Non-symmetric kernel-based greedy approximation

Gabriele Santin
Joint work with Bernard Haasdonk

Institute for Applied Analysis and Numerical Simulation

MAIA2016 - Luminy, 18-23.9.2016

University of Stuttgart
Germany

Setting: $\Omega \subset \mathbb{R}^d$ compact, points $Y_n := \{y_1, \dots, y_n\} \subset \Omega$, $f(y_1), \dots, f(y_n)$ samples.

- $K : \Omega \times \Omega \to \mathbb{R}$ continuous, SPD and sym. kernel (possibly RBF)
- $\mathcal{H}(\Omega)$ native space of K on Ω (K is the reproducing kernel)
- $s_n(f) := \sum_{j=1}^n \alpha_j K(\cdot, y_j)$ interpolant
- $V_n := V(Y_n) := \text{span}\{K(\cdot, y), y \in Y_n\}$
- Interpolation conditions $s_n(f)(y_i) = f(y_i)$
- Interpolation operator $f \mapsto s_n(f)$, from $\mathcal{H}(\Omega)$ to V_n

The vector of coefficients c exist and it is unique since

$$A = [K(y_i, y_j)]_{i,j=1}^n$$

is SPD.

The interpolation operator is the projection

$$\Pi_{V_n} : \mathcal{H}(\Omega) \to V_n$$

(best approximation in $\mathcal{H}(\Omega)$)

Setting: $\Omega \subset \mathbb{R}^d$ compact, points $Y_n := \{y_1, \dots, y_n\} \subset \Omega$, $f(y_1), \dots, f(y_n)$ samples.

- $K : \Omega \times \Omega \to \mathbb{R}$ continuous, SPD and sym. kernel (possibly RBF)
- $\mathcal{H}(\Omega)$ native space of K on Ω (K is the reproducing kernel)
- $s_n(f) := \sum_{j=1}^n \alpha_j K(\cdot, y_j)$ interpolant
- $V_n := V(Y_n) := \text{span}\{K(\cdot, y), y \in Y_n\}$
- Interpolation conditions $s_n(f)(y_i) = f(y_i)$
- Interpolation operator $f \mapsto s_n(f)$, from $\mathcal{H}(\Omega)$ to V_n

The vector of coefficients c exist and it is unique since

$$A = [K(y_i, y_j)]_{i,j=1}^n$$

is SPD.

The interpolation operator is the projection

$$\Pi_{V_n} : \mathcal{H}(\Omega) \to V_n$$

(best approximation in $\mathcal{H}(\Omega)$)

Setting: $\Omega \subset \mathbb{R}^d$ compact, points $Y_n := \{y_1, \dots, y_n\} \subset \Omega$, $f(y_1), \dots, f(y_n)$ samples.

- $K : \Omega \times \Omega \to \mathbb{R}$ continuous, SPD and sym. kernel (possibly RBF)
- $\mathcal{H}(\Omega)$ native space of K on Ω (K is the reproducing kernel)
- $s_n(f) := \sum_{j=1}^{n} \alpha_j K(\cdot, y_j)$ interpolant
- $V_n := V(Y_n) := \text{span}\{K(\cdot, y), y \in Y_n\}$
- Interpolation conditions $s_n(f)(y_i) = f(y_i)$
- Interpolation operator $f \mapsto s_n(f)$, from $\mathcal{H}(\Omega)$ to V_n

The vector of coefficients c exist and it is unique since

$$A = [K(y_i, y_j)]_{i,j=1}^{n}$$

is SPD.

The interpolation operator is the projection

$$\Pi_{V_n} : \mathcal{H}(\Omega) \to V_n$$

(best approximation in $\mathcal{H}(\Omega)$)

Setting: $\Omega \subset \mathbb{R}^d$ compact, points $Y_n := \{y_1, \ldots, y_n\} \subset \Omega$, $f(y_1), \ldots, f(y_n)$ samples.

- $K : \Omega \times \Omega \to \mathbb{R}$ continuous, SPD and sym. kernel (possibly RBF)
- $s_n(f) := \sum_{j=1}^{n} \alpha_j K(\cdot, y_j)$ interpolant
- Interpolation conditions $s_n(f)(y_i) = f(y_i)$
- $\mathcal{H}(\Omega)$ native space of K on Ω (K is the reproducing kernel)
- $V_n := V(Y_n) :=$ span$\{K(\cdot, y), y \in Y_n\}$
- Interpolation operator $f \mapsto s_n(f)$, from $\mathcal{H}(\Omega)$ to V_n

The vector of coefficients c exist and it is unique since

$$A = [K(y_i, y_j)]_{i,j=1}^{n}$$

is SPD.

The interpolation operator is the projection

$$\Pi_{V_n} : \mathcal{H}(\Omega) \to V_n$$

(best approximation in $\mathcal{H}(\Omega)$)

Setting: $\Omega \subset \mathbb{R}^d$ compact, points $Y_n := \{y_1, \dots, y_n\} \subset \Omega$, $f(y_1), \dots, f(y_n)$ samples.

- $K : \Omega \times \Omega \to \mathbb{R}$ continuous, SPD and sym. kernel (possibly RBF)
- $\mathcal{H}(\Omega)$ native space of K on Ω (K is the reproducing kernel)
- $s_n(f) := \sum_{j=1}^n \alpha_j K(\cdot, y_j)$ interpolant
- $V_n := V(Y_n) := \text{span}\{K(\cdot, y), y \in Y_n\}$
- Interpolation conditions $s_n(f)(y_i) = f(y_i)$
- Interpolation operator $f \mapsto s_n(f)$, from $\mathcal{H}(\Omega)$ to V_n

The vector of coefficients c exist and it is unique since

$$A = [K(y_i, y_j)]_{i,j=1}^n$$

is SPD.

The interpolation operator is the projection

$$\Pi_{V_n} : \mathcal{H}(\Omega) \to V_n$$

(best approximation in $\mathcal{H}(\Omega)$)

Structure

Nested point sets:	Y_1	$\subset$	Y_2	$\subset$	$\dots$	$\subset$	Y_N
Nested subspaces:	V_1	$\subset$	V_2	$\subset$	$\dots$	$\subset$	V_N
Projections:	f_1		f_2		$\dots$		f_N

Use of Newton basis $v_1, \dots, v_n$:⟵ **Müller-Sch09**

- nested o.n.b. of V_n
- easy construction via Gram-Schmidt over kernel basis
- coefficients from partial Cholesky decomposition of the kernel matrix

Data dependent point-selection rules:

- **[DeM-Sch-Wen05]** f-greedy: $y_i = \arg\max_{y \in Y_N \setminus Y_{i-1}} |f_{i-1}(y)|$
- **[Wirtz-Haas13]** f/P-greedy: $y_i = \arg\max_{y \in Y_N \setminus Y_{i-1}} |f_{i-1}(y)| \,/\, |P_{i-1}(y)|$

Structure

Nested point sets:	Y_1	$\subset$	Y_2	$\subset$	$\dots$	$\subset$	Y_N
Nested subspaces:	V_1	$\subset$	V_2	$\subset$	$\dots$	$\subset$	V_N
Projections:	f_1		f_2		$\dots$		f_N

Use of Newton basis $v_1, \dots, v_n$:⟵ **Müller-Sch09**

- nested o.n.b. of V_n
- easy construction via Gram-Schmidt over kernel basis
- coefficients from partial Cholesky decomposition of the kernel matrix

Data dependent point-selection rules:

- **[DeM-Sch-Wen05]** f-greedy: $y_i = \arg\max_{y \in Y_N \setminus Y_{i-1}} |f_{i-1}(y)|$
- **[Wirtz-Haas13]** f/P-greedy: $y_i = \arg\max_{y \in Y_N \setminus Y_{i-1}} |f_{i-1}(y)| / |P_{i-1}(y)|$

Structure

Nested point sets:	Y_1	$\subset$	Y_2	$\subset$	$\dots$	$\subset$	Y_N
Nested subspaces:	V_1	$\subset$	V_2	$\subset$	$\dots$	$\subset$	V_N
Projections:	f_1		f_2		$\dots$		f_N

Use of Newton basis $v_1, \dots, v_n$:⟵ **Müller-Sch09**

- nested o.n.b. of V_n
- easy construction via Gram-Schmidt over kernel basis
- coefficients from partial Cholesky decomposition of the kernel matrix

Data dependent point-selection rules:

- **[DeM-Sch-Wen05]** f-greedy: $y_i = \arg\max_{y \in Y_N \setminus Y_{i-1}} |f_{i-1}(y)|$
- **[Wirtz-Haas13]** f/P-greedy: $y_i = \arg\max_{y \in Y_N \setminus Y_{i-1}} |f_{i-1}(y)| \,/\, |P_{i-1}(y)|$

Notation, motivations and setting

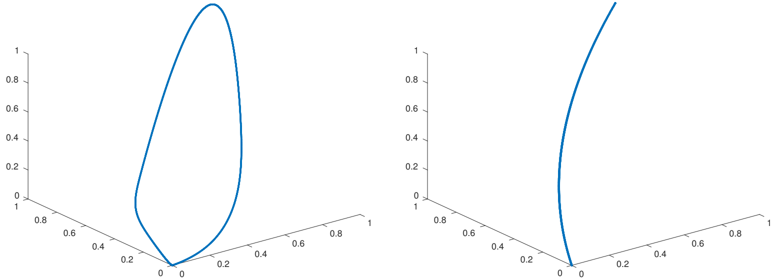
Application example

Matrix-valued kernel for $f : \Omega \to \mathbb{R}^q$, $q \geq 1$ ([4], [8])

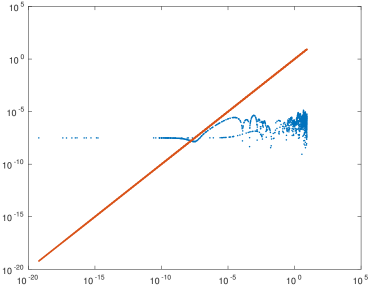
- $\mathcal{H}_K(\Omega)^q := \{f : \Omega \to \mathbb{R}^q, f_j \in \mathcal{H}(\Omega)\}$
- $(f, g)_q := \sum_{j=1}^{q} (f_j, g_j)$
- Use the same subspace V_n over components

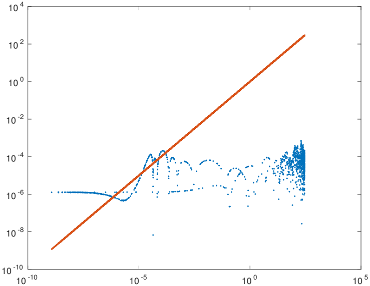
Blood-flow simulation in vascular networks
(with Tobias Köppl, Dep. Hydromechanics and Modelling of Hydrosystems)

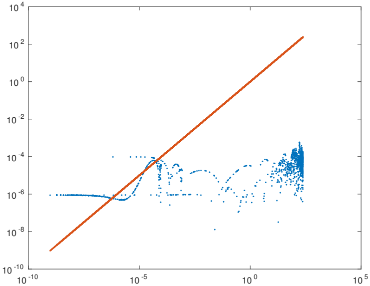
- 1D - transport eqs.
- Newton solver for bifurcations
- Sampling of characteristic curves
- Surrogate kernel model of the bifurcations
- $\mathbb{R}^3 \to \mathbb{R}^3$ model (possibly $\mathbb{R}^q \to \mathbb{R}^q$)



Motivation I







- $M = 12742$ samples
- $n = 70$ selected centers
- issues for small data

Outline

1. General definition of non-symmetric problem.
2. Some preliminary theoretical result.
3. Construction of approximants.
4. Construction of greedy approximants.
5. Numerical results.

Characterization

Setting

- Functionals $\{\lambda_i\}_{i=1}^M, \{\mu_j\}_{j=1}^N \subset \mathcal{H}(\Omega)^*$ ($M \leq N$, each set is lin. indep.)
- Test and trial spaces

$$V_M := \{\lambda_i^y K(\cdot, y), i = 1, \dots, M\}, \quad U_N := \{\mu_j^x K(x, \cdot), j = 1, \dots, N\},$$

- Kernel matrix $(A_{MN})_{ij} := (\lambda_i^y K(\cdot, y), \mu_j^x K(x, \cdot))_{\mathcal{H}(\Omega)^*} = \lambda_i^y \mu_j^x K(x, y)$

Goals (Given $f \in \mathcal{H}(\Omega)$ and $n \leq M$)

- determine subspaces $V_n \subset V_M$, $U_n \subset U_N$, such that

$$s_n(f) := \sum_{j=1}^n \alpha_j u_j, \quad (s_n(f), v_i) = (f, v_i), \ 1 \leq i \leq n$$

(bases $\{u_j\}_{j=1}^n$ of U_n and $\{v_i\}_{i=1}^n$ of V_n)
- Solve

$$A_n \alpha = b$$

$A_n \in \mathbb{R}^{n \times n}$, $(A_n)_{ij} := (v_i, u_j)$ $b \in \mathbb{R}^n$, $b_i := (f, v_i)$ vector of data.

Examples: non-sym interp., non-sym coll. ... (other functs? multiscale interp.?)

- Functionals $\{\lambda_i\}_{i=1}^M, \{\mu_j\}_{j=1}^N \subset \mathcal{H}(\Omega)^*$ ($M \leq N$, each set is lin. indep.)
- Test and trial spaces

$$V_M := \{\lambda_i^y K(\cdot, y), i = 1, \dots, M\}, \quad U_N := \{\mu_j^x K(x, \cdot), j = 1, \dots, N\},$$

- Kernel matrix $(A_{MN})_{ij} := (\lambda_i^y K(\cdot, y), \mu_j^x K(x, \cdot))_{\mathcal{H}(\Omega)^*} = \lambda_i^y \mu_j^x K(x, y)$

Goals (Given $f \in \mathcal{H}(\Omega)$ and $n \leq M$)

- determine subspaces $V_n \subset V_M$, $U_n \subset U_N$, such that

$$s_n(f) := \sum_{j=1}^n \alpha_j u_j, \quad (s_n(f), v_i) = (f, v_i), \ 1 \leq i \leq n$$

(bases $\{u_j\}_{j=1}^n$ of U_n and $\{v_i\}_{i=1}^n$ of V_n)
- Solve

$$A_n \alpha = b$$

$A_n \in \mathbb{R}^{n \times n}$, $(A_n)_{ij} := (v_i, u_j)$ $b \in \mathbb{R}^n$, $b_i := (f, v_i)$ vector of data.

Examples: non-sym interp., non-sym coll. ... (other functs? multiscale interp.?)

Characterization

Setting

- Functionals $\{\lambda_i\}_{i=1}^M, \{\mu_j\}_{j=1}^N \subset \mathcal{H}(\Omega)^*$ ($M \le N$, each set is lin. indep.)
- Test and trial spaces

$$V_M := \{\lambda_i^y K(\cdot, y), i = 1, \dots, M\}, \quad U_N := \{\mu_j^x K(x, \cdot), j = 1, \dots, N\},$$

- Kernel matrix $(A_{MN})_{ij} := (\lambda_i^y K(\cdot, y), \mu_j^x K(x, \cdot))_{\mathcal{H}(\Omega)^*} = \lambda_i^y \mu_j^x K(x, y)$

Goals (Given $f \in \mathcal{H}(\Omega)$ and $n \le M$)

- determine subspaces $V_n \subset V_M$, $U_n \subset U_N$, such that

$$s_n(f) := \sum_{j=1}^n \alpha_j u_j, \quad (s_n(f), v_i) = (f, v_i),\ 1 \le i \le n$$

(bases $\{u_j\}_{j=1}^n$ of U_n and $\{v_i\}_{i=1}^n$ of V_n)

- Solve

$$A_n \alpha = b$$

$A_n \in \mathbb{R}^{n \times n}$, $(A_n)_{ij} := (v_i, u_j)$ $b \in \mathbb{R}^n$, $b_i := (f, v_i)$ vector of data.

Examples: non-sym interp., non-sym coll. ... (other functs? multiscale interp.?)

Principal angles

Let $\gamma_1, \ldots, \gamma_N$ be the cosines of the principal angles between the V_M, U_N. For $M + 1 \leq n \leq N$, we set $\gamma_n := 0$. Denote by $\{\psi_i\}_i$, $\{\varphi_j\}_j$ a pair of principal bases .

Characterization and existence

Given $n \leq M$, there exist subspaces $V_n \subset V_M$, $U_n \subset U_N$ such that the approximation problem with test space V_n and trial space U_n has a unique solution, if and only if $\gamma_n > 0$.

In this case, given any pair of such subspaces, the operator $s_n : \mathcal{H}(\Omega) \to U_n$ is a projection with $\ker(s_n) = V_n^\perp$, and norm

$$\|s_n\| := \sup_{f \in \mathcal{H}(\Omega)} \frac{\|s_n(f)\|}{\|f\|} = \frac{1}{\gamma_n(V_n, U_n)}.$$

In particular, it is an orthogonal projection if and only if $V_n = U_n$. Moreover, there exist such subspaces that are spanned by n elements of the original bases of V_M and U_N.

Principal angles

Let $\gamma_1, \dots, \gamma_N$ be the cosines of the principal angles between the V_M, U_N. For $M+1 \leq n \leq N$, we set $\gamma_n := 0$. Denote by $\{\psi_i\}_i$, $\{\varphi_j\}_j$ a pair of principal bases .

Characterization and existence

Given $n \leq M$, there exist subspaces $V_n \subset V_M$, $U_n \subset U_N$ such that the approximation problem with test space V_n and trial space U_n has a unique solution, if and only if $\gamma_n > 0$.
In this case, given any pair of such subspaces, the operator $s_n : \mathcal{H}(\Omega) \to U_n$ is a projection with $\ker(s_n) = V_n^\perp$, and norm

$$\|s_n\| := \sup_{f \in \mathcal{H}(\Omega)} \frac{\|s_n(f)\|}{\|f\|} = \frac{1}{\gamma_n(V_n, U_n)}.$$

In particular, it is an orthogonal projection if and only if $V_n = U_n$. Moreover, there exist such subspaces that are spanned by n elements of the original bases of V_M and U_N.

Def: Power Function

$$P_n(\lambda) := \sup_{f \in \mathcal{H}(\Omega),\ \|f\| \le 1} \lambda(f - s_n(f)), \quad \lambda \in \mathcal{H}(\Omega)^*$$

Power Function use:

$$|\lambda(f - s_n)| \le P_n(\lambda)\|f\| \text{ for all } \lambda \in \mathcal{H}(\Omega)^* \quad \text{and } P_n(\lambda) = 0 \text{ iff } \lambda \in V_n^*.$$

Error comparison (for $f \in \mathcal{H}(\Omega)$, use principal bases for U_n, V_n):

$$\|f - \Pi_{V_n}(f)\|^2 = \sum_{k=1}^{n} \frac{1}{1-\gamma_k^2}\left((f,\varphi_k) - \gamma_k(f,\psi_k)\right)^2 + \|\Pi_{V_n^\perp}(f)\|^2$$

$$\|f - s_n(f)\|^2 = \sum_{k=1}^{n} \frac{1}{1-\gamma_k^2}\left((f,\varphi_k) - \frac{1}{\gamma_k}(f,\psi_k)\right)^2 + \|\Pi_{V_n^\perp}(f)\|^2.$$

In particular, $\|f - s_n(f)\| \le \|f - \Pi_{V_n}(f)\|$ if

$$0 \le |(f,\psi_k)| \le 2\left(\frac{\gamma_k}{1+\gamma_k^2}\right)|(f,\varphi_k)|, \ 1 \le k \le n.$$

Def: Power Function

$$P_n(\lambda) := \sup_{f\in\mathcal{H}(\Omega),\ \|f\|\le 1} \lambda(f - s_n(f)), \quad \lambda \in \mathcal{H}(\Omega)^*$$

Power Function use:

$|\lambda(f - s_n)| \le P_n(\lambda)\|f\|$ for all $\lambda \in \mathcal{H}(\Omega)^*$ and $P_n(\lambda) = 0$ iff $\lambda \in V_n^*$.

Error comparison (for $f \in \mathcal{H}(\Omega)$, use principal bases for U_n, V_n):

$$\|f - \Pi_{V_n}(f)\|^2 = \sum_{k=1}^{n} \frac{1}{1-\gamma_k^2} ((f,\varphi_k) - \gamma_k(f,\psi_k))^2 + \|\Pi_{V_n^\perp}(f)\|^2$$

$$\|f - s_n(f)\|^2 = \sum_{k=1}^{n} \frac{1}{1-\gamma_k^2} \left((f,\varphi_k) - \frac{1}{\gamma_k}(f,\psi_k)\right)^2 + \|\Pi_{V_n^\perp}(f)\|^2.$$

In particular, $\|f - s_n(f)\| \le \|f - \Pi_{V_n}(f)\|$ if

$$0 \le |(f,\psi_k)| \le 2\left(\frac{\gamma_k}{1+\gamma_k^2}\right)|(f,\varphi_k)|, \quad 1 \le k \le n.$$

Def: Power Function

$$P_n(\lambda) := \sup_{f\in\mathcal{H}(\Omega),\ \|f\|\le 1} \lambda\left(f - s_n(f)\right), \quad \lambda \in \mathcal{H}(\Omega)^*$$

Power Function use:

$$|\lambda(f - s_n)| \le P_n(\lambda)\|f\| \text{ for all } \lambda \in \mathcal{H}(\Omega)^* \quad \text{and } P_n(\lambda) = 0 \text{ iff } \lambda \in V_n^*.$$

Error comparison (for $f \in \mathcal{H}(\Omega)$, use principal bases for U_n, V_n):

$$\|f - \Pi_{V_n}(f)\|^2 = \sum_{k=1}^{n} \frac{1}{1-\gamma_k^2}\left((f,\varphi_k) - \gamma_k(f,\psi_k)\right)^2 + \|\Pi_{V_n^\perp}(f)\|^2$$

$$\|f - s_n(f)\|^2 = \sum_{k=1}^{n} \frac{1}{1-\gamma_k^2}\left((f,\varphi_k) - \frac{1}{\gamma_k}(f,\psi_k)\right)^2 + \|\Pi_{V_n^\perp}(f)\|^2.$$

In particular, $\|f - s_n(f)\| \le \|f - \Pi_{V_n}(f)\|$ if

$$0 \le |(f,\psi_k)| \le 2\left(\frac{\gamma_k}{1+\gamma_k^2}\right)|(f,\varphi_k)|, \ 1 \le k \le n.$$

Now fixed V_n , U_n , A_n.

$$v_i := \sum_{k=1}^{n} c_{ki} \lambda_k^y K(y, \cdot), \quad u_j := \sum_{h=1}^{n} d_{hj} \mu_h^x K(\cdot, x).$$

$$(C_v)_{ij} := c_{ij}, \; (C_u)_{ij} := d_{ij}$$

$$(V_v)_{ij} := (\mu_i^x K(x, \cdot), v_j) = \mu_i(v_j), \quad (V_u)_{ij} := (\lambda_i^y K(\cdot, y), u_j) = \lambda_i(u_j).$$

Bases from matrix factorization

Any basis $\{u_j\}_{j=1}^n$ of U_n is uniquely defined by the matrix of coefficients C_u or by the matrix of evaluations V_u , and in particular

$$A_n = V_u \, C_u^{-1}. \tag{1}$$

The same holds for any basis $\{v_j\}_{j=1}^n$ of V_n, with instead $A_n^T = V_v \, C_v^{-1}$.

Now fixed V_n , U_n , A_n.

$$v_i := \sum_{k=1}^{n} c_{ki} \lambda_k^y K(y, \cdot), \quad u_j := \sum_{h=1}^{n} d_{hj} \mu_h^x K(\cdot, x).$$

$$(C_v)_{ij} := c_{ij}, \ (C_u)_{ij} := d_{ij}$$

$$(V_v)_{ij} := (\mu_i^x K(x, \cdot), v_j) = \mu_i(v_j), \quad (V_u)_{ij} := (\lambda_i^y K(\cdot, y), u_j) = \lambda_i(u_j).$$

Bases from matrix factorization

Any basis $\{u_j\}_{j=1}^n$ of U_n is uniquely defined by the matrix of coefficients C_u or by the matrix of evaluations V_u , and in particular

$$A_n = V_u \, C_u^{-1}. \tag{1}$$

The same holds for any basis $\{v_j\}_{j=1}^n$ of V_n, with instead $A_n^T = V_v \, C_v^{-1}$.

Def: biorthonormal bases

Bases $\{v_j\}_{j=1}^n$ of V_n and $\{u_j\}_{j=1}^n$ of U_n are biorthonormal (or dual) if

$$(v_i, u_j) = \delta_{ij}, \quad 1 \leq i, j \leq n.$$

Bi-o.n. bases from matrix factorization

Bases $\{v_j\}_{j=1}^n$ and $\{u_j\}_{j=1}^n$ are biorthonormal if and only if

$$A_n = C_v^{-T} C_u^{-1}. \tag{2}$$

Moreover,

$$C_u = V_v^{-T}, \quad C_v = V_u^{-T}. \tag{3}$$

and in particular for any basis of U_n there exists a unique dual basis of V_n.

Def: biorthonormal bases

Bases $\{v_j\}_{j=1}^n$ of V_n and $\{u_j\}_{j=1}^n$ of U_n are biorthonormal (or dual) if

$$(v_i, u_j) = \delta_{ij}, \quad 1 \leq i, j \leq n.$$

Bi-o.n. bases from matrix factorization

Bases $\{v_j\}_{j=1}^n$ and $\{u_j\}_{j=1}^n$ are biorthonormal if and only if

$$A_n = C_v^{-T} C_u^{-1}. \tag{2}$$

Moreover,

$$C_u = V_v^{-T}, \quad C_v = V_u^{-T}. \tag{3}$$

and in particular for any basis of U_n there exists a unique dual basis of V_n.

Approximation and Power Function

Given a basis $\{u_j\}_{j=1}^n$ of U_n, and $f \in \mathcal{H}(\Omega)$

$$s_n(f) = \sum_{j=1}^{n} (f, v_j) u_j,$$

where $\{v_j\}_{j=1}^n$ is the unique basis of V_n which is dual to $\{u_j\}_{j=1}^n$.
For any $\lambda \in \mathcal{H}(\Omega)^*$,

$$P_n(\lambda)^2 = \lambda^x \lambda^y K(x,y) - \sum_{j=1}^{n} \lambda(u_j) \lambda(2v_j - s_n(v_j)).$$

- Incrementally selected indexes

$$I_n := \{i_1, \dots, i_n\} \subset I_M := \{1, \dots, M\},\ J_n := \{j_1, \dots, j_n\} \subset J_N := \{1, \dots, N\}$$

- Basis elements

$$\lambda^y_{i_1} K(\cdot, y), \dots, \lambda^y_{i_n} K(\cdot, y) \text{ and } \mu^x_{j_1} K(\cdot, x), \dots, \mu^x_{j_n} K(\cdot, x)$$

- Nested subspaces

$$V_1 \subset V_2 \subset \cdots \subset V_n \subset V_M \text{ and } U_1 \subset U_2 \subset \cdots \subset U_n \subset U_N,$$

Goals:

- Nested, bi-o.n.bases

$$U_{k-1} = \text{span}\{u_1, \dots, u_{k-1}\},\ \ U_k = \text{span}\{u_1, \dots, u_k\}$$

and similarly for V_n.

- Check solvability in a cheap way ($\gamma_n > 0$ is not enough)

- Incrementally selected indexes

$$I_n := \{i_1, \dots, i_n\} \subset I_M := \{1, \dots, M\}, \; J_n := \{j_1, \dots, j_n\} \subset J_N := \{1, \dots, N\}$$

- Basis elements

$$\lambda^y_{i_1} K(\cdot, y), \dots, \lambda^y_{i_n} K(\cdot, y) \;\text{ and }\; \mu^x_{j_1} K(\cdot, x), \dots, \mu^x_{j_n} K(\cdot, x)$$

- Nested subspaces

$$V_1 \subset V_2 \subset \cdots \subset V_n \subset V_M \;\text{ and }\; U_1 \subset U_2 \subset \cdots \subset U_n \subset U_N,$$

Goals:

- Nested, bi-o.n.bases

$$U_{k-1} = \text{ span } \{u_1, \dots, u_{k-1}\}, \;\; U_k = \text{ span } \{u_1, \dots, u_k\}$$

  and similarly for V_n.
- Check solvability in a cheap way ($\gamma_n > 0$ is not enough)

Generalized Newton Bases

A basis $\{u_j\}_j$ of U_n is nested if and only if C_u is upper triangular. A basis $\{u_j\}_j$ and its dual basis $\{v_i\}_i$ are both nested bases if and only if

$$C_v = L^{-T}, \quad C_u = U^{-1},$$

or equivalently

$$V_v = U^T, \quad V_u = L,$$

where L, U are lower- and upper-triangular matrices, without constraints on the diagonals and such that

$$A_n = L\,U$$

is a LU-decomposition of the matrix A_n, which corresponds to a partial LU-decomposition of the full matrix A_{MN} after rearrangement of columns and rows.

Use rediduals

$$r_0 := f, \quad r_{n-1} = (r_{n-1}, v_n)u_n + r_n, \ n \geq 1.$$

Incremental approximation with gen. Newton bases

We have $r_n \in V_n^\perp$ and

$$s_n(f) = \sum_{k=1}^{n} (r_{k-1}, v_n)u_n = \sum_{k=1}^{n} \lambda_{i_n}(r_{k-1})u_n,$$

and in particular

$$s_n(f) = s_{n-1}(f) + (r_{n-1}, v_n)u_n.$$

For the Power Function:

$$P_n(\lambda)^2 - P_{n-1}(\lambda)^2 = \|v_n\|^2 \lambda(u_n)^2 - 2\lambda(u_n)\lambda\left(v_n - s_{n-1}(v_n)\right).$$

Use rediduals

$$r_0 := f, \quad r_{n-1} = (r_{n-1}, v_n)u_n + r_n, \; n \geq 1.$$

Incremental approximation with gen. Newton bases

We have $r_n \in V_n^\perp$ and

$$s_n(f) = \sum_{k=1}^{n} (r_{k-1}, v_n)u_n = \sum_{k=1}^{n} \lambda_{i_n}(r_{k-1})u_n,$$

and in particular

$$s_n(f) = s_{n-1}(f) + (r_{n-1}, v_n)u_n.$$

For the Power Function:

$$P_n(\lambda)^2 - P_{n-1}(\lambda)^2 = \|v_n\|^2 \lambda(u_n)^2 - 2\lambda(u_n)\lambda\left(v_n - s_{n-1}(v_n)\right).$$

Use rediduals

$$r_0 := f, \quad r_{n-1} = (r_{n-1}, v_n) u_n + r_n, \ n \geq 1.$$

Incremental approximation with gen. Newton bases

We have $r_n \in V_n^{\perp}$ and

$$s_n(f) = \sum_{k=1}^{n} (r_{k-1}, v_n) u_n = \sum_{k=1}^{n} \lambda_{i_n}(r_{k-1}) u_n,$$

and in particular

$$s_n(f) = s_{n-1}(f) + (r_{n-1}, v_n) u_n.$$

For the Power Function:

$$P_n(\lambda)^2 - P_{n-1}(\lambda)^2 = \|v_n\|^2 \lambda(u_n)^2 - 2\lambda(u_n)\lambda\left(v_n - s_{n-1}(v_n)\right).$$

Matrix free and cheap rank check :

- Compute v_n:

$$v_n := \lambda^y_{i_n} K(\cdot, y) - \sum_{k=1}^{n-1} (\lambda^y_{i_n} K(\cdot, y), u_k) v_k = \lambda^y_{i_n} K(\cdot, y) - \sum_{k=1}^{n-1} \lambda_{i_n}(u_k) v_k$$

- Candidate u_n:

$$\tilde{u}_n := \mu^x_{j_n} K(\cdot, x) - \sum_{k=1}^{n-1} (\mu^x_{j_n} K(\cdot, x), v_k) u_k = \mu^x_{j_n} K(\cdot, x) - \sum_{k=1}^{n-1} \mu_{j_n}(v_k) u_k$$

- Check solvability:

$$(v_n, \tilde{u}_n) = \lambda^y_{i_n} \mu^x_{j_n} K(y, x) - \sum_{k=1}^{n-1} \mu_{j_n}(v_k) \lambda_{i_n}(u_k) = \mu_{j_n}(v_n)$$

- If so, normalize:

$$u_n := \tilde{u}_n / (v_n, \tilde{u}_n)$$

Matrix free and cheap rank check :

- Compute v_n:

$$v_n := \lambda^y_{i_n} K(\cdot, y) - \sum_{k=1}^{n-1} (\lambda^y_{i_n} K(\cdot, y), u_k) v_k = \lambda^y_{i_n} K(\cdot, y) - \sum_{k=1}^{n-1} \lambda_{i_n}(u_k) v_k$$

- Candidate u_n:

$$\tilde{u}_n := \mu^x_{j_n} K(\cdot, x) - \sum_{k=1}^{n-1} (\mu^x_{j_n} K(\cdot, x), v_k) u_k = \mu^x_{j_n} K(\cdot, x) - \sum_{k=1}^{n-1} \mu_{j_n}(v_k) u_k$$

- Check solvability:

$$(v_n, \tilde{u}_n) = \lambda^y_{i_n} \mu^x_{j_n} K(y, x) - \sum_{k=1}^{n-1} \mu_{j_n}(v_k) \lambda_{i_n}(u_k) = \mu_{j_n}(v_n)$$

- If so, normalize:

$$u_n := \tilde{u}_n / (v_n, \tilde{u}_n)$$

Matrix free and cheap rank check :

- Compute v_n:

$$v_n := \lambda^y_{i_n} K(\cdot, y) - \sum_{k=1}^{n-1} (\lambda^y_{i_n} K(\cdot, y), u_k) v_k = \lambda^y_{i_n} K(\cdot, y) - \sum_{k=1}^{n-1} \lambda_{i_n}(u_k) v_k$$

- Candidate u_n:

$$\tilde{u}_n := \mu^x_{j_n} K(\cdot, x) - \sum_{k=1}^{n-1} (\mu^x_{j_n} K(\cdot, x), v_k) u_k = \mu^x_{j_n} K(\cdot, x) - \sum_{k=1}^{n-1} \mu_{j_n}(v_k) u_k$$

- Check solvability:

$$(v_n, \tilde{u}_n) = \lambda^y_{i_n} \mu^x_{j_n} K(y, x) - \sum_{k=1}^{n-1} \mu_{j_n}(v_k) \lambda_{i_n}(u_k) = \mu_{j_n}(v_n)$$

- If so, normalize:

$$u_n := \tilde{u}_n / (v_n, \tilde{u}_n)$$

Matrix free and cheap rank check :

- Compute v_n:

$$v_n := \lambda^y_{i_n} K(\cdot, y) - \sum_{k=1}^{n-1} (\lambda^y_{i_n} K(\cdot, y), u_k) v_k = \lambda^y_{i_n} K(\cdot, y) - \sum_{k=1}^{n-1} \lambda_{i_n}(u_k) v_k$$

- Candidate u_n:

$$\tilde{u}_n := \mu^x_{j_n} K(\cdot, x) - \sum_{k=1}^{n-1} (\mu^x_{j_n} K(\cdot, x), v_k) u_k = \mu^x_{j_n} K(\cdot, x) - \sum_{k=1}^{n-1} \mu_{j_n}(v_k) u_k$$

- Check solvability:

$$(v_n, \tilde{u}_n) = \lambda^y_{i_n} \mu^x_{j_n} K(y, x) - \sum_{k=1}^{n-1} \mu_{j_n}(v_k) \lambda_{i_n}(u_k) = \mu_{j_n}(v_n)$$

- If so, normalize:

$$u_n := \tilde{u}_n / (v_n, \tilde{u}_n)$$

- Any LU-piv. rule is valid
- f - greedy with rank maximization
  (minimize residual, maximize spaces extension)

$$
\begin{aligned}
i_n &:= \arg\max_{i \in I_M \setminus I_{n-1}} |\lambda_i(r_{n-1})| \\
j_n &:= \arg\max_{j \in J_N \setminus J_{n-1}} |\mu_j(v_n)|
\end{aligned}
$$

- f - greedy with dual stabilization
  (minimize residual, minimize dual residual with coefficients as data)

$$
\begin{cases}
V_u c = b \\
V_v d = -V_v^T c
\end{cases}
$$

(comes from [3] - regularized solution of under-determined system)

- Any LU-piv. rule is valid
- f - greedy with rank maximization
  (minimize residual, maximize spaces extension)

$$\begin{aligned} i_n &:= \underset{i \in I_M \setminus I_{n-1}}{\arg\max} |\lambda_i(r_{n-1})| \\ j_n &:= \underset{j \in J_N \setminus J_{n-1}}{\arg\max} |\mu_j(v_n)| \end{aligned}$$

- f - greedy with dual stabilization
  (minimize residual, minimize dual residual with coefficients as data)

$$\begin{cases} V_u c &= b \\ V_v d &= -V_v^T c \end{cases}$$

(comes from [3] - regularized solution of under-determined system)

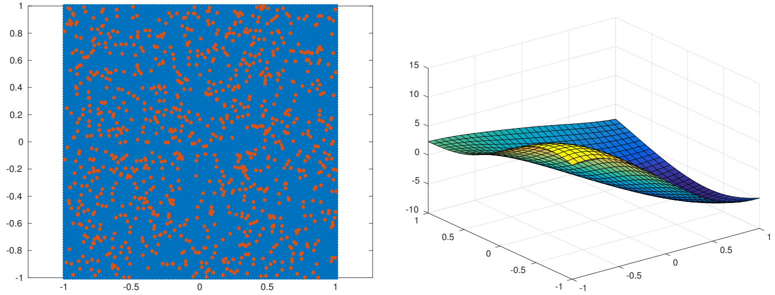


Figure: Gaussian Kernel with $\epsilon = 1$, $f \in \mathcal{H}(\Omega)$, $\Omega = [-1, 1]^2$.

Numerical experiments

Example I

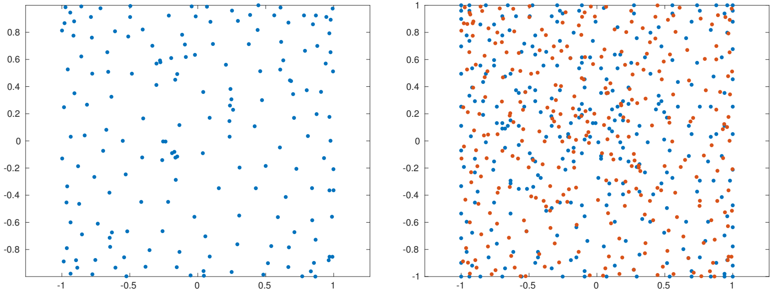


Figure: Gaussian Kernel with $\epsilon = 1$, $f \in \mathcal{H}(\Omega)$, $\Omega = [-1, 1]^2$.

Numerical experiments

Example I

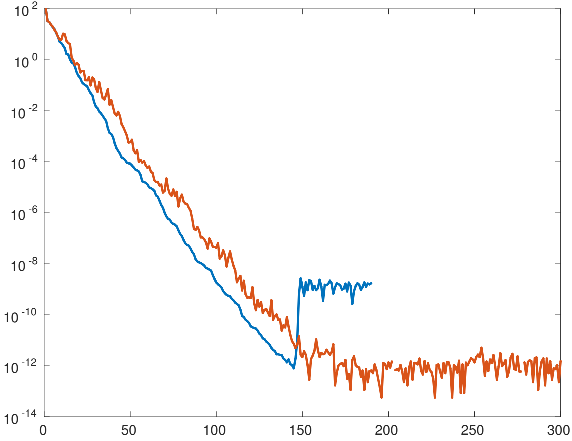


Figure: Gaussian Kernel with $\epsilon = 1$, $f \in \mathcal{H}(\Omega)$, $\Omega = [-1, 1]^2$.

Numerical experiments

Example II

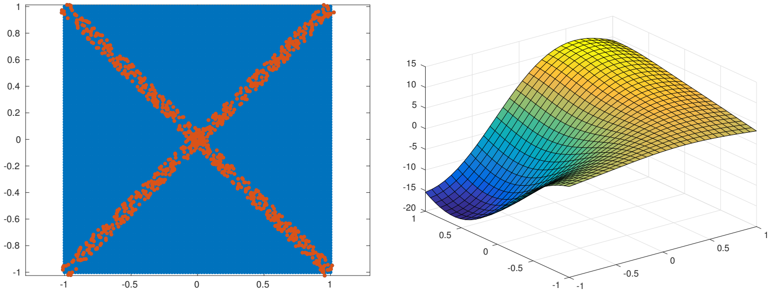


Figure: Gaussian Kernel with $\epsilon = 1$, $f \in \mathcal{H}(\Omega)$, $\Omega = [-1, 1]^2$.

Numerical experiments

Example II

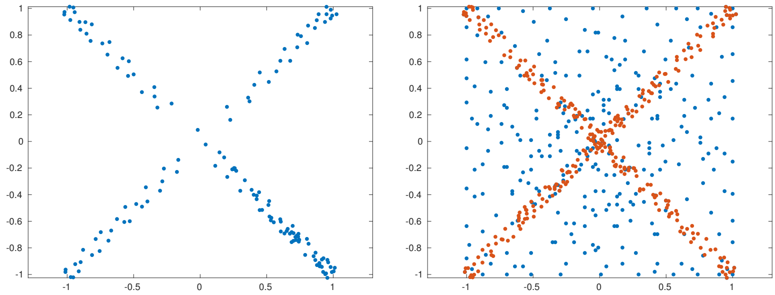


Figure: Gaussian Kernel with $\epsilon = 1$, $f \in \mathcal{H}(\Omega)$, $\Omega = [-1, 1]^2$.

Numerical experiments

Example II

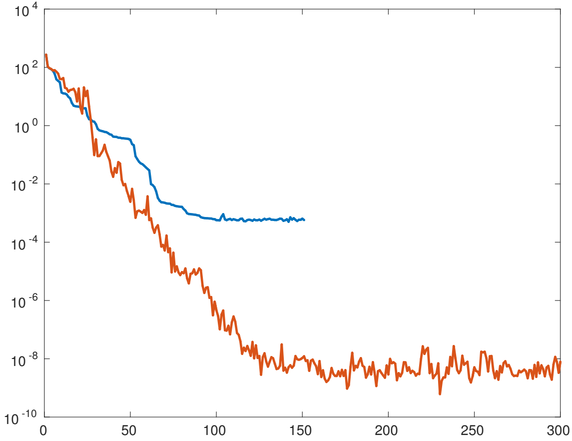
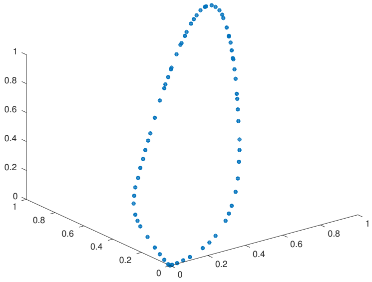
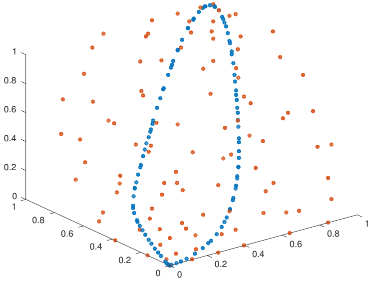


Figure: Gaussian Kernel with $\epsilon = 1$, $f \in \mathcal{H}(\Omega)$, $\Omega = [-1, 1]^2$.

Numerical experiments

Example III





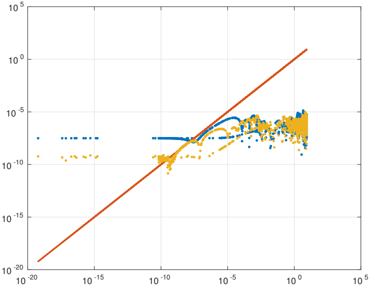
- $M = 12742$ samples
- $N = 200000$ points on a uniform grid
- Random permutation of data: 80% train, 10% val., 10% test (is this restrictive?)

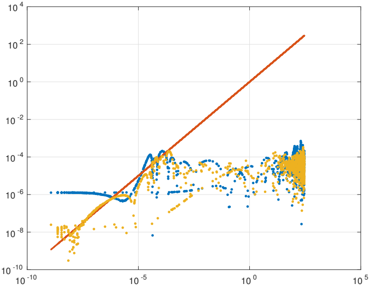
- $\varepsilon \in [0.1, 4]$, validate on val. set

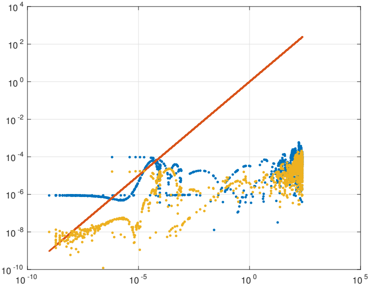
- Sym: $n = 70$ selected samples/centers
- Non-Sym: $n = 102$ selected samples and centers

Numerical experiments

Example III







Thank you!

[1] S. De Marchi, R. Schaback, and H. Wendland. "Near-optimal data-independent point locations for radial basis function interpolation". *Adv. Comput. Math.* 23.3 (2005), pp. 317–330.

[2] B. Haasdonk and G. Santin. "Non-symmetric kernel greedy interpolation." University of Stuttgart, in preparation, 2016.

[3] L. Ling and R. Schaback. "An improved subspace selection algorithm for meshless collocation methods". *Internat. J. Numer. Methods Engrg.* 80.13 (2009), pp. 1623–1639.

[4] C. A. Micchelli and M. Pontil. "On learning vector-valued functions". *Neural Comput.* 17.1 (2005), pp. 177–204.

[5] S. Müller and R. Schaback. "A Newton basis for kernel spaces". *J. Approx. Theory* 161.2 (2009), pp. 645–655.

[6] E. Novak and H. Woźniakowski. *Tractability of multivariate problems. Vol. 1: Linear information.* Vol. 6. 2008, pp. xii+384.

[7] R. Schaback and H. Wendland. "Approximation by Positive Definite Kernels". *Advanced Problems in Constructive Approximation.* Ed. by M. Buhmann and D. Mache. Vol. 142. 2002, pp. 203–221.

[8] D. Wirtz and B. Haasdonk. "A Vectorial Kernel Orthogonal Greedy Algorithm". *Dolomites Research Notes on Approximation* 6 (2013). Proceedings of DWCAA12, pp. 83–100.