An Improved Vectorial Kernel Orthogonal Greedy Algorithm

at

DWCAA 2012

Dipl. Math. Dipl. Inf. Daniel Wirtz
Monday, September 10th

Dipl. Math. Dipl. Inf. Daniel Wirtz
Jun.-Prof. Dr. Bernard Haasdonk
Institute of Applied Analysis and Numerical Simulation, University of Stuttgart
Pfaffenwaldring 57, D-70569 Stuttgart
daniel.wirtz@mathematik.uni-stuttgart.de
Model reduction with kernels

- Nonlinear dynamical systems
  - Spatially discretized partial differential equations
  - Analog macromodeling / Multicomponent electric circuits
  - Systems biology
  - ...

- Multiscale models
  - Human spine: Macroscale multibody system with microscale intervertebral spine discs
  - Nonclassical shock waves propagation in porous media
  - Neuromuscular Cell models
  - ...

Our window on Kernel Methods
Model reduction
Nonlinear dynamical systems

General nonlinear parametrized dynamical system

\[ x'(t) = f(x(t), t, \mu) + B u(t), \quad x(0) = x_0(\mu), \]  

- System state \( x(t) \in \mathbb{R}^d \) for time \( t \in [0, T] \), often \( d \gg 1 \)
- Parameters \( \mu \in \mathcal{P} \subset \mathbb{R}^p \) and initial state \( x_0(\mu) \in \mathbb{R}^d \)
- Input function \( u : [0, T] \rightarrow \mathbb{R}^m \) and \( B \in \mathbb{R}^{d \times m} \)

Key idea: Approximate \( f \) by kernel expansion

\[ f(x, t, \mu) \approx \tilde{f}(x, t, \mu) = \sum_{i=1}^{N} c_i \Phi(x, x_i) \Phi_t(t, t_i) \Phi_P(\mu, \mu_i) \]  

- Coarse macroscale model with inputs $y \in \mathcal{P}_C \subseteq \mathbb{R}^{pc}$
- Detailed microscale model with inputs $x \in \mathcal{P}_D \subseteq \mathbb{R}^{pd}$

**Typical model interaction**

![Diagram](attachment:image.png)

- Communication cycle between models can be seen as (vectorial) function $f$
- Try to approximate $f$ by kernel expansion $\tilde{f}$
Kernels, RKHS

- Domain $\Omega \subseteq \mathbb{R}^d$ closed, sym. pos. def. kernel $\Phi : \Omega \times \Omega \rightarrow \mathbb{R}$.
- For $X = \{x_1, \ldots, x_m\} \subset \Omega$ have $\mathcal{H}^X := \langle \{\Phi(x, \cdot) \mid x \in X\} \rangle$.
- $\Phi$ induces native/reproducing kernel Hilbert space (RKHS)

$$\mathcal{H} := \overline{\mathcal{H}^\Omega}, \quad \langle \Phi(x, \cdot), \Phi(y, \cdot) \rangle_{\mathcal{H}} = \Phi(x, y).$$

- Vectorial space $\mathcal{H}^q := \{f : \Omega \rightarrow \mathbb{R}^q \mid f_j \in \mathcal{H}, j = 1 \ldots q\}, q \in \mathbb{N}$ with standard scalar product & norm

$$\langle f, g \rangle_{\mathcal{H}^q} := \sum_{j=1}^{q} \langle f_j, g_j \rangle_{\mathcal{H}}, \quad \|f\|_{\mathcal{H}^q} = \sqrt{\langle f, f \rangle_{\mathcal{H}^q}}.$$
Nonlinear approximation
Overall setting

- Have $\tilde{f}$ from previous examples
- Given $\Omega$ and $q \in \mathbb{N}$, assume $\tilde{f} \in \mathcal{H}^q$ for suitable RKHS.
- Assume to have training data

$$X = \{x_1, \ldots, x_m\} \subset \Omega, \quad Y := \{y_k = f(x_k) \mid x_k \in X\}$$

Approximation goal

$$\tilde{f}(x) \approx f(x) = \sum_{i=1}^{N} c_i \Phi(x, x_i) \in \mathcal{H}^q, \quad N \ll |X|$$

- Find suitable kernel $\Phi$
- Find suitable centers $x_i \in X$ (shared over $q$)
- Find suitable coefficients $c_i \in \mathbb{R}^q$
Orthogonal projection

- Linear subspace $S \subseteq \mathcal{H}$, e.g. $S = \mathcal{H}^X$, $X = \{x_1, \ldots, x_m\} \subset \Omega$
- Orthogonal projection operator $\mathcal{P}_S : \mathcal{H} \to S$, $f \mapsto \mathcal{P}_S[f]$:
  $$\langle f - \mathcal{P}_S[f], g \rangle_{\mathcal{H}} = 0 \quad \forall \ g \in S$$
- Is pointwise interpolation for $S = \mathcal{H}^X$ and certain conditions
- Component-wise extension to $\mathbb{R}^q$ denoted by $P^q_S : \mathcal{H}^q \to S^q$

Orthonormal remainders

For $x \in \Omega$ the $\mathcal{H}^X$-orthonormal remainder $\phi_x$ of $\Phi(x, \cdot)$ is

$$\tilde{\phi}_x := \Phi(x, \cdot) - \mathcal{P}_{\mathcal{H}^X}[\Phi(x, \cdot)],$$

$$\phi_x := \begin{cases} \tilde{\phi}_x / \| \tilde{\phi}_x \|_{\mathcal{H}}, & \Phi(x, \cdot) \notin \mathcal{H}^X, \\ 0, & \Phi(x, \cdot) \in \mathcal{H}^X. \end{cases}$$
Key ideas: Greedy approximation

- Build hierarchical linear subspaces with increasingly good approximations of $\tilde{f}$
- Greedy step: Select an element $x \in \Omega$ whose induced $\Phi(x, \cdot)$ yields the smallest approximation error after extension

$$
\inf_{x \in \Omega} \min_{g \in (\mathcal{H}^X \oplus \langle \Phi(x, \cdot) \rangle)^q} \| f - g \|_{\mathcal{H}^q}^2
\quad (3)
$$

- Can show: Infima of (3) are taken at any $x^*$ that satisfies

$$
x^* \in \arg \max_{x \in \Omega} \sum_{j=1}^{q} \langle f_j, \phi_x \rangle_{\mathcal{H}}^2
$$

- (3) = $\inf_{x \in \Omega} \| f - P^q_{\mathcal{H}^X \oplus \langle \Phi(x, \cdot) \rangle} [f] \|_{\mathcal{H}^q}^2 = C - \sup_{x \in \Omega} \sum_{j=1}^{q} \langle f_j, \phi_x \rangle_{\mathcal{H}}^2$

- $\sup$ is actually $\max$ as $\Omega$ closed and $\langle f_j, \phi_x \rangle_{\mathcal{H}}$ continuous
Nonlinear approximation
Vectorial Kernel Orthogonal Greedy Algorithm

From the previous ideas we obtain the following algorithm:

**VKOG-Algorithm**

Let \( f \in \mathcal{H}^q \), define \( X_0 := \emptyset, f^0 := 0 \) and for \( m > 0 \) the sequences

\[
  x_m \in \arg \max_{x \in \Omega} \sum_{j=1}^{q} \langle f_j, \phi_x^{m-1} \rangle_{\mathcal{H}}^2,
\]

\( X_m := X_{m-1} \cup \{x_m\} \),

\( f^m := \mathcal{P}^q_{\mathcal{H}X_m}[f] \).

\( \phi_x^m \) is orthonormal remainder of \( \Phi(x, \cdot) \) w.r.t. \( X_m \) for any \( m, x \).

In practice:

- Have finite training data \( X \subset \Omega \), maximize (4) over \( X \setminus X_m \)
- \( f^m \) is component- and pointwise interpolation of \( f \) on \( X_m \)
- Stop if \( \|f - f^m\|_{\infty(X)} < tol \) or \( m = M_{max} \)
Nonlinear approximation

Relations to existing vectorial greedy algorithms

Turns out this is very similar to a vectorial algorithm “WSOGA2” in [Leviatan & Temlyakov(2005), §3]

“Weak Simultaneous Orthogonal Greedy Algorithm 2”

Let $\Phi$ be a normalized s.p.d. kernel spanning the RKHS $\mathcal{H}$ on $\Omega \subset \mathbb{R}^d$ and $q \in \mathbb{N}$. Further let $f \in \mathcal{H}^q$, define $X_0 := \emptyset$, $f^0 := 0$ and for $m > 0$ the sequences

$$x_m \in \arg \max_{x \in \Omega} \sum_{j=1}^{q} \langle f_j - f_{j-1}^m, \Phi(x, \cdot) \rangle_{\mathcal{H}}^2,$$  \hspace{1cm} (7)$$

$$X_m := X_{m-1} \cup \{x_m\},$$  \hspace{1cm} (8)$$

$$f^m := \mathcal{P}_{\mathcal{H}X_m}^q [f].$$  \hspace{1cm} (9)$$

For comparability, here for RKHS case and in strong formulation
Nonlinear approximation

So what's the difference?

Let's have a look at the extension selection criteria.

**VKOGA**

\[
\langle f_j, \phi_m \rangle^2_{H} = \frac{\langle f_j, \Phi(x, \cdot) - P_{H \times m-1} [\Phi(x, \cdot)] \rangle^2_{H}}{\| \tilde{\phi}_m \|_{H}^2}
\]

\[
\langle f_j - f_m^{-1} j, \Phi(x, \cdot) \rangle^2_{H} = \langle f_j(x) - f_m^{-1}(x) \rangle^2_{H} / \| \tilde{\phi}_m \|_{H}^2
\]

**WSOGA2**

Recall \( f_m^{-1} = P_{H \times m-1} [f_j] \).

\[
\langle f_j - f_m^{-1}, \Phi(x, \cdot) \rangle^2_{H} = \langle f_j(x) - f_m^{-1}(x) \rangle^2_{H}
\]

Note that

\[
\| \tilde{\phi}_m \|_{H}^2 \in [0, 1]
\]

\[
= \| \Phi(x, \cdot) - P_{H \times m-1} [\Phi(x, \cdot)] \|_{H}^2
\]
Nonlinear approximation

Illustration of selection criteria

Figure: VKOGA criteria (dashed-green) selects spatially very different point vs. WSOGA2 (solid-green)
Nonlinear approximation
Illustration of selection criteria

Figure: Selected VKOGA point (left cross) close to existing center: “penalization” of errors closer to centers
Nonlinear approximation
Theoretic analysis of selection

Local optimality

For extension choices $x^o$ (VKOGA) and $x^c$ (WSOGA2):

$$\langle f, \phi_{x^c} \rangle_H^2 \leq \max_{x \in \Omega} \langle f, \phi_x \rangle_H^2 = \langle f, \phi_{x^o} \rangle_H^2$$

- Locally maximal “gain” w.r.t. $H$-norm
Convergence rates for VKOQA

Let $M > 0$ and

$$\mathcal{H}_M := \left\{ f \in \mathcal{H} \mid f = \sum_{k=0}^{\infty} \alpha_k \Phi(x_k, \cdot), \sum_{k=0}^{\infty} |\alpha_k| \leq M \right\} \quad (10)$$

Then for any $f \in \mathcal{H}_M^q$, $f^m$ converges to $f$ no slower than

$$\|f - f^m\|_{\mathcal{H}_q} \leq \sqrt{qM} \left(1 + \frac{m}{q}\right)^{-\frac{1}{2}}, \quad m \geq 0 \quad (11)$$

More on vec. greedy methods & convergence: [Temlyakov(2008), Leviatan & Temlyakov(2005), Leviatan & Temlyakov(2006)]
Nonlinear approximation
A-posteriori bound

- Generated sequence $x_1, \ldots, x_m$ contains more information
- Define

$$c_m := \max_{x \in \Omega} \left\| \tilde{\phi}_x^{m-1} \right\|_{\mathcal{H}}^2, \quad m > 0$$  \hspace{1cm} (12)

A-posteriori approximation error bound

$$\| f - f^m \|_{\mathcal{H}^q} \leq \sqrt{qM} \left( 1 + \frac{1}{q} \sum_{k=1}^{m} \frac{1}{c_k} \right)^{-\frac{1}{2}}, \quad m \geq 0$$  \hspace{1cm} (13)

- As $c_m \leq 1$, using $c_m \equiv 1 \forall m$ yields the a-priori convergence rate.
- Rather of theoretical interest
Experiments
Experiment #1 - Setup

Experiment #1 Setup - 1D

- Domain $\Omega = [-10, 10]$
- Training samples $X \subset \Omega$, $|X| = 201$
- $\mathcal{H}$ induced by Gaussian $\Phi$ with $\gamma = 2.8888$
  (s.t. $\Phi(x, y) < 0.05$ for $||x - y|| > \frac{1}{4} \text{diam}(\Omega)$)
- Test function

$$f = \sum_{i=1}^{N} c_i \Phi(x_i, \cdot)$$

with $N = 12$ unif. distr. $x_i \in [-10, 10]$ and $c_i \in [-150, 150]$
- Stopping at maximum relative error $< 10^{-2}$ or $m = 100$
Experiments

Experiment #1 - Center selection

Figure: Selected centers from both algorithms. VKOGA: 9, WSOGA2: 12. Stop at $10^{-2}$ rel. error on $X$
Experiment #2 - Setup - 5D

- Domain $\Omega = [-10, 10]^5$
- Training/validation samples $X, X_v \subset \Omega$, $|X| = 2500$, $|X_v| = 1000$
- $\mathcal{H}$ induced by Gaussian $\Phi$ with $\gamma = 8.9526$
  (s.t. $\Phi(x, y) < 0.4$ for $||x - y|| > \frac{1}{4} \text{diam}(\Omega)$)
- Test function
  $$f = \sum_{i=1}^{N} c_i \Phi(x_i, \cdot)$$

  with $N = 20$ unif. distr. $x_i \in [-10, 10]^5$ and $c_i \in [0, 15]^5$
- Stopping at maximum relative error $< 10^{-3}$ or $m = 200$
Experiments
Experiment #2 - Absolute Errors

Figure: Absolute approximation errors on training data $X$ (left) and validation set $X_v$ (right)
Experiments
Experiment #2 - Relative Errors

Figure: Relative approximation errors on training data $X$ (left) and validation set $X_v$ (right)
Experiments
Experiment #2 - $\mathcal{H}$-norm error decay and bounds

Figure: $H$-norm decay and upper convergence bounds with $M = 163.238$
Conclusion
What did we see?

Summary

- Vectorial kernel-based approximation algorithm
- Greedy locally optimal extension selection
- Related to WSOGA2 [Leviatan & Temlyakov(2005), §3]
- Works well for medium size dimensions
Outlook & open questions

- Weighted algorithm: Bounding $\| \phi_x \|_\mathcal{H} \geq \epsilon$, $\epsilon \in [0, 1]$
  - Continuous switch between VKOGA and WSOGA2
- Stability analysis
  - Adopting preconditioning techniques, e.g. [Schaback(2008)]
- Relations to more general settings
- Extensions to higher dimensions

Thank you for your attention!

