M. Dihlmann · M. Drohmann · B. Haasdonk

Model reduction of parametrized evolution problems using the reduced basis method with adaptive time partitioning

Stuttgart, May 2011

a Institute of Applied Analysis and Numerical Simulation, University of Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany
{dihlmann, haasdonk}@mathematik.uni-stuttgart.de
www.ians.uni-stuttgart.de/agh

b Institute of Computational and Applied Mathematics, University of Münster, Einsteinstr. 62, 48149 Münster, Germany
martin.drohmann@math.uni-muenster.de

Abstract

Modern simulation scenarios require real-time or many query responses from a simulation model. This is the driving force for increased efforts in model order reduction for high-dimensional dynamical systems or partial differential equations. This demand for fast simulation models is even more critical for parametrized problems. There exist several snapshot-based methods for model order reduction of parametrized problems, e.g. proper orthogonal decomposition (POD) or reduced basis (RB) methods. An often faced problem is that the produced reduced models for a given accuracy tolerance are still of too high dimension. This is especially the case for evolution problems where the model shows high variability during time evolution. We will present an approach to gain control over the online complexity of a reduced model by an adaptive time domain partitioning. Thereby we can prescribe simultaneously a desired error tolerance and a limiting size of the dimension of the reduced model. This leads to fast and accurate reduced models. The method will be applied to an advection problem.

Keywords

model order reduction; reduced basis method; evolution problem; parametrized partial differential equation; adaptive time partitioning

MOR and A-Posteriori Error Estimation for Parameterized Kernel-Based Systems

at

CAAM,

Dipl. Math. Dipl. Inf. Daniel Wirtz

Monday, April 16th

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
Remark on residual
Introduction
Model reduction methodology
Error estimation
Pictures!
Nonlinear approximation
  - Support vector machines
  - VKOGA: Greedy algorithms
Outlook
RB a-posteriori error estimation
remarks on riesz representant of residual

- Have \( a(e(\mu), v; \mu) = a(u_h(\mu), v; \mu) - a(u_N(\mu), v; \mu) = f(v) - a(u_N(\mu), v; \mu) =: r(v; \mu) \forall v \in X_h \)

- Riesz representant \( \hat{e}(\mu) : a(\hat{e}(\mu), v; \mu) = r(v; \mu) \forall v \in X_h \)

- From separability:
  \[
  r(v; \mu) = \sum_{n=1}^{Q_N} E_n(\mu) h_n(v), \quad Q_N = Q_f + Q_{aN}
  \]

- Have \( Q_N \) riesz representants \( \hat{g}_n : h_n(v) = (\hat{g}_n, v) \forall v \in X_h \)

- Thus \( \hat{e}(\mu) = \sum_{n=1}^{Q_N} E_n(\mu) \hat{g}_n \)

- Have representation \( \hat{g}_n = \sum_{i=1}^{\lfloor X_h \rfloor} \alpha_i \varphi_i, \quad \{\varphi_i\}_i \) basis of \( X_h \)

- Gives \( Q_N \) large systems
  \[
  h_n(\varphi_i) = \sum_{j=1}^{\lfloor X_h \rfloor} \alpha_j^n(\varphi_i, \varphi_j)
  \]

- In matrix notation
  \[
  X^T \alpha_n = h_n, \quad X_{ij} = (\varphi_i, \varphi_j)
  \]

- Finally
  \[
  (\hat{g}_n, \hat{g}_m) = h_n(\hat{g}_m) = h_n X^{-1} h_m
  \]

Remarks on error estimators
Nonlinear PDE case

- Combination of EI/RB methods for nonlinear PDEs
- Includes a-posteriori error estimation for explicit discretizations
  - Assume $I_d - L_E(\mu, t^k)$ Lipschitz continuous
  - Uses extended interpolation space $\mathcal{V}_{M+1}$
  - Offline/online decomposable
- [Drohmann et al. (2010) Drohmann, Haasdonk & Ohlberger, SISC 2012]: RB/EI for nonlinear PDEs
- Explicit/implicit discretizations, with EI for directional derivatives
  - Lipschitz continuity
  - Extended EI base $M + M'$ for error estimation
  - Residuals from Newton steps
  - ...
Biochemical system: cell apoptosis simulation (M. Daub)

- Described by PDE
- Spatial discretization yields large scale dynamical system
- Possibly parameterized, i.e. TNF receptor inputs

Figure: Cell death model, presented by Dr. K. Nickel
Introduction

Base parameterized dynamical system

General system structure considered

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

- System state \( x(t) \in \Omega \subseteq \mathbb{R}^d \) and output \( y(t) \in \mathbb{R}^k \) for \( t \in [0, T] \)
- Parameters \( \mu \in \mathcal{P} \subseteq \mathbb{R}^p \)
- Input function \( u : [0, T] \rightarrow \mathbb{R}^m \) and \( B(t, \mu) \in \mathbb{R}^{d \times m} \)
- Initial state \( x_0(\mu) \in \mathbb{R}^d \)
- Output conversion \( C(t, \mu) \in \mathbb{R}^{k \times d} \)
Affine-parametric $B, C, x_0$

Important assumption on structure for input, output and initial value:

\[
B(t, \mu) = \sum_{i=1}^{Q_B} \theta_i^B(t, \mu) B_i, \quad C(t, \mu) = \sum_{i=1}^{Q_C} \theta_i^C(t, \mu) C_i \tag{4}
\]

\[
x_0(\mu) = \sum_{i=1}^{Q_{x_0}} \theta_i^{x_0}(t, \mu) x_i^0 \tag{5}
\]

- $Q_B, Q_C, Q_{x_0} \in \mathbb{N}$ small, e.g. $\in \{1 \ldots 20\}$
- Scalar coefficient functions
  $\theta_i^B(t, \mu), \theta_i^C(t, \mu), \theta_i^{x_0}(t, \mu) : [0, T] \times \mathcal{P} \rightarrow \mathbb{R}$
- Input mapping matrices $B_i \in \mathbb{R}^{d \times m}$
- Output mapping matrices $C_i \in \mathbb{R}^{k \times d}$
- Initial state vectors $x_i^0 \in \mathbb{R}^d$
System kernel expansion \( f(x, t, \mu) : \Omega \times [0, T] \times \mathcal{P} \rightarrow \mathbb{R} \)

System nonlinearity is a *kernel expansion* of the form

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

(6)

- Expansion centers \( x_i \in \mathbb{R}^d, t_i \in [0, T], \mu_i \in \mathcal{P}, i = 1 \ldots N \)
- \( \Phi_s, \Phi_t, \Phi_P \) are scalar state, time and parameter *kernels*
- Coefficient vectors \( c_i \in \mathbb{R}^d, i = 1 \ldots N \)
- Kernel examples:
  - Linear: \( \Phi(x, y) = \langle x, y \rangle \)
  - Gaussian: \( \Phi(x, y) = e^{-\frac{||x-y||^2}{\gamma^2}}, \gamma > 0 \)

- Initial idea for approach:
Introduction
Background on kernel approach

Kernels - Basic definition

- Positive definite function \( \Phi : \Omega \times \Omega \rightarrow \mathbb{R} \), commonly \( \Omega \subset \mathbb{R}^d \)
- Basic algebraic combinations of kernels are kernels
- Every s.p.d. kernel induces a Reproducing Kernel Hilbert Space:

\[ \mathcal{H} = \overline{\langle \{ \Phi(x, \cdot) \mid x \in \Omega \} \rangle} \]

- Scalar product \( \langle \Phi(x, \cdot), \Phi(y, \cdot) \rangle_{\mathcal{H}} = \Phi(x, y), \quad x, y \in \Omega \)

Kernel approximation context

- Find representations (6) for more general nonlinear functions
- Least-Squares [Wendland(2005)]
- Kernel interpolation / best-subspace-approximation [Wendland(2005)]
- Support Vector Machines [Schölkopf & Smola(2002), Steinwart & Christman(2008)]
- More later!
Model reduction
Model reduction
Reduction via subspace projection

Projection subspace given by $V, W \in \mathbb{R}^{d \times r}$ biorthogonal ($W^t V = I_r$).

Ritz-Galerkin projection of the full system (1)

\[
\begin{align*}
    z'(t) &= W^t f(Vz(t), t, \mu) + W^t B(t, \mu)u(t) \\
    z(0) &= W^t x_0(\mu) \\
    y^r(t) &= C(t, \mu)Vz(t)
\end{align*}
\]

- Reduced variable $z(t) \in \mathbb{R}^r$ at times $t \in [0, T]$, $r << d$
- We use a.o. **POD-Greedy** for $V, W$ computation [Haasdonk & Ohlberger(2008), Haasdonk(2011)]
Model reduction
Efficient projection of kernel expansions

- \( W^t f(V z(t), t, \mu) : \mathbb{R}^r \to \mathbb{R}^r \), but costly as \( V z(t) \in \mathbb{R}^d \)
- Certain kernels allows lossless, efficient evaluation

Translation- and rotation invariant kernels

Assume

- \( \Phi_s(x, y) = \phi(||x - y||) \) for some scalar \( \phi : \mathbb{R}^+ \to \mathbb{R} \)
- Centers \( x_i = V z_i \), i.e. \( x_i \in \langle v_1, \ldots, v_r \rangle \) for \( V = [v_1, \ldots, v_r] \)

Then

\[
\Phi_s(V \hat{z}, x_i) = \phi(||V \hat{z} - V z_i||) = \phi(||\hat{z} - z_i||_{V^t V}) =: \Phi^r_s(\hat{z}, z_i),
\]

with \( V^t V \in \mathbb{R}^{r \times r} \) precomputable.

- Different class (scalar product kernels) proposed in [Phillips et al. (2003) Phillips, Afonso, Oliveira & Silveira]
Model reduction
Reduced model with all low-dimensional components

Reduced form of full system (7)

\[
\begin{align*}
z'(t) &= \sum_{i=1}^{N} \tilde{c}_i \Phi_s^r(z(t), z_i) \Phi_t(t, t_i) \Phi_P(\mu, \mu_i) + \sum_{i=1}^{Q_B} \theta_i^B(t, \mu) \tilde{B}_i u(t) \\
z(0) &= \sum_{i=1}^{Q_{x_0}} \theta_{x_0}^i(t, \mu) \tilde{x}_i^0 =: z_0(\mu), \quad y^r(t) = \sum_{i=1}^{Q_C} \theta_C^i(t, \mu) \tilde{C}_i z(t)
\end{align*}
\]

- Complexity independent of \(d\)
- Small expansion coefficient vectors \(\tilde{c}_i = W^t c_i, i = 1 \ldots N\)
- Reduced components
  \(\tilde{B}_i = W^t B_i \in \mathbb{R}^{r \times m}, \quad \tilde{C}_i = C_i V \in \mathbb{R}^{r \times k}, \quad \tilde{x}_i^0 = W^t x_i^0\)
- Allows full offline/online decomposition
Error estimation
Error estimation
Definitions

- A-posteriori error estimation for linear parameterized systems in [Haasdonk & Ohlberger(2011)]
- Reconstructed full state space variable $x^r(t) := Vz(t)$.
- Reduction error: $e(t; \mu) := x(t) - x^r(t)$, $t \in [0, T]$
- $e(t; \mu)$ denotes dependency on given $\mu \in \mathcal{P}$

Error system for $e(t; \mu)$

$$
e'(t; \mu) = f(x(t), t, \mu) - VW^t f(Vz(t), t, \mu) + (I_d - VW^t)B(t, \mu)u(t)$$
$$e(0; \mu) = (I_d - VW^t)x_0(\mu)$$

- Assume $G \in \mathbb{R}^{d \times d}$ induces norm on $\Omega$.
- Let $E_{x_0}(\mu) := \| (I_d - VW^t)x_0(\mu) \|_G$ be the initial error.
Error estimation
Global Lipschitz Estimator

Estimating the norm of \( e'(t; \mu) \) and the comparison lemma [Hale(1969), p.32] yield an a-posteriori error estimator.

**Theorem (Global Lipschitz constant Estimator (GLE))**

Let \( \Phi_s \) be a \( L_{\Phi_s} \)-Lipschitz continuous kernel (w.r.t. first variable). Then the state space error is bounded via

\[
\| e(t; \mu) \|_G \leq \Delta_{GLE}(t, \mu) \quad \forall \ t \in [0, T], \mu \in \mathcal{P} \text{ with}
\]

\[
\Delta_{GLE}(t, \mu) := \int_0^t \alpha(s, \mu) e^s \int_s^t \beta(r, \mu) dr \ ds + E_{x_0}(\mu),
\]

\[
\alpha(t, \mu) := \left\| (I_d - VW)^t \left( f(x^r(t), t, \mu) + B(t, \mu)u(t) \right) \right\|_G,
\]

\[
\beta(t, \mu) := L_{\Phi_s} \sum_{i=1}^N \| c_i \|_G \Phi_t(t, t_i) \Phi_P(\mu, \mu_i).
\]
Error estimation
Improvement possibilities

Obvious: Large $\beta(t, \mu)$ makes estimation overly conservative!

Improvement ideas

- Utilize locality since $x^r(t)$ is available
- Use coarse a-priori bound $||e(t; \mu)|| \leq \Theta(t, \mu)$ for even more locality as it means $x(t) \in B_{\Theta(t, \mu)}(x^r(t))$
- Instead of maximum $\Phi_s$ gradient ($= L\Phi_s$) use local secant gradients

Local secant gradients computable for Bell functions $\phi$:

- $\phi \in C^2(\mathbb{R}^+_0, \mathbb{R}^+_0)$, $\phi$ bounded, $\phi' \leq 0$
- $\exists r_0 > 0 : \phi''(r)(r - r_0) \forall r \neq r_0$ (unique saddle point)
- $\phi$ induces radial kernel $\Phi_s(x, y) := \phi(||x - y||_G)$
- Example: Gaussian is induced by bell function $\phi(r) = e^{-r^2/\gamma^2}$
Error estimation
Illustration of local secant gradient error estimation for Gaussian

Figure: Illustration of local Lipschitz constants, using $d_i(t) := ||x^T(t) - x_i||_G$, $i \in \{1 \ldots N\}$
Error estimation
Local secant gradient error estimation

**Theorem (Local Secant Lipschitz error Estimator (LSLE))**

Let \( \Phi_s(x, y) = \phi(||x - y||) \), \( x, y \in \mathbb{R}^d \) for a bell function \( \phi \) and \( e(t; \mu) \leq \Theta(t, \mu) \) \( \forall t, \mu \). Then \( ||e(t; \mu)|| \leq \Delta_{LSLE}^\Theta(t, \mu) \) as with GLE but new \( \beta(t, \mu) \):

\[
\beta(t, \mu) := \sum_{i=1}^{N} L_\Theta(d_i(t)) ||c_i||_G \Phi_t(t, t_i) \Phi_P(\mu, \mu_i),
\]

(11)

\[
L_\Theta(d_i(t)) := \left| \frac{\phi(d_i(t)) - \phi(r_{\Theta,i})}{d_i(t) - r_{\Theta,i}} \right|,
\]

(12)

\[
r_{\Theta,i} := \begin{cases} 
  d_i(t) + \text{sign}(r_i - d_i(t)) \Theta(t, \mu) & r_i \notin \Omega(t, \mu) \\
  r_i & r_i \in \Omega(t, \mu).
\end{cases}
\]

(13)

\[
r_i := \arg \min_{r \in \mathbb{R}_0^+} \frac{\phi(d_i(t)) - \phi(r)}{d_i(t) - r}
\]

(14)

with \( \Omega(t, \mu) := B_{\Theta(t, \mu)}(d_i(t)) \cap \mathbb{R}_0^+ \)
Error estimation
Output error and computational aspects

Output error bounds

The output error can be bounded by

\[ \| e_y(t; \mu) \| \leq C_o(\mu) \| e(t; \mu) \|_G \left[ \leq C_o(\mu) \Delta_{GLE}(t, \mu) \right] \]

\[ C_o(\mu) := \sup_{t \in [0, T]} \| C(t, \mu) \|_G \]

Computation of \( r_i \)

- \( r_i \) efficiently computable via penalized Newton iteration
- \( r_0 \{ \leq \} d_i(t) \Rightarrow r_i \{ \leq \} r_0 \) can be used
- Low number of iterations for subsequent timesteps \( t_n \rightarrow t_{n+1} \)
- For details see [Wirtz & Haasdonk(2012)]
Error estimation
Application of coarse a-priori bound

Iterative scheme using a-priori bounds

- Start with $\Theta_0(t, \mu) \equiv \infty$
- Set $\Theta_n(t, \mu) := \Delta_{LSLE}^{\Theta_{n-1}}(t, \mu), \ n \geq 1$
- Let

$$\Gamma_n := \{ t \in [0, T] \mid \exists \ i \in \{1 \ldots N\} : \ r_i \notin B_{\Theta_n(t)}(d_i(t)) \}.$$ (15)

- Then, if $\Gamma_1 \neq \emptyset$:

$$\Gamma_{n+1} \neq \emptyset \ \forall \ n \in \mathbb{N},$$

$$\Delta_{LSLE}^{\Theta_{n+1}}(t) \begin{cases} < \Delta_{LSLE}^{\Theta_n}(t), & t > \inf \Gamma_{n+1} \\ = \Delta_{LSLE}^{\Theta_n}(t), & t \leq \inf \Gamma_{n+1} \end{cases} \ \forall \ n \in \mathbb{N}. \quad (16)$$

- Allows balancing of computational costs vs. sharpness
Error estimation
Offline-online decomposition

**Offline phase**
- Large matrices and coefficients:
  \[ \alpha(t, \mu) = \| (I_d - VW^t) (f(x^r(t), t, \mu) + B(t, \mu)u(t)) \| \]

**Online phase**
- Solve small auxiliary ODE

\[ v'(t) = \beta(t, \mu)v(t) + \alpha(t, \mu) \]  \hspace{1cm} (17)
\[ v(0) = E_{x_0}(\mu) \]  \hspace{1cm} (18)

All this work for non-parameterized systems: [Wirtz & Haasdonk(2012)]
Numerical experiments
Numerical experiments and results

Two more estimator versions

Two more estimator variants before we look at Numerical results:

**Time-discrete heuristic variant ("LsLE TD")**

- Motivated from convergence result of iterative scheme
- \( \Delta^{\Theta}_{\infty} \) reproduces itself when used as a-priori bound
- Numerically: Use error from previous time-step as a-priori bound

**Expensive lower bound ("Lower Bound")**

- Let \( \beta(t, \mu) = \frac{||f(x(t),t,\mu) - f(x^r(t),t,\mu)||}{||x(t) - x^r(t)||} \)
- Smallest possible \( \beta \) estimation, since this is the best local Lipschitz constant for any \( t \in [0, T], \mu \in \mathcal{P} \)
- Expensive but lower bound to any estimator using the above structure.
Numerical experiments and results
The setting

- Let $d = 240000$, $T = 20$, $G = I_d$ and $1 := (1 \ldots 1)^t \in \mathbb{R}^d$
- Parameter domain $\mathcal{P} = [0, 1] \times [0, 10] \times [-1, 1]$
  - $\mu_2$ is an “expansion parameter” influencing the system’s inner dynamics
  - $\mu_3$ sets the “initial value”
  - $\mu_1$ will be used/discussed later.
- Kernel expansion with $N = 20$, $\Phi_t \equiv 1$ and
  $$x_i := \frac{50(i-1)}{N-1} 1, \ \mu_i := \frac{10(i-1)}{N-1} (0, 1, 0)^t, \ i = 1 \ldots N$$
- $\Phi_s$ and $\Phi_P$ are Gaussians with $\gamma_s = 224$, $\gamma_P = 5.3733$
  - $\gamma$’s chosen for locality, e.g. $\Phi_s(x_i, x_j) < 10^{-5} \forall |i - j| \geq 2$
  - $\Phi_P$ only uses $\mu_2$
- Expansion coefficients $c_i = \exp(-x_i/15) \in \mathbb{R}^d$, $i = 1 \ldots N$
- $x_0(\mu) = \mu_3 1$, $C(t, \mu) = (1, \ldots, 1)^t/d$ (mean value)
- Explicit Euler integration scheme $\Delta t = 0.05$
- $V = W = (1, \ldots, 1)^t/\sqrt{d}$ plus rotation of 200 dims by $\theta = 0.05$
Numerical experiments and results
Error estimator comparison

Comparison of different estimated output errors $\|e_y(t)\|$ using $\mu_2 = 5, \mu_3 = -0.2$.

Figure: Absolute (l) and relative (r) $L^2$ state space errors of estimators using $\mu_2 = 5, \mu_3 = -0.2$
Numerical experiments and results
Error estimator runtimes

Figure: Estimation at $t = T$ against computation times for different error estimators
### Numerical experiments and results

**Error estimator statistics**

<table>
<thead>
<tr>
<th>Name</th>
<th>$\Delta(20)$</th>
<th>Time</th>
<th>Overestimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>True error</td>
<td>$3.650e-03$</td>
<td>21.43s</td>
<td>$1.000e+00$</td>
</tr>
<tr>
<td>GLE</td>
<td>$3.682e+15$</td>
<td>0.62s</td>
<td>$1.009e+18$</td>
</tr>
<tr>
<td>LSLE</td>
<td>$3.251e+01$</td>
<td>2.05s</td>
<td>$8.907e+03$</td>
</tr>
<tr>
<td>LSLE, 1 It</td>
<td>$1.568e-01$</td>
<td>2.79s</td>
<td>$4.295e+01$</td>
</tr>
<tr>
<td>LSLE, 2 It</td>
<td>$2.839e-02$</td>
<td>3.11s</td>
<td>$7.779e+00$</td>
</tr>
<tr>
<td>LSLE, 5 It</td>
<td>$2.801e-02$</td>
<td>4.13s</td>
<td>$7.674e+00$</td>
</tr>
<tr>
<td>LSLE TD</td>
<td>$2.801e-02$</td>
<td>0.90s</td>
<td>$7.674e+00$</td>
</tr>
<tr>
<td>Lower bound</td>
<td>$3.652e-03$</td>
<td>44.02s</td>
<td>$1.001e+00$</td>
</tr>
</tbody>
</table>

**Table:** Estimator statistics at $T = 20$
Numerical experiments and results

Error estimator runtimes

- Parameter sweep for $\mu_2$ (outputs over time against $\mu_2$)
- Error bounds by LSLE TD estimator in transparent light-red

Figure: Parameter sweep for $\mu_2 \in [0, 10]$, $\mu_1 = 0$, $\mu_3 = -0.2$
Numerical experiments and results
Adding external control

In order to show the influence of external input we choose an affine-parametric input matrix

\[ B(t, \mu) = \mu_1 \begin{pmatrix} 1 & 0 \end{pmatrix} + (1 - \mu_1) \begin{pmatrix} 0 & 1 \end{pmatrix} \in \mathbb{R}^{d \times 2} \]

and equip the system with two inputs

\[ u_1(t) = \begin{pmatrix} \frac{2}{5} \sin \left( \frac{t}{3} \right) \end{pmatrix} e^{- (12-t)^2}, \quad u_2(t) = \begin{pmatrix} \frac{1}{2} \sin \left( \frac{t}{2} \right) \\ 4e^{-7(12-t)^2} - \frac{1}{2} e^{-(5-t)^2} \end{pmatrix}. \]

- Both represent each an oscillating and a local stimulation of different kind
- \( \mu_1 \) is a smooth switch between oscillation and local stimulus
Numerical experiments and results

Parameter sweep for $\mu_1$ using $u_1$

Figure: Parameter sweep for input shift $\mu_1 \in [0, 1]$ and $u_1$, LSLE TD estimator
Numerical experiments and results

Parameter sweep for $\mu_1$ using $u_2$

Figure: Parameter sweep for input shift $\mu_1 \in [0, 1]$ and $u_2$, LSLE TD estimator
Numerical experiments and results

Error estimator runtimes

**Figure**: Parameter sweep for input shift and expansion parameter, $\mu_1 \in [0, 1]$, $\mu_2 \in [-8, 7]$ and using $u_1$
Nonlinear approximation

\[ f \longrightarrow \hat{f} \]
Approximation generation
Many choices!

**Solve one problem, get three new ones!**
- Kernel choice
- Training center selection
- Coefficient computation

**Center computation - possible choices**
- Random vectors $x_i \in \Omega$
- Trajectories of System states $x_1 = x(t_1), \ldots, x_N = x(t_l)$
- If $l \gg N$, select suitable subset of trajectories
- ...
Reduction methods
Computation of approximation

**Coefficient computation - possible choices**

- Least-Squares [Wendland(2005)]
- Kernel interpolation [Wendland(2005)]
- Support Vector Machines
  [Schölkopf & Smola(2002), Steinwart & Christman(2008)]
- Additionally: Reduced Set Expansions
  [Schölkopf & Smola(2002), §18.3]
- ...

Recent work:

**Adaptive approximation**

- Make data-driven selections
- Work-in-progress
Adaptive approximation
Support Vector Machines
Adaptive strategies from machine learning:

Support Vector Machines (SVM)

- Established method

- Consider training data \( D := \{(x_i, f(x_i)) \mid i = 1 \ldots N\} \) generated by distribution \( P_f \).

- For \( \epsilon > 0 \) we have an \( \epsilon \)-insensitive loss function

\[
L(y, f(x)) = |y - f(x)|_\epsilon := \max\{0, |y - f(x)| - \epsilon\}
\]

- Try to minimize the empirical expected risk

\[
R_{L,D}(g) = \frac{1}{N} \sum_{i=1}^{N} L(y_i, g(x_i)).
\]

with a \( g \in \mathcal{H} \)
Regularized optimization target

\[ \min_{f \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i)) + \lambda \| f \|_{\mathcal{H}}^2 \]

Primary problem

With \( C := \frac{1}{2 \lambda N} \) and slack variables \( \xi^+, - \):

\[ \min_{f \in \mathcal{H}, \xi^+, - \in \mathbb{R}^N} \frac{1}{2} \| f \|_{\mathcal{H}}^2 + C \sum_{i=1}^{N} (\xi_i^+ + \xi_i^-) \] (19)

\[ 0 \geq -\xi^+,- \] (20)

\[ 0 \geq y_i - f(x_i) - \epsilon - \xi_i^+ \] (21)

\[ 0 \geq -(y_i - f(x_i)) - \epsilon - \xi_i^+ \] (22)
Adaptive approximation

SVM
Adaptive approximation
Greedy methods

Current work; basic notation

- \( \Omega \subset \mathbb{R}^d, \Phi : \Omega \times \Omega \to \mathbb{R} \) s.p.d. kernel
- \( X = \{x_1, \ldots, x_m\} \subseteq \Omega, q \in \mathbb{N} \) and \( f \in \mathcal{H}^q \)
- \( \mathcal{H}^X := \langle \{\Phi(x, \cdot) \mid x \in X\} \rangle \)
- Orthogonal remainders

\[
\tilde{\phi}_x := \Phi(x, \cdot) - P_{\mathcal{H}^X} [\Phi(x, \cdot)], \quad \phi_x := \begin{cases} 
\frac{\tilde{\phi}_x}{\|\tilde{\phi}_x\|_{\mathcal{H}}}, & \Phi(x, \cdot) \notin \mathcal{H}^X, \\
0, & \Phi(x, \cdot) \in \mathcal{H}^X.
\end{cases}
\]

Optimal subspace extension

\[
\arg \min_{x \in \Omega} \min_{g \in (\mathcal{H}^X \oplus \langle \Phi(x, \cdot) \rangle)^q} \|f - g\|_{\mathcal{H}^q}^2 = \arg \max_{x \in \Omega} \sum_{j=1}^q \langle f_j, \phi_x \rangle_{\mathcal{H}}^2.
\]
Adaptive approximation
Greedy methods

Current work: VKOGA

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 $X_0 := \emptyset$, $f^0 := f$ and for $m > 0$ the sequences

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

(23)

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

(24)

$$f^m := P_{\mathcal{H}X_m}^q [f],$$

(25)

where $\phi_x^m$ denotes the orthonormal remainder of $\Phi(x, \cdot)$ with respect to $X_m$ for any given $m, x$. 
Adaptive approximation
Greedy methods

Convergence rates

Let $M > 0$ and

\[
\mathcal{H}^M := \left\{ f \in \mathcal{H} \mid f = \sum_{k=0}^{\infty} \alpha_k \Phi(x_k, \cdot), \sum_{k=0}^{\infty} |\alpha_k| \leq M \right\}
\]  \hspace{1cm} (26)

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

\[
\|f - f^m\|_{\mathcal{H}^q} \leq \sqrt{q}M \left(1 + \frac{m}{q}\right)^{-\frac{1}{2}}, \quad m \geq 0
\]  \hspace{1cm} (27)

More on (vectorial) greedy methods: [Temlyakov(2008), Leviatan & Temlyakov(2005), Leviatan & Temlyakov(2006)]
Adaptive approximation
VKOGA - Application

Stopping condition: Max $L^2$-error on training data: 1.000000e−002
Total $L^2$-error: 5.148545e−003
Adaptive approximation
VKOGA - Convergence rates

Theoretical upper bound

H-approximation error at step m

2 4 6 8 10 12 14

10^−5
10^−4
10^−3
10^−2
10^−1
10^0
10^1

2 4 6 8 10 12 14

H-approximation error at step m

Theoretical upper bound

10^1
10^0
10^−1
10^−2
10^−3
10^−4
10^−5
Secant gradient estimation:

$$\beta(t) = \sum_{i=1}^{N} \| c_i \|_G L_{\phi,C(t)} (V \hat{z}(t), x_i)$$

Large $N$ together with large $\| c_i \|_G$ are still a problem

Adaptive approximation generation / comparison
  - Greedy-type adaptive algorithms [Temlyakov(2008)]
  - Support Vector Machines without offset [Steinwart et al.(2011)Steinwart, Hush & Scovel]

Combinations with EI/DEIM approach [Barrault et al.(2004)Barrault, Maday, Nguyen & Patera, Chaturantabut & Sorensen(2009)] for very large scale systems
Combination of kernel methods & EI/DEIM

- Instead of
  \[ f(x, t, \mu) \approx \sum_{k=1}^{N} \alpha_k \Phi(x_k, t_k, \mu_k, x, t, \mu), \alpha_k \in \mathbb{R}^d \]

- Apply EI before:
  \[ f(x, t, \mu) \approx \sum_{i=1}^{M} \xi_i f_i(\mathcal{P}_i(x), t, \mu), \mathcal{P}_i \text{ loc. sel.} \]

- Replace
  \[ f_i(\mathcal{P}_i(x), t, \mu) \approx \sum_{k=1}^{N} \alpha^i_k \Phi(x_k, \mathcal{P}_i(x), \ldots), \dim x_k \ll d \]

- Advantage: Don't need inner access to \( f_i(\mathcal{P}_i(x), t, \mu) \), just learn
  \( (\mathcal{P}_i(x), t, \mu) \mapsto f_i(x, t, \mu) \)

- Error estimation for combined reduced system
Possible topics / questions

**Error estimations with/from DEIM**

- A-priori analysis for DEIM also for Operator-EI in [Haasdonk *et al.* (2007) Haasdonk, Ohlberger & Rozza]?
- i.e. a-priori for DEIM without using POD
- A-posteriori error estimation for DEIM
  - Analogous to [Wirtz & Haasdonk (2012)]?
  - Analogous to [Drohmann *et al.* (2010) Drohmann, Haasdonk & Ohlberger]?

