete version of T_Φ , is defined just by replacing the $L_2(\Omega)$ product with its discrete version $\ell_{2,w}(X)$ in T_Φ . The operator T_N has N eigenvectors, which turns out to be exactly the basis $\{u_j\}_{j=1}^N$. The functions u_j are essentially a discrete version of the $\{\varphi_j\}_{j>0}$. Moreover property (4) of the previous theorem suggest that

$$\sum_{j=1}^N w_j \Phi(x_j, x_j) = \text{tr}(\sqrt{W} \cdot A \cdot \sqrt{W}) = \sum_{j=1}^N \sigma_j^2 = \varphi(0) |\Omega| = \int_{\Omega} \Phi(x, x) dx$$

The integral is exactly approximated by the cubature rule (it is suppose to be exact at least for constant function)

Weighted SVD bases

Remark

The norm of the point-wise error operator, namely the Power function $\mathcal{P}_{\Phi, X}(x)$, can be expressed in the simpler form

$$\mathcal{P}_{\Phi, X}(x)^2 = \varphi(0) - \sum_{j=1}^N u_j(x)^2$$

which involves only the basis. Furthermore, it is possible to give an expansion of the kernel when it acts on functions of $\mathcal{N}_{\Phi}(X)$. That is

$$\Phi(x, y) = \sum_{j=1}^N u_j(x)u_j(y) \quad \forall x, y \in \Omega$$

This is useful when it is required to use a degenerate kernel to approximate the original one .

Weighted discrete least-squares approximation

The goal is to project the unknown function $f \in \mathcal{N}_\phi(\Omega)$ into a proper set $\mathcal{N}_\phi(X)$. In other words we will use a smaller basis $\mathcal{U}' \subsetneq \mathcal{U}$. This is done in order to obtain better result in terms of stability and computational cost, without serious loss of convergence speed.

Definition

Given a function $f \in \mathcal{N}_\phi(\Omega)$, a discrete subset $X \subset \Omega$, a set of cubature weights \mathcal{W} associated with X , a weighted SVD basis \mathcal{U} for $\mathcal{N}_\phi(X)$ and a natural number $M \leq N = |X|$, the weighted discrete least-squares approximation of order M of f is the function $\Lambda_M[f]$ that satisfied the condition

$$\Lambda_M[f] = \arg \min_{g \in \{u_1, \dots, u_M\}} \|f - g\|_{\ell_{2,w}(X)}$$

Weighted discrete least-squares approximation

Theorem (how to compute Λ_M)

The weighted discrete least-squares approximation of a function $f \in \mathcal{N}_\Phi(\Omega)$ is given by

$$\Lambda_M[f](x) = \sum_{j=1}^M (f, u_j)_\Phi u_j(x) = \sum_{j=1}^M \frac{(f, u_j)_{\ell_2, w(X)}}{\sigma_j^2} u_j(x)$$

that is $\Lambda_M[f]$ is nothing else but a truncation to the first M terms of $P_X[f]$

Weighted discrete least-squares approximation

Remark

To compute the weighted least-squares approximant $\Lambda_M[f]$ it suffices to use the first M elements of the interpolant $P_X[f]$, which correspond to the biggest singular values σ_j^2 , $j = 1, \dots, N$. This approach makes sense since in the singular values $\{\sigma_j^2\}_{1 \leq j \leq N}$ of A accumulate to zero very fast, being in particular a discrete approximation of the eigenvalues of the compact operator $T_\phi[f]$. The rate of decay of the singular values is faster if the kernel is smoother and if the shape parameter is smaller. Hence the truncated basis provides a good approximation and allows to deal with a smaller problem. From a linear algebra point of view, this corresponds to solve the (weighted) linear system associated to the interpolation problem using a **total least-square** method. This is opposite to the case of the standard basis of translates, where the choice of the elements of the basis to neglect corresponds to the choice of a restricted subset $Y \subset X$.

Stability and convergence for weighted discrete least-squares approximation

Concerning the convergence we can define the norm of the pointwise-evaluation operator related to the approximation operator, which is the equivalent to the power function, is given by

$$(\mathcal{P}_{\Phi, X}^M)^2 = \|\varepsilon_x^M\|_{\mathcal{N}_{\Phi}(\Omega)^*}^2 = \varphi(0) - \sum_{j=1}^M u_j(x)^2$$

It is clear that is a simple truncation of the power function. Concerning the convergence, using the standard error bound, we have

$$|\Lambda_M[f] - f(x)|^2 \leq \left(\varphi(0) - \sum_{j=1}^M u_j(x)^2 \right) \|f\|_{\Phi}^2 \quad \forall f \in \mathcal{N}_{\Phi}(\Omega)$$

Stability and convergence for weighted discrete least-squares approximation

Note that by replacing an exact interpolant with a weighted discrete least-squares approximant, we can obtain better result in terms of stability. In particular, the stability estimate can also be refined for the particular case of a weighted SVD-basis, in fact

$$|\Lambda_M[f](x)| \leq \left(\sum_{j=1}^M u_j(x)^2 \right)^{1/2} \|f\|_{\Phi} \leq \sqrt{\varphi(0)} \|f\|_{\Phi}, \quad \forall f \in \mathcal{N}_{\Phi}(\Omega)$$

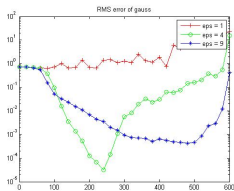
Hence, for convergent approximations, namely for approximations for which the power function decres to zero, namely we have necessary

$$\sum_{j=1}^M u_j(x)^2 \rightarrow \varphi(0).$$

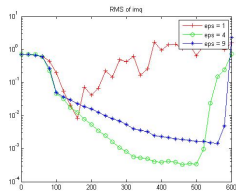
That is, the stability bound is maximized by $\varphi(0)$.

Comparison between the interpolant and the weighted least-squares approximant

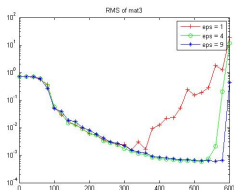
- We compare the approximation error produced using the full interpolant and some reduced weighted least-squares approximant starting from 600 trigonometric gaussian centers, and then truncating the basis for $M \in \{0, 20, \dots, 600\}$;
- We reconstruct the oscillatory function $f(x, y) = \cos(20(x + y))$ on the disk Ω with center $C = (\frac{1}{2}, \frac{1}{2})$ and radius $R = 1/2$;
- The experiment has been repeated for the gaussian kernel, the inverse multiquadric (IMQ) and the cubic matern (3MAT) for a shape parameter $\varepsilon = 1, 4, 9$
- We choose the trigonometric gaussian points as centers because they provide high-accuracy cubature rules while being sufficiently uniformly distributed in Ω ;



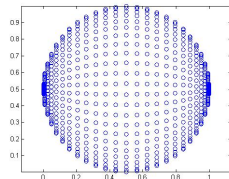
(a)



(b)



(c)



(d)

Figura: RMS errors for the reconstruction of f on Ω (d) using $\Lambda_M[f]$ for different values of M and different shape parameters, using gaussian kernel (a), the IMQ (b) and the 3MAT (c)

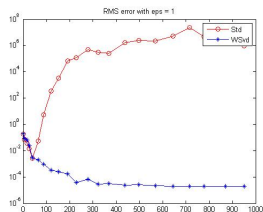
Comparison between the interpolant and the weighted least-squares approximant

- Depending on the choice of the shape parameter ε , the kernels present a fast decay to zero (Gaussian), a medium decay (IMQ) and a slow decay (3MAT);
- For the 3MAT kernel the interpolant remains stable, except for the last iterations, for $\varepsilon = 4, 9$, while for $\varepsilon = 1$ is more unstable;
- For the IMQ kernel the interpolant becomes unstable for $\varepsilon = 1$, while remain stable, except for the last iterations, for $\varepsilon = 4, 9$;
- For the Gaussian kernel the interpolant becomes unstable for $\varepsilon = 1, 4$, while remain stable, except for the last iterations, for $\varepsilon = 9$;
- The eigenvalues of the Gaussian kernel for $\varepsilon = 1$ are almost all under the machine precision; moreover the gaussian becomes too flat, and there is no hope to reconstruct an oscillatory function.

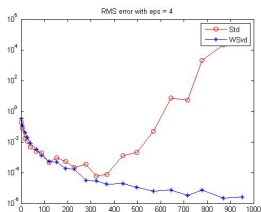
Comparison with the standard basis

- In this example we try to reconstruct the Franke's function with the IMQ-kernel on the lens Ω defined as the intersection of two disk with centers $C = (-\sqrt{2}/2, 0)$ and $c = (\sqrt{2}/2, 0)$ and radii $R = r = 1$;
- The test compares the results obtained with the interpolant based on the standard basis and the new basis, centred on an equally-spaced points and on a trigonometric-gaussian set of points, respectively and then is repeated for $\varepsilon = 1, 4, 9$;

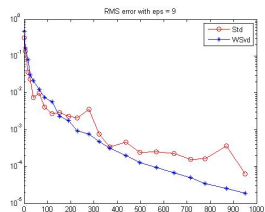
Comparison with the standard basis



(a)



(b)



(c)

Figura: RMS errors of the reconstruction of f on the lens Ω using the IMQ kernel with the standard basis and the new basis with shape parameter $\varepsilon = 1$ (a), $\varepsilon = 4$ (b), $\varepsilon = 9$ (c).

Comparison with the standard basis

- for $\varepsilon = 9$ there is only a small difference between the two basis;
- for $\varepsilon = 1, 4$, although for small data sets X the two basis does not behave so differently, when N becomes bigger the standar basis becomes unstable;
- the weigthed least-square approximant $\Lambda_M[f]$ based on the new basis presents a convergent behavior for each shape parameter, therefore there is no need to choose a particular ε to guarantee convergence, even if it slow;

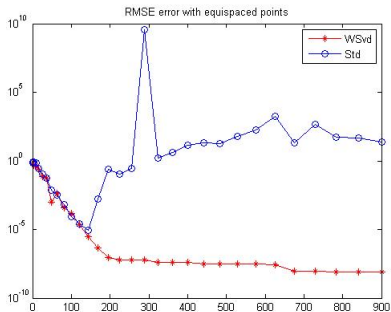
Comparison with a "stable" standard basis

- We want to reconstruct the function

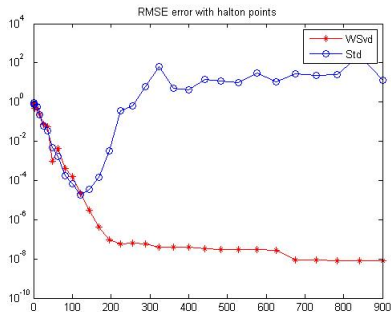
$$f(x) = -2\Phi_4(x, (0.5, 0.5)) + \Phi_4(x, (0, 0)) + 3\Phi_4(x, (0.7, 0.7))$$

where $\Phi_4(x, y)$ is the gaussian kernel with $\varepsilon = 4$ on the square $\Omega = [0, 1]^2$;

- To improve the stability of the standard basis we want to find the optimal ε^* which guarantees to minimize the residual error. We use the **leave-one-out cross validation strategy**;
- The RMS errors are plotted using equally-spaced points and Halton points as centers for the standard basis, while for the new basis, in both cases, we use the product Gauss-Legendre points.
- A good choice of the shape parameter reduces the instability of the standard interpolant, although it does not suffice to avoid it completely



(a)



(b)

Figura: RMS errors for the reconstruction of f on the square Ω using the gaussian kernel with the standard basis of translates with optimal shape parameter ε^* and the new basis with $\varepsilon = 4$. The standard interpolant is computed using equally spaced points (a) and Halton points (b).

The new basis

As seen before we can compute an orthonormal basis \mathcal{U} of $\mathcal{N}_\phi(X)$. Its main property is that it allows to extract a sub basis $\{u_1, \dots, u_M\}$, with $M < N$, which gives an approximant $\Lambda_M[f]$ as good as the interpolant, while being much more stable.

Problem: **The main problem is the efficiency of the basis computation.** Since the basis is found by a singular value decomposition, we have to compute all the elements u_j , $1 \leq j \leq N$, and then to leave out the last $N - M$ terms, the ones for which $\sigma_j^2 < \tau$, with τ a prescribed tolerance.

Objective: Find a way to slightly modify our basis in order to compute its most significant part using some tools from the theory of Krylov subspaces.

The Lanczos method and the approximation of the SVD

We want to find the solution of the linear system $Ax=b$ with A the kernel matrix and $b = (f(x_i))_{1 \leq i \leq N}$.

Let $\mathcal{K}_M(A, b) = \text{span} \{b, Ab, \dots, A^{M-1}b\}$ be the Krylov subspace of order M generated by A and b .

The Lanczos method computes an orthonormal basis $\{p_1, \dots, p_M\}$ of $\mathcal{K}_M(A, b)$ through a Gram-Schmidt orthonormalization.

The Lanczos method and the approximation of the SVD

In matrix form the algorithm is

$$AP_M = P_{M+1}\bar{H}_M, \quad \bar{H}_M = \begin{bmatrix} H_M \\ \bar{h}e_M^T \end{bmatrix},$$

where

- $P_M = [p_1, \dots, p_M] \in \mathbb{R}^{N \times M}$;
- H_M is a $M \times M$ tridiagonal matrix;
- $\bar{h} \in \mathbb{R}$;
- $e_M \in \mathbb{R}^M$ is the m -th unit vector.

The Lanczos method and the approximation of the SVD

The solution x can be approximated as

$$x = P_M y, \quad \text{where } y \in \mathbb{R}^M : \bar{H}_M y = \|b\|_2 e_1.$$

If A has a good low-rank approximation, we expect that a good approximation of x can be computed using M components with $M \ll N$.

The Lanczos method and the approximation of the SVD

Let's consider the singular value decomposition of

$$\bar{H}_M = U_M \bar{\Sigma}_M V_M^T, \quad \text{where } \bar{\Sigma} = \begin{bmatrix} \Sigma_M^2 \\ 0 \end{bmatrix},$$

where

- $U_M \in \mathbb{R}^{(M+1) \times (M+1)}$ unitary matrix;
- $V_M \in \mathbb{R}^{M \times M}$ unitary matrix;
- Σ_M^2 diagonal matrix having the singular values as its entries.

Remark

Since the last row of $\bar{\Sigma}$ is the zero vector, the decomposition does not change if we remove this row and the last column of U_M . From now on we will denote by U_M the matrix without the last column so that the decomposition becomes $\bar{H}_M = U_M \Sigma_M^2 V_M^T$.

Construction of the basis

Our idea is using the SVD of \bar{H}_M for constructing the new basis.

We can construct a new set of functions $\{\tilde{u}_j\}_{1 \leq j \leq M} \in \mathcal{N}_\Phi(X)$ similarly to the construction of the WSVD basis. We stress that the set $\{\tilde{u}_j\}$ does not span $\mathcal{N}_\Phi(X)$ unless $M = N$ when the SVD of \bar{H}_M equals to that of A .

Definition

Let A be the kernel matrix for the set X on N distinct points. Let \bar{H}_M , P_M , V_M , U_M , Σ_M^2 be as introduced as before. Hence, the basis $\bar{U} = \{\bar{u}_1, \dots, \bar{u}_M\}$ is characterized by the matrix of change of basis

$$C_{\bar{U}_M} = P_M V_M \Sigma_M^{-1}$$

or by the collocation matrix

$$V_{\bar{U}_M} = P_{M+1} U_M \Sigma_M$$

Construction of the basis

Remark

In this case the basis strongly depends on the particular function $f \in \mathcal{N}_\Phi(\Omega)$ used to construct the Krylov subspace. This dependence influences the behavior of the approximant.

Lemma

Let \tilde{U}_M be the square matrix obtained from U_M removing the last row u_M^T . Then \tilde{U}_M and V_M coincide except for the last row, namely only the m -th row d_M^T of the difference is a non zero row vector i.e.

$$\tilde{U}_M = V_M + e_M d_M^T$$

Theorem

Let the basis $\bar{\mathcal{U}}$ be defined as in definition before. Then

- i) the basis in $\ell_{2,w}(X)$ -orthogonal with $\|\bar{u}_j\|_{\ell_{2,w}}^2 = \sigma_j^2$;
- ii) the basis is near orthonormal on $\mathcal{N}_\Phi(\Omega)$, meaning that $(\bar{u}_i, \bar{u}_j)_\Phi = \delta_{ij} + r_{ij}^{(M)}$, where

$$R_M = (r_{ij}^{(M)})_{1 \leq i, j \leq M} = \Sigma_M^{-1} v_M d_M^T \Sigma_M$$

is a rank one matrix for $1 \leq M < N$, and $r_{ij}^{(M)} = 0$ when $M = N$;

- iii) when $M = N$, $\bar{\mathcal{U}} = \mathcal{U}$.

Remark

The magnitude of r_{ij} is close to the machine precision except if both i and j are close to M . That is to say that **the first elements of the basis are orthonormal from a numerical point of view.**

Approximation

Let's now compute the approximant. If we take the function $f \in \mathcal{N}_\Phi(\Omega)$ from which the basis is constructed, we get the approximant as a projection with respect to the $\ell_{2,w}(X)$ inner product

$$\Lambda_M[f](x) = \sum_{j=1}^M \sigma_j^{-2} (f, \bar{u}_j)_{\ell_{2,w}(X)} \bar{u}_j \quad \forall x \in \Omega. \quad (2)$$

But it can be expressed in terms of the $\mathcal{N}_\Phi(\Omega)$ inner product as in the WSVD basis:

Theorem

If the basis \bar{U}_M is constructed from $f \in \mathcal{N}_\Phi(\Omega)$, for all $j = 1, \dots, M$ we have

$$(f, \bar{u}_j)_{\ell_{2,w}(X)} = \sigma_j^2 (f, \bar{u}_j)_\Phi. \quad (3)$$

Hence the approximant $\Lambda_M[f](x)$ is a projection of $\mathcal{N}_\Phi(\Omega)$, or equivalently

$$\Lambda_M[f](x) = \sum_{j=1}^M (f, \bar{u}_j)_\Phi \bar{u}_j \quad \forall x \in \Omega. \quad (4)$$

Approximation

Remark

If we take an other function $\tilde{f} \in \mathcal{N}_\Phi(\Omega)$, the equality (3) holds with a residual term on the right hand side. This is due to the fact that in the case $\tilde{f} \neq f$ the terms depending on p_M, p_{M+1} are not cancelled.

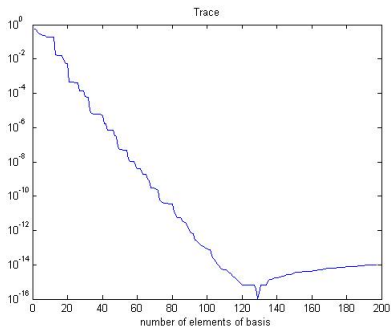
On the other hand, the equation (2) **depends only on the relation between the left inverse of $V_{\bar{U}_M}$, that is $\Sigma_M^{-2} V_{\bar{U}_M}$, and its transpose, not on the connection with $C_{\bar{U}}$** . This means that **we can compute the approximant of a function $\tilde{f} \neq f$ also using the basis constructed starting from f** .

A native space example

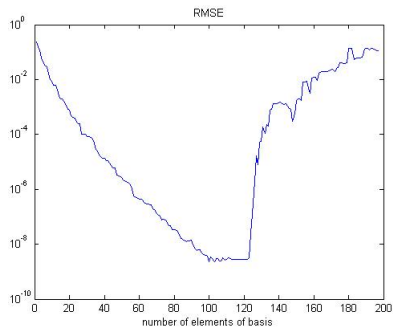
- Domain Ω is the unit circle
- We want to reconstruct the function
 $f(x) = \Phi_1(x, p_1) + 2 \Phi_1(x, p_2) - 2 \Phi_1(x, p_3) + 3 \Phi_1(x, p_4)$
 where Φ_1 is the gaussian kernel with $\varepsilon = 1$.
- We would like a stopping rule for the Lanczos iteration: a reasonably good choice is when, for a certain tolerance $\tau > 0$, we have

$$\left| \frac{1}{N} \sum_{j=1}^M (H_M)_{jj} - 1 \right| < \tau.$$

It seems good enough to control the iteration in the case the functions lies in the native space of the kernel.



(a)



(b)

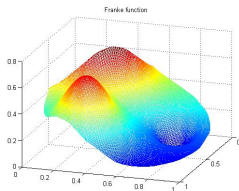
Figura: Decay of the residual (a) compared with the corresponding RMSE (b)

A general example

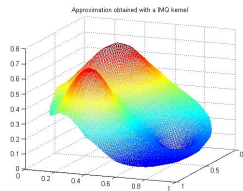
We want to approximate the Franke's function, we choose as kernel the inverse multiquadratic one (IMQ) with $\varepsilon = 2$ and as domain a lune, namely defined by the difference of 2 disk of radius 0.5 with centers in $(0,0)$ and $(0.5,0.5)$. To make the test general we use a set of randomly distributed data sites X with $n = 60^2$.

Remark

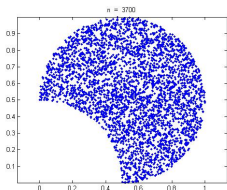
In this case the tolerance of the stopping rule have been set to $\tau = 10^{-10}$, since a smaller value lead to an increase of the RMS error. **The failure of the stopping rule in this case is due to the fact that the Franke's function does not belong to the native space of the kernel**



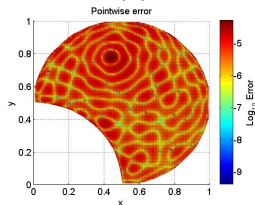
(a)



(b)



(c)



(d)

Figura: Franke's function (a); approximation obtained with a IMQ kernel with $\varepsilon = 2$ (b) using 60^2 randomly distributed data points (c); pointwise error computed on a grid of 60×60 equally spaced points (d)

THANK YOU FOR YOUR ATTENTION