SAND2013-4510C

Stochastic estimation of reduced basis errors based on machine learning concepts

K. Carlberg   M. Drohmann

Sandia National Laboratories, Livermore, CA, (mdrohma@sandia.gov)

ModRed CIRM, Marseille, France, June 2013

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000

Error estimates for model order reduction methods

Model reduction implies error to high fidelity model: $e := u_h - u_{\mathrm{red}}$.

Error bounds

Error bounds η should be

- **rigorous**: $\eta \geq \|e\|$,
- **effective**: $\frac{\eta}{\|e\|} \leq C_{\mathsf{eff}}$ or "not too big"

  and
- **efficient**, i.e. quickly computable (as fast as reduced model)

Error estimates for model order reduction methods

Model reduction implies error to high fidelity model: $e := u_h - u_{\mathrm{red}}$.

Error bounds

Error bounds η should be

- **rigorous**: $\eta \geq \|e\|$,
- **effective**: $\frac{\eta}{\|e\|} \leq C_{\mathsf{eff}}$ or "not too big"

  and
- **efficient**, i.e. quickly computable (as fast as reduced model)

Often there is a trade-off between **efficiency** and **rigor/effectivity**.

Effectivity of error estimates

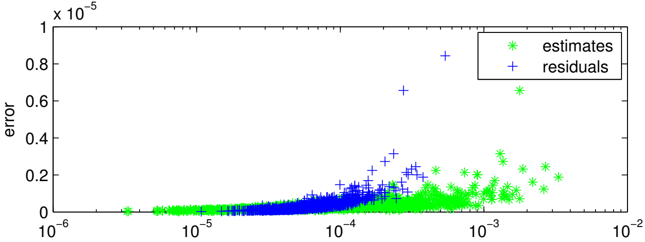
Effectivity. . .

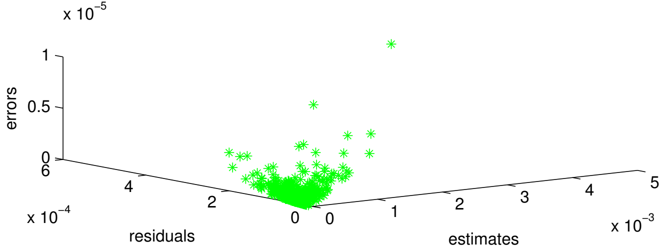
- . . . is necessary for the **selection of good** reduced bases.
- . . . ensures (certified) reduced solutions at **low** reduced basis size.

But. . .

Effective error estimates can be very **expensive**.

OBSERVATION





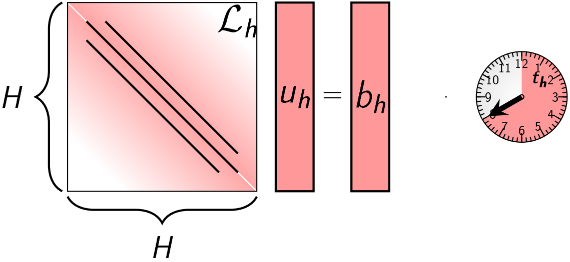
Outline

1. A posteriori error estimators
   - Reduced basis scheme
   - Ingredients
   - Effectivity of error estimators
2. Stochastic error estimate
3. Numerical experiments
   - Thermalblock-Problem
4. Summary and future work

Outline

High–dimensional model

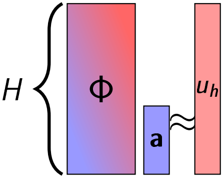
$(\partial_t u_h)+$



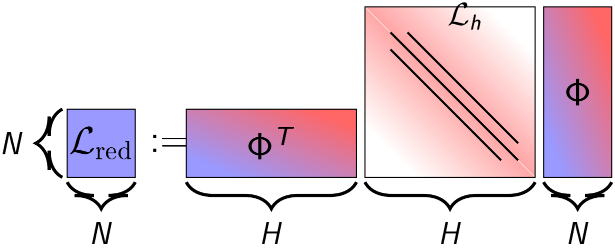
Offline phase

- Find reduced basis functions $\Phi := \{\varphi_1, \ldots, \varphi_N\}$,

- such that

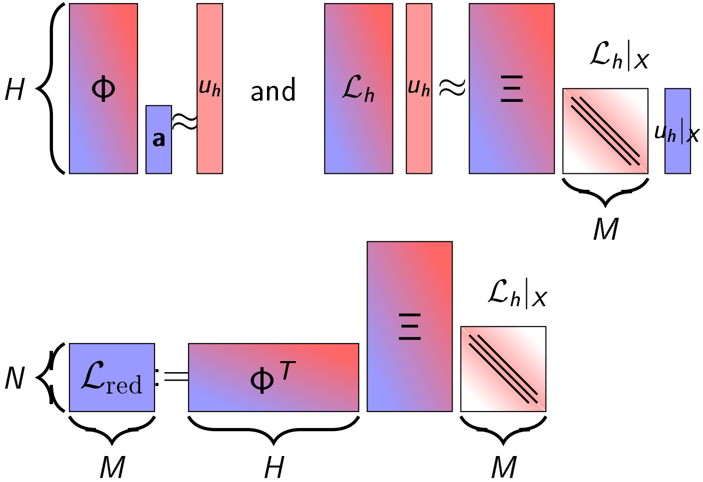


and



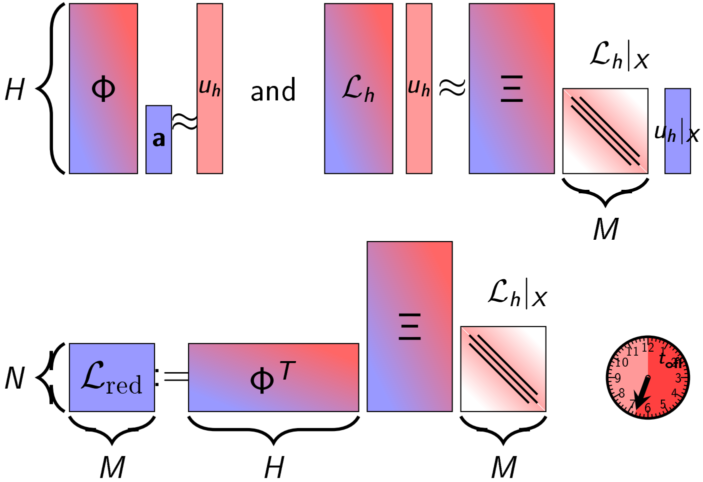
Offline phase

- Find reduced basis functions $\Phi := \{\varphi_1, \dots, \varphi_N\}$,
- interpolation points $X := \{x_1, \dots, x_M\}$ and functions $\Xi := \{\xi_1, \dots, \xi_M\}$,
- such that



Offline phase

- Find reduced basis functions $\Phi := \{\varphi_1, \dots, \varphi_N\}$,
- interpolation points $X := \{x_1, \dots, x_M\}$ and functions $\Xi := \{\xi_1, \dots, \xi_M\}$,
- such that



Reduced basis model

$$(\partial_t u_{\mathrm{red}} +) \quad N \left\{ \underbrace{\mathcal{L}_{\mathrm{red}}}_{M} \underbrace{\Phi|_X}_{N} \, \mathbf{a} = b_{\mathrm{red}} \right.$$

Error: $e := u_h - u_{\mathrm{red}}$ with $u_{\mathrm{red}} := \Phi \mathbf{a}$

A posteriori error estimator (DHO, 2012)

Estimator

$$\left\| u_h^k(\boldsymbol{\mu}) - u_{\mathrm{red}}^k(\boldsymbol{\mu}) \right\| \leq \eta_{N,M,M'}^k(\boldsymbol{\mu})$$

A posteriori error estimator

Theorem (A posteriori error estimator)

Assumptions:

- *Operator(s) fulfill "Lipschitz" properties:*
  - $\left\| u - v + \Delta t \mathcal{L}_I [u] - \Delta t \mathcal{L}_I [v] \right\| \geq \frac{1}{C_{I,\Delta t}} \|u - v\|_{\mathcal{W}_h}$
  - $\left\| u - v - \Delta t \mathcal{L}_E [u] + \Delta t \mathcal{L}_E [v] \right\| \leq C_{E,\Delta t} \|u - v\|_{\mathcal{W}_h}$
- *M'-trick: Empirical interpolations exact for larger number of interpolation points $M + M'$ and the initial data is $u_0(\boldsymbol{\mu}) \in \mathrm{span}\Phi$*

A posteriori error estimator

Theorem (A posteriori error estimator cont.)

Assumptions:

- *Operator(s) fulfill "Lipschitz" properties:*
  - $\|u - v + \Delta t \mathcal{L}_I[u] - \Delta t \mathcal{L}_I[v]\| \geq \frac{1}{C_{I,\Delta t}} \|u - v\|_{\mathcal{W}_h}$
  - $\|u - v - \Delta t \mathcal{L}_E[u] + \Delta t \mathcal{L}_E[v]\| \leq C_{E,\Delta t} \|u - v\|_{\mathcal{W}_h}$
- *M'-trick: Empirical interpolations exact for larger number of interpolation points $M + M'$ and the initial data is $u_0(\boldsymbol{\mu}) \in \mathrm{span}\Phi$*

Then:

$$\left\| u_{\mathrm{red}}^k(\boldsymbol{\mu}) - u_h^k(\boldsymbol{\mu}) \right\| \leq \eta_{N,M,M'}^k(\boldsymbol{\mu})$$

with

$$\eta_{N,M,M'}^k(\boldsymbol{\mu}) := \sum_{i=0}^{k-1} C_{I,\Delta t}^{k-i+1} C_{E,\Delta t}^{k-i} \left(\left\| R_{I+E,M,M'}^{k+1}(\boldsymbol{\mu}) \right\| + \left\| \Delta t R^{k+1}(\boldsymbol{\mu}) \right\| + \varepsilon^{New} \right)$$

A posteriori error estimator

Theorem (A posteriori error estimator cont.)

Then:

$$\left\| u_{\mathrm{red}}^k(\boldsymbol{\mu}) - u_h^k(\boldsymbol{\mu}) \right\| \leq \eta_{N,M,M'}^k(\boldsymbol{\mu})$$

with

$$\eta_{N,M,M'}^k(\boldsymbol{\mu}) := \sum_{i=0}^{k-1} C_{I,\Delta t}^{k-i+1} C_{E,\Delta t}^{k-i} \left(\left\| R_{I+E,M,M'}^{k+1}(\boldsymbol{\mu}) \right\| + \left\| \Delta t R^{k+1}(\boldsymbol{\mu}) \right\| + \varepsilon^{New} \right)$$

The residuals $R_{*,M}$ measure the empirical interpolation error, e.g.

$$R_{*,M,M'}^{k+1,\nu} := \sum_{m=M}^{M+M'} l_m^* \left[u_{\mathrm{red}}^{k+1,\nu} \right] \xi_m = \mathcal{I}_{M+M'}[\mathcal{L}] - \mathcal{I}_M[\mathcal{L}]$$

A posteriori error estimator

Theorem (A posteriori error estimator cont.)

Then:

$$\left\| u_{\mathrm{red}}^k(\boldsymbol{\mu}) - u_h^k(\boldsymbol{\mu}) \right\| \leq \eta_{N,M,M'}^k(\boldsymbol{\mu})$$

with

$$\eta_{N,M,M'}^k(\boldsymbol{\mu}) := \sum_{i=0}^{k-1} C_{I,\Delta t}^{k-i+1} C_{E,\Delta t}^{k-i} \left(\left\| R_{I+E,M,M'}^{k+1}(\boldsymbol{\mu}) \right\| + \left\| \Delta t R^{k+1}(\boldsymbol{\mu}) \right\| + \varepsilon^{New} \right)$$

empirical interpolation error	
Galerkin projection error	
time evolution	
Newton step error	

A posteriori error estimator

Theorem (A posteriori error estimator cont.)

Then:

$$\left\| u_{\mathrm{red}}^k(\boldsymbol{\mu}) - u_h^k(\boldsymbol{\mu}) \right\| \leq \eta_{N,M,M'}^k(\boldsymbol{\mu})$$

with

$$\eta_{N,M,M'}^k(\boldsymbol{\mu}) := \sum_{i=0}^{k-1} C_{I,\Delta t}^{k-i+1} C_{E,\Delta t}^{k-i} \left(\left\| R_{I+E,M,M'}^{k+1}(\boldsymbol{\mu}) \right\| + \left\| \Delta t R^{k+1}(\boldsymbol{\mu}) \right\| + \varepsilon^{New} \right)$$

empirical interpolation error	(quasi-) rigorous, M' affects efficiency and effectivity
Galerkin projection error	rigorous and efficient
time evolution	selection affects effectivity
Newton step error	neglectable

Effective error bound

- (Compute many interpolation points M'.)
- Define Lipschitz–constants as tight as possible by approximation of $C_{I,\Delta t} = \sup_{k=0,\dots,K} \left\| \left(D\left(\mathrm{Id} + \mathcal{L}_{I,\Delta t}\right)|_{u_h^k}\right)^{-1} \right\|$, e.g. as

Computation of Lipschitz constants

- Operator–norm approximation:

$$C_{I,\Delta t} \approx \inf_{v \in X_{\text{test}}} \frac{\left\| D\left(\mathrm{Id} + \mathcal{L}_{I,\Delta t}\right)|_{u_h^k}[v] \right\|}{\|v\|}$$

$$C_{E,\Delta t} \approx \sup_{v \in X_{\text{test}}} \frac{\left\| D\left(\mathrm{Id} + \mathcal{L}_{E,\Delta t}\right)|_{u_h^k}[v] \right\|}{\|v\|},$$

where X_{text} is a set of intermediate Newton step solutions.

- $C_{I,low} \approx C_I(\boldsymbol{\mu})$, $C_{E,upper} \approx C_E(\boldsymbol{\mu})$ for all $\boldsymbol{\mu} \in \mathcal{P}$

Remark: Output of interests

- If application has output of interest: $s(u_h)$ (s linear functional),
- the error bound $\eta^s \geq |s(u_h) - s(u_{\mathrm{red}})|$
- can be improved by solving bounds for adjoint problems.

Elliptic problems

$$A(\boldsymbol{\mu})\,[u_h(\boldsymbol{\mu}), \phi] = b(\phi) \qquad \text{for all } \phi \in \mathcal{W}_h$$
$$A(\boldsymbol{\mu})\,[u_{\mathrm{red}}(\boldsymbol{\mu}), \phi] = b(\phi) \qquad \text{for all } \phi \in \mathcal{W}_{\mathrm{red}}$$

Elliptic problems

$$A(\boldsymbol{\mu})\left[u_h(\boldsymbol{\mu}), \phi\right] = b(\phi) \qquad \text{for all } \phi \in \mathcal{W}_h$$
$$A(\boldsymbol{\mu})\left[u_{\mathrm{red}}(\boldsymbol{\mu}), \phi\right] = b(\phi) \qquad \text{for all } \phi \in \mathcal{W}_{\mathrm{red}}$$

Error estimator

$$\eta(\boldsymbol{\mu}) := \frac{R(\boldsymbol{\mu})}{\beta(\boldsymbol{\mu})}$$

with

- residual $R(\boldsymbol{\mu}) := \sup_{\phi \in \mathcal{W}_h} A(\boldsymbol{\mu})\left[u_{\mathrm{red}}(\boldsymbol{\mu})\right] - b(\phi)$ and
- coercivity constant $\beta(\boldsymbol{\mu})$.

Effectivity of error estimators

Factors

- Coercivity / Lipschitz constants
- Propagation over time
- Output functional

Solutions from literature:

Method	*strategy*	*cost*
SCM	finds tight coercivity constant	increased basis size / offline time
EV approximation	compute approximation of eigenvalues from reduced matrices	increased basis size / offline time
Space-time grid	tighter control of error propagation over time	rewrite of the numerical scheme

Outline

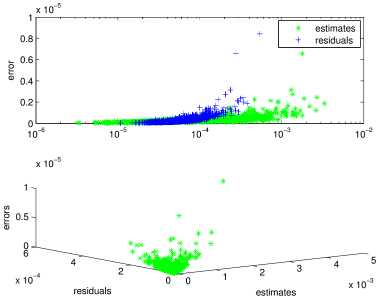
Machine learning solution

Observation

Mapping f from residuals to errors

$$R(\mu) \qquad \to e(\mu)$$

usually behaves very smoothly.



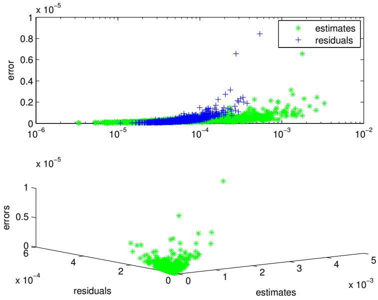
Machine learning solution

Observation

Mapping f from residuals to errors

$$X(\mu) := (R(\mu), I(\mu)) \to e(\mu)$$

usually behaves very smoothly, for indicators $I(\mu)$.



Machine learning solution (cont.'d)

- *Assumption:* Mapping f from residuals to errors behaves randomly.
- Find stochastic process $\bar{f}$ such that

$$\bar{f}(X) \sim \{e(\boldsymbol{\mu}) \mid e(\boldsymbol{\mu}) = f(X) \text{ with } \boldsymbol{\mu} \in \mathcal{P}\}.$$

Machine learning solution (cont.'d)

- *Assumption:* Mapping f from residuals to errors behaves randomly.
- Find stochastic process $\bar{f}$ such that

$$\bar{f}(X) \sim \{e(\boldsymbol{\mu}) \mid e(\boldsymbol{\mu}) = f(X) \text{ with } \boldsymbol{\mu} \in \mathcal{P}\}.$$

- We do not know better: Assume **Gaussian process**

$$\bar{f}(X) \sim \mathcal{N}(m(X), \Sigma).$$

Construction of Gaussian process

1. Compute training samples $(X(\boldsymbol{\mu}), e(\boldsymbol{\mu}))$ for $\boldsymbol{\mu} \in P_{\text{train}} \subset \mathcal{P}$
2. Infer Gaussian process predictions $f(x^*)$ at new points x^* with
   A. GP Kernel methods (Rasmussen&Williams, 2005)
   B. Relevance vector machine (Tipping, 2001)

GP Kernel method

Assumption

$$\mathbf{x} \in \mathbb{R}^N, \quad \mathrm{cov}(f(\mathbf{x})) = K(\mathbf{x}) + \sigma^2 I$$

with **kernel**

$$(K(\mathbf{x}))_{1 \leq n,m \leq N} = \exp\left(\frac{(x_n - x_m)^2}{l^2}\right).$$

GP Kernel method

Assumption

$$\mathbf{x} \in \mathbb{R}^N, \quad \mathrm{cov}(f(\mathbf{x})) = K(\mathbf{x}) + \sigma^2 I$$

with **kernel**

$$(K(\mathbf{x}))_{1 \leq n,m \leq N} = \exp\left(\frac{(x_n - x_m)^2}{l^2}\right).$$

1. Infer **hyper–parameters** σ^2 and l.
2. **Predict** $f(x^*) = \mathcal{N}(m(x^*), \Sigma)$ for test/validation values x^*.

ASSUMPTION

$$f(x) = \sum_{m=1}^{M} w_m \phi_m(x) + \epsilon$$

with fixed basis functions ϕ_m (e.g. polynomials), priors $w \sim \mathcal{N}(0, A)$ and noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$.

Relevance vector machine

Assumption

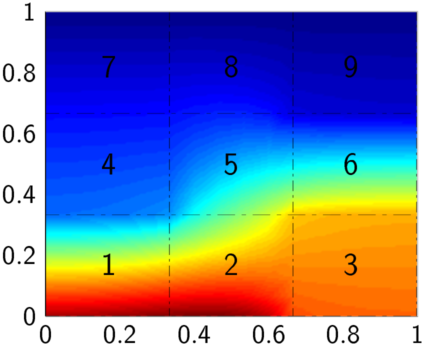
$$f(x) = \sum_{m=1}^{M} w_m \phi_m(x) + \epsilon$$

with fixed basis functions ϕ_m (e.g. polynomials), priors $w \sim \mathcal{N}(0, A)$ and noise $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$.

1. Infer **hyper–parameters** σ^2 and A.
2. **Predict** $f(x^*) = \mathcal{N}(m(x^*), \Sigma)$ for test/validation values x^*.

Outline

Example: A thermal block

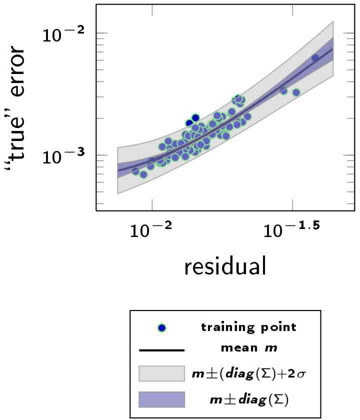


Thermal block

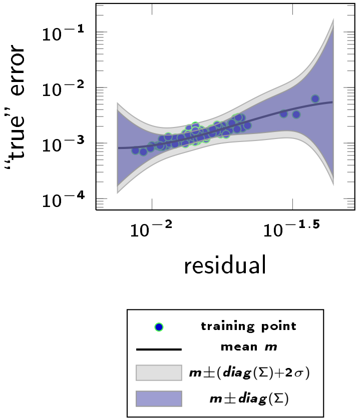
- **Heat equation** $-a(x;\mu)\Delta u = f$
- **with** $a(x;\mu) = \sum_{i=1}^{9} \mu_i \chi_{\Omega_i}(x)$
- **9 parameters**

Visualization of Gaussian processes

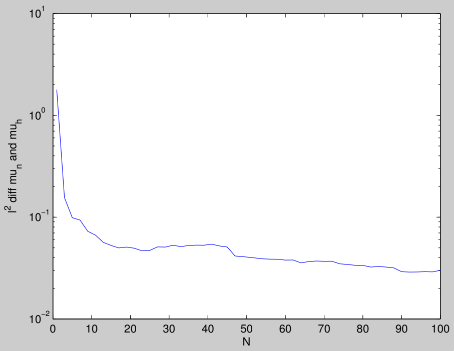
GP-Kernel method

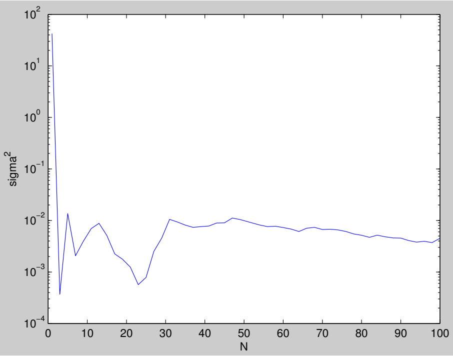


RVM method (polynomial basis)



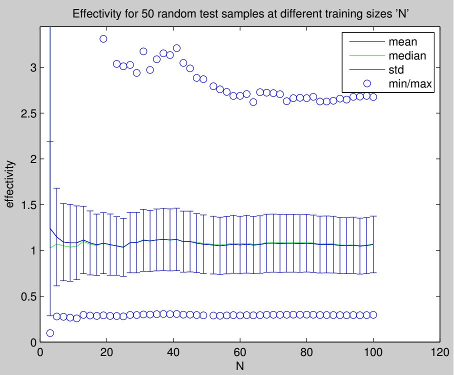
Convergence of $m_N(x)$ and σ_N^2





- Stabilizes after 30–40 training samples.
- The estimator seems to converge.

Effectivity



-	m_{47}	m_{100}	η
mean	1.08	1.07	424.41
median	1.09	1.07	256.41
std	0.32	0.31	433.72
min	0.3	0.3	68.25
max	2.89	2.68	3,550.8

Stochastic error estimates

- are more effective (400x), but
- underestimate the error sometimes.

Outline

1. A posteriori error estimators
   - Reduced basis scheme
   - Ingredients
   - Effectivity of error estimators
2. Stochastic error estimate
3. Numerical experiments
   - Thermalblock-Problem
4. Summary and future work

Summary and Future work

Summary

- **Effective** a posteriori error bounds are **costly**.
- Given a sufficient number of high dimensional computations, **cheap** and **effective** error bounds can be constructed.
- Based on **Bayesian inference**.

Summary and Future work

Summary

- **Effective** a posteriori error bounds are **costly**.
- Given a sufficient number of high dimensional computations, **cheap** and **effective** error bounds can be constructed.
- Based on **Bayesian inference**.

Future work

- Use (non-Gaussian) in order to reduce likelihood for underestimating.
- Non–linear and time dependent examples with more complex indicators.
- Try to improve effectivity with further indicators (e.g. eigenvalues of reduced matrices).

References

[Drohmann et al.,2012] M. Drohmann, B. Haasdonk, and M. Ohlberger.
Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation.
SIAM J. Sci Comp, 34:A937–A969, 2012.

[Drohmann, 2012] M. Drohmann.
Reduced basis approximation for nonlinear evolution equations.
PhD thesis, University of Münster, 2012.

[Tonn, 2012] T. Tonn.
Reduced-Basis Method (RBM) for Non-Affine Elliptic Parametrized PDEs.
PhD thesis, University of Ulm, 2012.

[Chaturantabut, Sorensen, 2010] S. Chaturantabut and D. C. Sorensen.
A state space error estimate for POD–DEIM nonlinear model reduction.
Technical report, CAAM, Rice U., December 2010.
Technical Report TR10-32.

[Haasdonk et al., 2008] B. Haasdonk, M. Ohlberger und G. Rozza.
A Reduced Basis Method for Evolution Schemes with Parameter-Dependent Explicit Operators
ETNA, Electronic Transactions on Numerical Analysis, vol. 32, 2008

[Barrault et al., 2004] M. Barrault, M., Y. Maday, N. Nguyen and A. Patera,
An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations
C. R. Math. Acad. Sci. Paris Series I, 2004, 339, 667-672