

An Adaptive Sampling Approach for the Reduced Basis Method

Sridhar Chellappa, Lihong Feng and Peter Benner

Abstract The offline time of the reduced basis method can be very long given a large training set of parameter samples. This usually happens when the system has more than two independent parameters. On the other hand, if the training set includes fewer parameter samples, the greedy algorithm might produce a ROM with large errors at the samples outside of the training set. We introduce a method based on a surrogate error model to efficiently sample the parameter domain such that the training set is adaptively updated starting from a coarse set with a small number of parameter samples. A sharp a posteriori error estimator is evaluated on a coarse training set. Radial Basis Functions are used to interpolate the error estimator over a separate fine training set. Points from the fine training set are added into the coarse training set at every iteration based on a user defined criterion. In parallel, parameter samples satisfying a defined tolerance are adaptively removed from the coarse training set. The approach is shown to avoid high computational costs by using a compact training set and to provide a reduced-order model with guaranteed accuracy over the entire parameter domain.

Sridhar Chellappa
Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany. e-mail: chellappa@mpi-magdeburg.mpg.de

Lihong Feng
Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany. e-mail: feng@mpi-magdeburg.mpg.de

Peter Benner
Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany and
Fakultät für Mathematik, Otto-von-Guericke Universität Magdeburg, Germany. e-mail: benner@mpi-magdeburg.mpg.de

1 Introduction

The need of modeling physical processes accurately very often leads to mathematical models of large-scale. Simulating such large-scale systems poses challenges on computer memory and computational power. Model Order Reduction (MOR) aims to speedup simulation of the large-scale full order model (FOM) by constructing a small-scale system, called the reduced-order model (ROM), which serves as a good approximation of the FOM.

For FOMs arising from discretization of parametric partial differential equations (PDEs), the Reduced Basis (RB) method is a popular choice [16]. It is suitable especially when the FOM needs to be solved repeatedly for a range of parameters. The RB method involves two stages. The first is an expensive *offline stage* in which a subspace of the solution manifold is constructed by simulations of the FOM at certain chosen parameters. Then, using Galerkin projection, the solution to the PDE is projected onto this subspace to generate the ROM. In the second *online stage*, instead of solving the FOM, one solves the ROM for any parameter of interest. The RB method works well when the solution manifold is supposed to have a small Kolmogorov n -width. In this work, we focus on the offline stage of the RB method.

The offline stage in the RB method requires the construction of a representative set of parameters from the domain of interest. This is called the *training set*. The subspace approximating the solution manifold is constructed by solving the FOM at a set of selected points from this training set, picked using a greedy algorithm [11]. The choice of a training set is nontrivial. On the one hand, if it includes too few parameters, the original solution manifold may not be adequately represented, leading to a poor ROM with large error. On the other hand, if it is too fine, the offline time can be prohibitively long. When the PDE involves several parameters, properly defining the training set can be a severe computational issue.

Several authors have attempted to address this issue. The earliest work to consider an adaptive sampling of the training set was [18], where the author proposes a multi-stage method. The greedy algorithm is run several times over randomly sampled small training sets to generate the RB basis. Then, the ROM is tested over a much larger training set and the greedy algorithm is re-run only on those points failing the tolerance criterion. The authors of [10] address the issue of large training sets by two approaches. The first one is a procedure to monitor the error over an additional validation parameter set. If a large error is detected, then the training set is further refined, either uniformly or locally. The second approach, similar to the one presented in [6], is based on partitioning the parameter domain adaptively, and generating a local basis for each partition. Other approaches for adaptive sampling are proposed in [12] and [14]. More recently, an approach based on Kriging interpolation and clustering is proposed in [15], to tackle the problem of high-dimensional parameter spaces. An interpolant of the residual norm is calculated over a fine grid of parameters. Then, k-means clustering is used to identify parameters that have high probability of presenting larger errors.

The approach we propose here is based on interpolating a sharp *a posteriori* output error estimator originally proposed in [5]. The RB method is initialized with a coarse

training set. We update this set progressively by adding or removing points from it. At each iteration of the greedy algorithm, we estimate the error of the ROM at every parameter in the coarse training set. Note that the evaluation of the estimator does not need to evaluate the FOM. In order to further reduce the computational costs, we then interpolate the estimated error over a fine training set and use the interpolant as the error surrogate, which will replace the error estimator for estimating the ROM error over the fine training set. To achieve this, a surrogate model based on Radial Basis Functions (RBFs) is used.

At each iteration of the greedy algorithm, the ROM errors at the parameters in the fine training set are checked by the error surrogate. Those parameters corresponding to large values of the error surrogate are selected and added into the coarse training set. The error surrogate is much cheaper to compute than the error estimator. Therefore, using the former instead of the latter to check the ROM error over the fine training set, will reduce the computational cost. Additionally, if any parameter in the coarse training set achieves the required ROM accuracy, we remove it from the coarse training set. Such an approach is able to construct a compact, representative training set by fully exploring the parameter domain with reduced computational cost. Forming the RBF interpolant is a relatively cheap operation with the most expensive contribution coming from the need to solve an $\ell \times \ell$ linear system of equations at each parameter sample, where ℓ is the cardinality of the coarse training set, which is expected to be small. The algorithm proposed in our work is based on the use of the RBF interpolant and offers certain advantages over the most relevant methods in [12] and [15]. Our proposed approach differs in the following aspects, when compared to [12]:

- We use a primal-dual a posteriori error estimator on a coarse training set and a cheaply computable error surrogate on the fine training set. No cheap error estimator is used on the fine training set in [12], potentially leading to a larger computational effort.
- At each iteration of the greedy algorithm, a saturation assumption is introduced in [12] in order to avoid calculation of the ROM at certain parameters. However, this requires estimation of a saturation constant which needs to be defined a priori by the user. In our method, no such constant needs to be estimated.

Furthermore, when compared to [15],

- We consider both addition and removal of parameter samples from the coarse training set. The method in [15] is restricted to adding new samples only.
- The use of Kriging interpolation involves the estimation of several hyper-parameters. This can be a computational bottleneck [19]. In the case of RBF, there is only one free parameter at most. If using some special kernel functions, e.g. polyharmonic spline kernels, there are no free parameters to tune.
- Finally, a nonlinear model was considered in [15] in order to demonstrate the adaptive sampling approach. An offline hyper-reduction step was used to reduce the complexity of the online nonlinearity evaluations, where a second training set for the nonlinear function is used in the hyper-reduction step, and is likely to

be separately given and fixed. In our approach, we employ the Adaptive POD-Greedy-(D)EIM algorithm proposed in [5]. This avoids the need for a potentially separate training set for the hyper-reduction phase. Instead, the (D)EIM basis generation is carried out as a part of the greedy loop. In this way, we propose a fully adaptive approach in Algorithm 4.

The rest of the paper is organized as follows. In Section 2, the basic idea behind the RB method is reviewed. An adaptive parameter sampling approach using the RBF interpolation is proposed and elaborated in Section 3. There, we begin with a brief introduction to the theory of RBFs. Then an adaptive sampling approach for the standard POD-Greedy algorithm is proposed to show a general framework for the proposed technique. In Subsection 3.3, a fully adaptive algorithm is proposed, which integrates the adaptive sampling technique with an adaptive basis construction method: the adaptive POD-Greedy-(D)EIM algorithm for general parametric nonlinear/nonaffine systems. A strategy of computing the shape parameter for RBF interpolation is detailed in Subsection 3.4. In Section 4, we show numerical results for a series of benchmark examples, validating the proposed algorithms. We conclude in Section 5 by summarizing our results and suggesting future research directions.

2 Reduced Basis Method

For the sake of completeness, we briefly highlight the key idea of the RB method. Consider a nonlinear parametric dynamical system arising from space-time discretization of a model of PDEs,

$$\begin{aligned} \mathbf{E}(\mu)x^{k+1}(\mu) &= \mathbf{A}(\mu)x^k(\mu) + f(x^k(\mu); \mu) + \mathbf{B}(\mu)u^k, \\ y^{k+1}(\mu) &= \mathbf{C}(\mu)x^{k+1}(\mu) \end{aligned} \quad (1)$$

where $x^k(\mu) \in \mathbb{R}^n$ is the state vector at the k -th time step t^k in the time interval $[0, T]$ divided as $t^0 < t^1 < \dots < t^k < \dots < t^K = T$. $\mathbf{E}(\mu)$, $\mathbf{A}(\mu) \in \mathbb{R}^{n \times n}$ are the parameter-dependent system matrices. $f(x^k(\mu); \mu) \in \mathbb{R}^n$ depicts the state, parameter dependent system nonlinearity. $\mathbf{B}(\mu) \in \mathbb{R}^{n \times q}$, $\mathbf{C}(\mu) \in \mathbb{R}^{m \times n}$ are the input and output matrices, respectively, while u^k denotes the input at the k -th time step and $y^{k+1}(\mu)$ denotes the quantities-of-interest. The vector of parameters $\mu \in \mathbb{R}^d$ belongs to a parameter domain $\mathcal{P} \subset \mathbb{R}^d$. The space discretization can be done through any method of choice such as Finite Element Method (FEM), Finite Volume Method (FVM) or the Finite Difference Method (FDM). For the time discretization, we adopt a semi-implicit scheme for ease of using the a posteriori error estimator in [5]. The RB method attempts to generate a subspace \mathcal{X} with dimension $r \ll n$, that best approximates the solution manifold. Let $\mathbf{V} \subset \mathbb{R}^{n \times r}$ be an orthogonal basis of \mathcal{X} . The ROM is then given via Galerkin projection,

$$\begin{aligned}\mathbf{E}_r(\mu)x_r^{k+1}(\mu) &= \mathbf{A}_r(\mu)x_r^k(\mu) + f_r(x_r^k(\mu); \mu) + \mathbf{B}_r(\mu)u^k, \\ y_r^{k+1}(\mu) &= \mathbf{C}_r(\mu)x_r^{k+1}(\mu),\end{aligned}\quad (2)$$

where $x_r^k \in \mathbb{R}^r$ is the reduced state vector at the k th time step. $\mathbf{E}_r(\mu) := \mathbf{V}^T \mathbf{E}(\mu) \mathbf{V} \in \mathbb{R}^{r \times r}$, $\mathbf{A}_r(\mu) := \mathbf{V}^T \mathbf{A}(\mu) \mathbf{V} \in \mathbb{R}^{r \times r}$ are the reduced system matrices. $\mathbf{B}_r(\mu) := \mathbf{V}^T \mathbf{B}(\mu) \in \mathbb{R}^{r \times q}$, $\mathbf{C}_r(\mu) := \mathbf{C}(\mu) \mathbf{V} \in \mathbb{R}^{m \times r}$ are the reduced input and output matrices, respectively. The reduced nonlinear term is given by $f_r(\mathbf{V}x_r^k(\mu); \mu) := \mathbf{V}^T f(\mathbf{V}x_r^k(\mu); \mu) \in \mathbb{R}^r$. An important assumption is that the system matrices and the input, output matrices have affine decomposition, i.e., any of them can be formulated as a sum of products of parameter dependent and parameter independent components as shown below,

$$\begin{aligned}\mathbf{E}(\mu) &= \sum_{i=1}^{Q_E} \theta_i^E(\mu) \mathbf{E}_i, & \mathbf{A}(\mu) &= \sum_{i=1}^{Q_A} \theta_i^A(\mu) \mathbf{A}_i, \\ \mathbf{B}(\mu) &= \sum_{i=1}^{Q_B} \theta_i^B(\mu) \mathbf{B}_i, & \mathbf{C}(\mu) &= \sum_{i=1}^{Q_C} \theta_i^C(\mu) \mathbf{C}_i.\end{aligned}$$

In the offline stage, the expensive parameter independent terms are evaluated once and stored for repeated use in the online stage.

The standard process of the RB method for time-dependent parametric systems is presented in Algorithm 1 which is known as the POD-Greedy algorithm. It constructs the projection matrix \mathbf{V} by selecting a set of samples $\{\mu_i\}_{i=1}^r$ from a given parameter training set, through a greedy algorithm. The FOM solutions $x^k(\mu)$, $k = 1, \dots, K$, at a newly selected sample μ_i are assembled into a matrix $\mathbf{X} := [x^1(\mu_i), \dots, x^K(\mu_i)]$. The matrix \mathbf{V} is iteratively enriched by projecting the solution vectors in \mathbf{X} onto the current subspace $\text{range}(\mathbf{V})$, see Steps 3 - 5 in Algorithm 1. Crucial to the success of the RB method is the *a posteriori* error estimator,

$$\Delta^{k+1}(\mu) \geq \|y^{k+1}(\mu) - y_r^{k+1}(\mu)\|.$$

It quantifies the error between the FOM and ROM for each parameter, without having to solve the FOM. The parameters at which the FOM is to be solved to enrich \mathbf{V} , is picked iteratively by maximizing the error estimator over the training set. Although Eq. (2) is of reduced order, the complexity of evaluating the nonlinear term $f_r(\mathbf{V}x_r^k(\mu); \mu)$ remains of order $\mathcal{O}(n)$. This is the so-called *lifting bottleneck*. Techniques like Empirical Interpolation Method (EIM) or its discrete variation, the Discrete Empirical Interpolation Method (DEIM), can be used to obtain a ROM whose computational complexity is independent of the full order n . For more details on this approach, we refer the reader to [2, 4]. In the following section we discuss the idea of adaptive parameter sampling.

3 Adaptive Parameter Sampling

The quality of the ROM depends on how well the training set represents the parameter domain. In the standard approach, if a finely sampled training set is used for

a parameter domain with high-dimension, it can be computationally expensive. If otherwise, a coarse training set is used, it will not guarantee that a reliable ROM can be derived. Consequently, one faces the situation of trial and error. The approach we propose here is based on a surrogate error model generated through RBF interpolation. It aims to overcome the drawbacks of the standard approach using a fixed training set.

3.1 Radial Basis Functions

RBF interpolation is an efficient way to approximate scattered data in high-dimensional space. Consider the domain $\Omega \subseteq \mathbb{R}^d$ and the function $h := \mathbb{R}^d \rightarrow \mathbb{R}$. Assume that h is difficult to evaluate, or known only at a few points in Ω . Let Φ be the *kernel function* having the property that $\Phi(\mu_1, \mu_2) = \Phi(\|\mu_1 - \mu_2\|)$, $\forall \mu_1, \mu_2 \in \Omega$. Typically, the Euclidean norm is used. Such kernels have the name of *radial basis functions* since they are radially symmetric. We seek an approximation $s : \mathbb{R}^d \rightarrow \mathbb{R}$ to the function h given as

$$s(\mu) := \sum_{i=1}^{\ell} c_i \Phi(\|\mu - \mu_i\|), \quad \forall \mu \in \Omega. \quad (3)$$

The coefficients $\{c_i\}_{i=1}^{\ell}$ are determined by enforcing the interpolation condition $h(\mu_i) = s(\mu_i)$, $i = 1, 2, \dots, \ell$. This amounts to solving the linear system of equations

$$\underbrace{\begin{bmatrix} \Phi(\mu_1, \mu_1) & \Phi(\mu_1, \mu_2) & \cdots & \Phi(\mu_1, \mu_\ell) \\ \Phi(\mu_2, \mu_1) & \Phi(\mu_2, \mu_2) & \cdots & \Phi(\mu_2, \mu_\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi(\mu_\ell, \mu_1) & \Phi(\mu_\ell, \mu_2) & \cdots & \Phi(\mu_\ell, \mu_\ell) \end{bmatrix}}_{\mathbf{R}} \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_\ell \end{bmatrix}}_c = \underbrace{\begin{bmatrix} h(\mu_1) \\ h(\mu_2) \\ \vdots \\ h(\mu_\ell) \end{bmatrix}}_d \quad (4)$$

The points $\{\mu_i\}_{i=1}^{\ell} \in \Omega$ are called the *centers* of the radial basis functions. Under the assumption that the centers are pairwise distinct, it can be proven that the RBF kernel matrix \mathbf{R} is positive definite for some suitable choice of radial basis functions Φ and thus Eq. (4) has a unique solution. We refer to [20, Chapter 6] for the proof and detailed treatment. Table 1 provides a list of commonly used radial basis functions. All of the radial basis functions in Table 1 have global support. There exist also radial basis functions with local support. For more details we refer the reader to [3].

In practice, the class of radial basis functions that have the positive definite property (\mathbf{R} being positive definite) is limited to a few, such as Gaussian and Inverse multiquadric. The class of admissible basis functions ($\Phi(\cdot)$) can be expanded by defining the so called *conditionally positive definite* functions, by which the positive definiteness can be satisfied by imposing some additional constraints given as,

Kernel	$\Phi(r) := \Phi(\ \mu - \mu_i\)$
Gaussian	$e^{-\sigma^2 r^2}$
Thin-plate splines	$r^2 \ln(r)$
Multiquadric	$\sqrt{r^2 + \sigma^2}$
Inverse multiquadric	$\frac{1}{\sqrt{r^2 + \sigma^2}}$

Table 1: Common radial basis kernels.

$$\sum_{j=1}^M c_j p_j(\mu) = 0, \quad i = 1, 2, \dots, \ell.$$

The functions p_1, p_2, \dots, p_M are a basis of the polynomial space with suitable degree. In practice, we choose M to be equal to the number of scalar parameters d . With the new conditions imposed, the radial basis interpolant now becomes,

$$s(\mu) := \sum_{i=1}^{\ell} c_i \Phi(\|\mu - \mu_i\|) + \sum_{j=1}^M \lambda_j p_j(\mu). \quad (5)$$

We then obtain a saddle-point system of dimension $N_{\text{RBF}} := (M + \ell) \times (M + \ell)$:

$$\begin{bmatrix} \mathbf{R} & \mathbf{P} \\ \mathbf{P}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} \quad (6)$$

With a proper choice of p_1, p_2, \dots, p_M , the augmented coefficient matrix is positive definite for all choices of radial basis functions in Table 1 and this ensures the uniqueness of the interpolant. For a detailed discussion of the rationale behind the idea of conditionally positive definite functions, we refer to [20, Chapter 6].

3.2 POD-Greedy algorithm with adaptive sampling

Now we detail the proposed approach of adaptively constructing the training set for the RB method. The proposed adaptive parameter sampling scheme is general and can be combined with the standard greedy algorithm for steady parametric systems, or with the POD-greedy algorithm (Algorithm 1) for time-dependent systems, to adaptively update the training set and reduced the offline costs. Since the framework is similar, we only present the adaptive parameter sampling approach with POD-Greedy for time-dependent problems. The standard POD-Greedy algorithm is presented in Algorithm 1. POD-greedy with adaptive sampling is presented as Algorithm 2, where it can be seen that the training set (Ξ_c) is updated at each iteration.

Algorithm 2 is initialized using a coarse training set Ξ_c , with cardinality N_c . At the end of each iteration, we compute the error estimator *only* over the *coarse* training set

Algorithm 1 Standard POD-Greedy**Input:** Training set Ξ , Tolerance (tol).**Output:** \mathbf{V}

- 1: Initialize. $\mathbf{V} = [], \mu^* \in \Xi$.
- 2: **while** $\Delta(\mu^*) > \text{tol}$ **do**
- 3: Compute full order solution at μ^* . $\mathbf{X} = [x(t^1, \mu^*), x(t^2, \mu^*), \dots, x(t^K, \mu^*)]$.
- 4: Form $\tilde{\mathbf{X}} := \mathbf{X} - \mathbf{V}\mathbf{V}^T\mathbf{X}$; Obtain $\tilde{\mathbf{X}} \xrightarrow{\text{svd}} \mathbf{U}\Sigma\mathbf{W}^T$.
- 5: Update $\mathbf{V} := \text{orth}[\mathbf{V}, \mathbf{U}(:, 1)]$.
- 6: Solve the ROM and compute the error estimator $\Delta(\mu)$, $\forall \mu \in \Xi$.
- 7: Next μ^* is chosen as, $\mu^* := \arg \max_{\mu \in \Xi} \Delta(\mu)$.
- 8: **end while**

Ξ_c , i.e., $\Delta(\mu)$, $\forall \mu \in \Xi_c$ (compare Step 6 in Algorithm 1 with Step 4 in Algorithm 2). We use this as the *input data* to form the radial basis interpolant over the fine training set, i.e., $s(\mu)$, $\forall \mu \in \Xi_f$, where Ξ_f is a finely sampled training set of cardinality $N_f \gg N_c$. The evaluation costs are $O((N_c + M)^3)$ for identifying the interpolant coefficients, and $O(N_f N_c)$ for interpolating over the fine training set. Here M is the number of polynomial functions added to make the augmented coefficient matrix in (Eq. (6)) conditionally positive definite. Since N_c, M are both small numbers, the cost of evaluating $s(\mu)$ is much smaller than the cost of computing $\Delta(\mu)$ which includes solving the ROM for each μ . We enrich the coarse training set by adding new parameters from Ξ_f to Ξ_c . We add those n_{add} new parameters that have the largest magnitude of $s(\mu)$, where n_{add} can be fixed and user-defined. One heuristic

way of varying n_{add} adaptively is $n_{\text{add}} = \log_{10} \left(\left\lceil \frac{\max_{\mu \in \Xi_c} \Delta(\mu)}{\text{tolerance}} \right\rceil \right)$. Additionally, we

also monitor the coarse training set to identify those points $\check{\mu} \in \Xi_c$ with $\Delta(\check{\mu}) < \text{tol}$. Those points are then removed from Ξ_c , meaning that their corresponding full solutions need not be computed to enrich the RB basis. In this way, the coarse training set gets updated iteratively and its size remains as compact as possible, avoiding many unnecessary full simulations of the FOM.

Remark 1 (Interpolating small values) The magnitudes of the errors we interpolate are often very small, especially at the latter iterations of the greedy algorithm. To ensure good, stable interpolation, we consider the input data to the RBF as the base 10 logarithm of the estimated errors. After interpolation, we perform the anti-logarithm and project the logarithm value back to the actual error value.

3.3 A fully adaptive POD-Greedy-(D)EIM algorithm

An adaptive version of Algorithm 1, called the Adaptive POD-Greedy-(D)EIM algorithm, is proposed in [5] for nonlinear systems, which aims to update the RB, (D)EIM basis with an adaptively chosen number of basis vectors at each iteration of the greedy algorithm. In case of a nonlinear or nonaffine system, the standard POD-

Algorithm 2 POD-Greedy with adaptive parameter sampling**Input:** Coarse training set Ξ_c , Fine training set Ξ_f , Tolerance (tol).**Output:** \mathbf{V} .

- 1: Initialize. $\mathbf{V} = []$, $\mu^* \in \Xi_c$.
- 2: **while** $\Delta(\mu^*) > \text{tol}$ **do**
- 3: Perform Steps 3-5 in Algorithm 1.
- 4: Solve the ROM and compute the error estimator $\Delta(\mu)$, $\forall \mu \in \Xi_c$.
- 5: Form RBF interpolant $s(\mu)$ of $\Delta(\mu)$ over Ξ_f .
- 6: Calculate n_{add} , pick $\{\mu_1, \dots, \mu_{\text{add}}\}$ from Ξ_f with largest errors measured by $s(\mu)$.
- 7: Update the training set. $\Xi_c := [\Xi_c \cup \{\mu_1, \dots, \mu_{\text{add}}\}]$ (Add samples to the current training set Ξ_c). Find all $\tilde{\mu}_i$, $\forall i = 1, \dots, n_{\text{del}}$ with $\Delta(\tilde{\mu}_i) < \text{tol}$. Set $\Xi_c := \Xi_c \setminus \{\tilde{\mu}_i, \forall i = 1, \dots, n_{\text{del}}\}$ (Remove unnecessary samples from the current training set Ξ_c).
- 8: Next μ^* is chosen as, $\mu^* := \arg \max_{\mu \in \Xi_c} \Delta(\mu)$.
- 9: **end while**

Algorithm 3 Adaptive POD-Greedy-(D)EIM algorithm**Input:** Training set Ξ , Tolerance (tol).**Output:** \mathbf{V} , DEIM matrices $\mathbf{U}_f, \mathbf{P}_f$.

- 1: Initialize. $\mathbf{V} = []$, $\mu^* \in \Xi$, $\delta_{\text{RB}} = 1$, $\ell_{\text{DEIM}} = 1$, $\delta_{\text{DEIM}} = 0$, $\mathbf{U}_f = []$, $\mathbf{P}_f = []$, $\mathbf{F} = []$.
- 2: **while** $\Delta(\mu^*) > \text{tol}$ **do**
- 3: Compute full order solution at μ^* . Form the snapshot matrices:
 $\mathbf{X} = [x(t^1, \mu^*), x(t^2, \mu^*), \dots, x(t^K, \mu^*)]$ and
 $\mathbf{F}^{\text{new}} = [f(x(t^1, \mu^*)), f(x(t^2, \mu^*)), \dots, f(x(t^K, \mu^*))]$.
- 4: Form $\tilde{\mathbf{X}} := \mathbf{X} - \mathbf{V}\mathbf{V}^T\mathbf{X}$; obtain $\tilde{\mathbf{X}} \xrightarrow{\text{svd}} \mathbf{U}\Sigma\mathbf{W}^T$.
- 5: Update $\mathbf{V} := \text{orth}([\mathbf{V}, \mathbf{U}(:, 1 : \delta_{\text{RB}})])$ if $\delta_{\text{RB}} \geq 0$; remove δ_{RB} vectors from \mathbf{V} if $\delta_{\text{RB}} < 0$.
- 6: Compute ℓ_{DEIM} (D)EIM interpolation basis vectors from snapshots of the nonlinear function,
 $\mathbf{F} := [\mathbf{F}, \mathbf{F}^{\text{new}}]$.
- 7: Solve ROM and compute the error estimator $\Delta(\mu)$, $\forall \mu \in \Xi$.
- 8: Based on $\Delta(\mu)$, decide the new number δ_{RB} and δ_{DEIM} using the adaptive scheme in [5].
 $\ell_{\text{DEIM}} = \ell_{\text{DEIM}} + \delta_{\text{DEIM}}$.
- 9: Next μ^* is chosen as, $\mu^* := \arg \max_{\mu \in \Xi} \Delta(\mu)$.
- 10: **end while**

Greedy algorithm usually precomputes the (D)EIM basis and the interpolation index matrix, \mathbf{U}_f and \mathbf{P}_f . Unlike the standard POD-Greedy algorithm (Algorithm 1), the Adaptive POD-Greedy algorithm (Algorithm 3) updates the (D)EIM matrix inside the greedy algorithm. This avoids many additional FOM simulations caused by the separate (D)EIM computation outside the POD-Greedy loop. To ensure a reliable ROM, an efficient a posteriori error estimator ($\Delta(\mu)$) is used. The algorithm updates the RB vectors and (D)EIM basis vectors at every iteration (Steps 5-8 in Algorithm 3). The number of RB basis vectors, δ_{RB} , and that of (D)EIM basis vectors, δ_{DEIM} , to be added to/removed from the current basis are computed adaptively, see Step 8 in Algorithm 3. For details of computing δ_{RB} and δ_{DEIM} , we refer the reader to [5]. We note that for this algorithm, the training set is fixed.

With the proposed adaptive sampling technique, we present a fully adaptive POD-Greedy-(D)EIM algorithm: Algorithm 4. It is a combination of Algorithm 3 with the adaptive sampling approach as presented in Algorithm 2. Since the adaptive

Algorithm 4 Adaptive POD-Greedy-(D)EIM with adaptive parameter sampling

Input: Coarse training set Ξ_c , Fine training set Ξ_f , Tolerance (tol).

Output: \mathbf{V} , DEIM basis $\mathbf{U}_f, \mathbf{P}_f$.

- 1: Initialize. $\mathbf{V} = [], \mu^* \in \Xi_c, \ell_{\text{RB}} = 1, \ell_{\text{DEIM}} = 1, \delta_{\text{DEIM}} = 0, \mathbf{U}_f = [], \mathbf{P}_f = []$.
- 2: **while** $\Delta(\mu^*) > \text{tol}$ **do**
- 3: Perform Steps 3-6 in Algorithm 3.
- 4: Solve ROM and compute the error estimator $\Delta(\mu), \forall \mu \in \Xi_c$.
- 5: Based on $\Delta(\mu)$, decide on the new number δ_{RB} and δ_{DEIM} using the adaptive scheme in [5].
 $\ell_{\text{DEIM}} = \ell_{\text{DEIM}} + \delta_{\text{DEIM}}$.
- 6: Perform Steps 5-8 in Algorithm 2
- 7: **end while**

sampling scheme is the same as in Algorithm 2, we omit the detailed explanation for Algorithm 4. One only needs to keep in mind that the error estimator is computed *only* over the *coarse* training set, see Step 4. To estimate the errors at samples in the fine training set, the error surrogate $s(\mu)$ is computed and used, see Step 6.

3.4 A strategy for shape parameter selection

In [15], the authors propose a technique to adaptively sample parameters based on Kriging interpolation, where several hyper-parameters need to be estimated. As an alternative, we use RBF to interpolate an efficient output error estimator. Based on the choice of the kernel used, RBF involves at most one free parameter the user has to specify. This is the case for Gaussian or Multiquadric kernels, where the shape parameter σ needs to be determined. For the case of Thin-plate spline kernels, there are no free parameters to choose. A good choice of the shape parameter is essential to ensure that the RBF kernel matrix (\mathbf{R}) remains well-conditioned [7]. In [17], a heuristic approach is introduced to determine the optimal shape parameter based on the idea of the Leave One Out Cross-Validation (LOOCV) strategy commonly used in the field of statistics [13]. The main idea of LOOCV is to utilize the available data and find a σ that best fits the data.

Below, we briefly describe how we use it to estimate the shape parameter. For more details on the method and its generalization - the k-fold cross validation, we refer to [13].

- (i) We start with the available ℓ centers and data points $(\mu_i, \Delta(\mu_i)), \forall i = 1, \dots, \ell$ for the RBF interpolation.
- (ii) For every $\mu_i, \forall i = 1, \dots, \ell$, compute a ‘less accurate’ radial basis interpolant $s^{\mu_i}(\mu)$ by removing the i^{th} row and i^{th} column of the RBF kernel matrix \mathbf{R} and the i^{th} row of the input data vector \mathfrak{d} . Note that $s^{\mu_i}(\mu_i)$ is actually an approximation of $h(\mu_i)$, whereas, $s(\mu_i) = h(\mu_i)$ (recall (Eq. (4))).
- (iii) Following this, we have the error between $h(\mu_i)$ and $s^{\mu_i}(\mu_i)$: $e^{\mu_i} := h(\mu_i) - s^{\mu_i}(\mu_i)$. Since $h(\mu_i) = s(\mu_i)$, we have

$$e^{\mu_i} = s(\mu_i) - s^{\mu_i}(\mu_i), \quad i = 1, \dots, \ell.$$

- (iv) Form the error vector $\mathbf{e} = [e^{\mu_1}, e^{\mu_2}, \dots, e^{\mu_\ell}]$.
- (v) Choose the optimal σ as,

$$\sigma^* = \arg \min_{\sigma \in [\sigma_{\min}, \sigma_{\max}]} \|\mathbf{e}(\sigma)\|_2.$$

We solve the minimization problem using the MATLAB[®] function `fminbnd`, as suggested in [7].

For the success of the adaptive sampling approach, the following criteria are crucial:

1. *Good data*: The input data to the RBF should be ‘good’. This means that we are interpolating something meaningful to the problem we are trying to approximate.
2. *Good shape parameter*: The RBF interpolation should be robust, ensuring that the interpolated values can act as ‘good’ surrogates of the actual values.

We ensure the first criterion by using an efficient output error estimator for the RB method from [5]. This is a residual based estimator suitable for nonlinear dynamical systems and tailored for output error estimation. Meanwhile, the second criterion is ensured by adopting the LOOCV strategy described above.

4 Numerical results

In this section, we validate the proposed adaptive parameter sampling method using three examples. The first is a nonlinear Burgers’ equation model with one parameter. The second is a two-parameter linear convection-diffusion model. The last one is a three-parameter model of a microthruster unit. All numerical tests were performed in MATLAB[®]2015a, on a laptop with Intel[®]Core[™]i5-7200U @ 2.5 GHZ, with 8 GB of RAM. For all the examples,

- The Adaptive POD-Greedy-(D)EIM algorithm proposed in [5] is used for basis construction. Therefore, we compare the results of Algorithm 3 with fixed training set with those of Algorithm 4 with adaptive training set.
- The maximal output error over the test set Ξ_{test} is defined as

$$\varepsilon_{\max} := \max_{\mu \in \Xi_{\text{test}}} \left(\frac{1}{K} \sum_{k=1}^K \|y^k(\mu) - \bar{y}^k(\mu)\| \right). \quad (7)$$

- The mean (over time steps) output error over Ξ_{test} is defined as

$$\varepsilon(\mu) := \left(\frac{1}{K} \sum_{k=1}^K \|y^k(\mu_i) - \bar{y}^k(\mu_i)\| \right), \quad \forall \mu \in \Xi_{\text{test}}. \quad (8)$$

- The semi-implicit Euler scheme is used for time discretization.
- If not particularly pointed out, the *initial* coarse training set Ξ_c for Algorithm 4 is the same as the fixed training set Ξ for Algorithm 3.

4.1 Burgers' equation

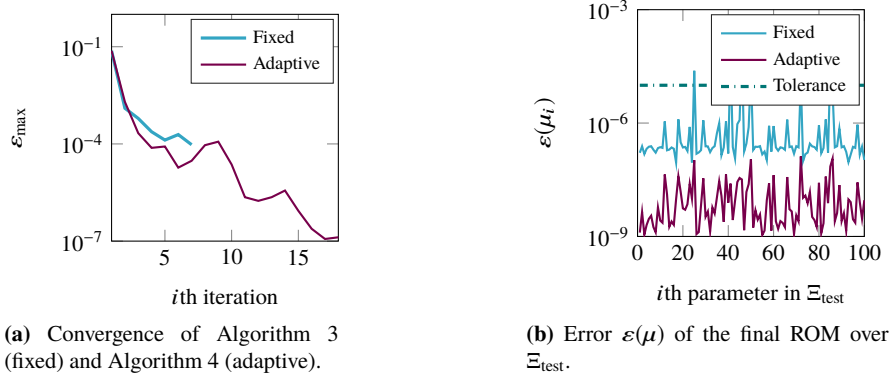
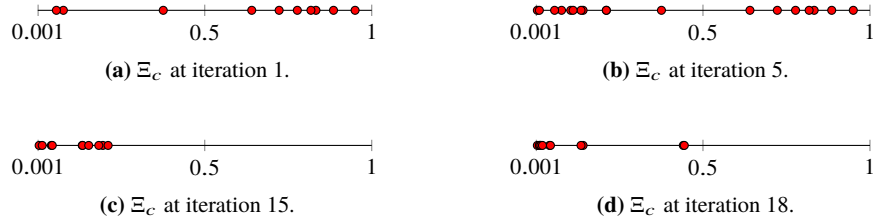
We consider the 1-D model of the viscous Burgers' equation. The PDE is defined in the spatial domain $w \in \Omega := [0, 1]$ and for time $T := [0, 2]$ as

$$\begin{aligned} \frac{dv}{dt} + v \frac{\partial v}{\partial w} &= q \frac{\partial^2 v}{\partial w^2} + s(w, t), \\ v(0, t) &= 0, \quad \frac{\partial v(1, t)}{\partial w} = 0, \end{aligned}$$

where $v(w, t)$ is the unknown variable and $q \in \mathcal{P} := [0.001, 1]$ is the viscosity parameter. The initial condition is set as zero and a constant input to the system is $s(w, t) \equiv 1$. The output is monitored on the last spatial point in the domain $y = v(1, t)$. The model has $N = 500$ equations after discretization in space.

We show the results of Algorithm 3 and Algorithm 4, respectively. The tolerance is set as $\text{tol} = 10^{-5}$. To implement the RB method without adaptive parameter sampling, i.e. Algorithm 3, we choose an initial training set consisting of 10 random parameter samples, using the `rng` command in MATLAB[®], with the seed value 112 for the random number generator `twister`. The fine training set Ξ_f for Algorithm 4 consists of 300 random samples from the same parameter domain, sampled using the seed value at 114 for the random number generator `twister`. Finally, 100 different random samples are considered for the test parameter set Ξ_{test} , sampled using `simdTwister` with the seed value 200. To construct the error surrogate model $s(\mu)$ in Algorithm 4, the Inverse Multiquadric (IMQ) kernel function is used for the RBF interpolation. Cross validation LOOCV has been applied to specify the shape parameter. Since the model is nonlinear, DEIM [4] is used in order to efficiently compute the nonlinear term.

In Fig. 1a shows the decay of the maximal output error, i.e. ε_{\max} , over the test set Ξ_{test} . It is calculated at every iteration of either Algorithm 3 or Algorithm 4. Evidently, using an adaptive training set leads to a better convergence. Algorithm 3 converges in 7 iterations, while Algorithm 4 requires 18 iterations to converge. However, as seen from the output error $\varepsilon(\mu)$ of the final ROM over the test set Ξ_{test} in Fig. 1b, Algorithm 4 produces a ROM with errors being uniformly below the tolerance, whereas, the ROM computed using Algorithm 3 has large errors above the tolerance. In Fig. 2, we show the evolution of the coarse training set Ξ_c for different iterations of Algorithm 4. It is seen that the algorithm tends to pick parameters in the low-viscosity regions (close to 0.001), as expected, since the solution of the PDE tends to be 'less smooth', requiring more basis functions to approximate.

**Fig. 1:** Results for the Burgers' equation.**Fig. 2:** Burgers' equation: training set evolution.

4.2 Convection-diffusion equation

Next, we consider a 1-D model of brain transport, originally discussed in [1] and also considered in [9, 21]. The transport is modelled as a linear convection-diffusion PDE defined in the spatial domain $w \in \Omega := [0, 1]$ and for time $T := [0, 1]$,

$$\frac{dv}{dt} = q_1 \frac{\partial^2 v}{\partial w^2} + q_2 \frac{\partial v}{\partial w} - q_2, \quad (9)$$

where $v(w, t)$ is the state vector and the two parameters $(q_1, q_2) \in \mathcal{P} := [0.001, 1] \times [0.5, 5]$ are the diffusion and convection constants, respectively. The boundary conditions are given by,

$$v(w, 0) = \begin{cases} 1, & w \leq 0.5 \\ 0, & \text{otherwise} \end{cases}, \quad v(0, t) = v(1, t) = 0.$$

We discretize the equation using the FDM on a grid yielding $n = 800$. The output is calculated as the average value of the state in a small interval Ω_o centered around the midpoint of the domain at $w = 0.5$.

$$y(t) := \frac{1}{|\Omega_o|} \int_{\Omega_o} v(w, t) dw, \quad \Omega_o := [0.495, 0.505].$$

The tolerance is set as $\text{tol} = 10^{-5}$. For Algorithm 3, we set Ξ as a set including 25 random samples from the parameter domain \mathcal{P} , picked using the `rng` command and making use of `twister`, with a seed of 112. The fine training set Ξ_f for Algorithm 4 includes 1600 equidistant samples in \mathcal{P} . We choose a test set Ξ_{test} consisting of 625 random samples. The test parameters are generated using the same random number generator as for the Burgers' equation example, viz., `simdTwister` with a seed of 200. The error surrogate model $s(\mu)$ in Algorithm 4 is constructed using IMQ kernel

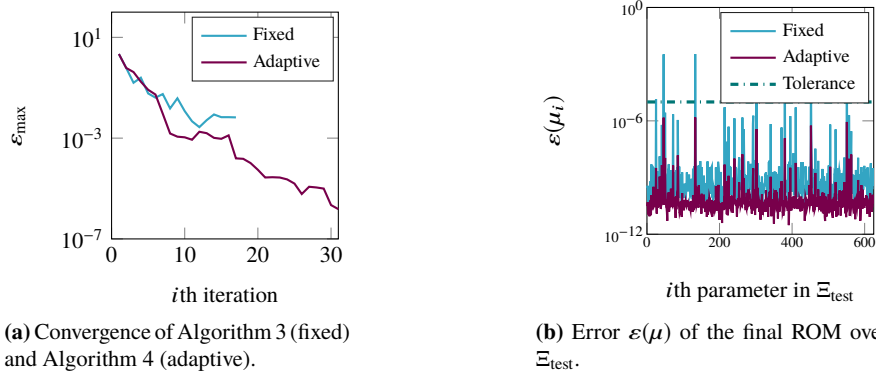


Fig. 3: Results for the Convection-diffusion equation.

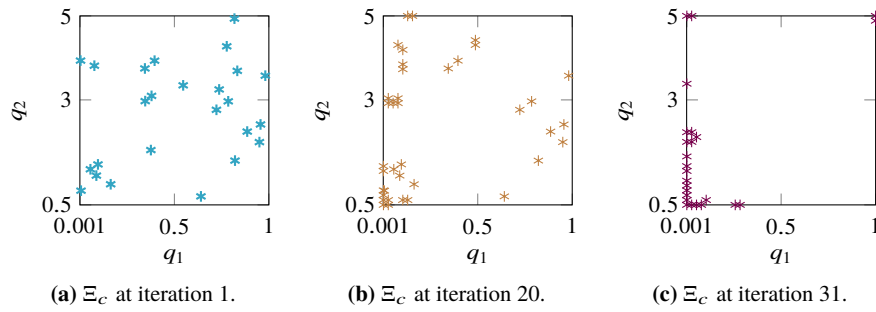


Fig. 4: Convection-diffusion equation: training set evolution.

function, where cross validation LOOCV has been applied to specify the shape parameter.

Fig. 3a shows the maximal error decay (ε_{\max}) over Ξ_{test} , computed at each iteration of Algorithm 3 and Algorithm 4, respectively. It is clear that adaptively enriching the training set leads to orders of magnitude faster convergence to the required accuracy. Fig. 3b plots the error $\varepsilon(\mu)$ (Eq. (8)) of the final ROM at every parameter in the test set Ξ_{test} , for the case of an adaptive training set and a fixed training set. We see that for some test parameters, the required tolerance is not met by the ROM computed using the fixed training set. In Fig. 4, the evolution of the training set is shown at different stages of Algorithm 4. It can be seen that samples from the left boundary of the parameter domain are added or kept. This has physical sense as this corresponds to the lower viscosity regions where the convective part of the solution dominates and the solution is ‘less smooth’. Thus, more basis functions are required for a good approximation.

In order to achieve a better ROM by using the fixed training set, we tried a Ξ with 225 random samples for Algorithm 3, leading to a ROM with error below the tolerance. In Table 2a, we provide the runtime results of both algorithms, where Algorithm 3 uses the refined training set with 225 samples. On the other hand, the initial coarse training set and the fine training set for Algorithm 4 remain the same. Algorithm 4 with adaptive sampling outperforms Algorithm 3, though both approaches produce accurate ROMs.

4.3 A thermal model

The final example is the Thermal model taken from the MORWiki benchmark collection¹. It models the heat transfer in a microthruster unit. The system is governed by the following semi-discretized ODE,

$$\mathbf{E}\dot{x} = \left(\mathbf{A} - \sum_{i=1}^3 h_i \mathbf{A}_i\right)x + \mathbf{B}u.$$

The model has been discretized in space using the FEM and the matrices are available from the Oberwolfach Benchmark Collection hosted at MORWiki. The dimension of the system is $n = 4257$. The parameters $\{h_i\}_{i=1}^3$ are the film coefficients at the top, bottom and side of the microthruster unit, respectively. They have a range between $[1, 10^8]$ in our simulations. The output is the temperature at the center of the polysilicon heater in the unit, which corresponds to the first row of the output matrix C .

Since this is a three-parameter problem with very large parameter domain, we use a large fixed training set (Ξ) with 6^3 logarithmically equidistant parameter samples for Algorithm 3 for a fair comparison. For Algorithm 4, the initial

¹ https://morwiki.mpi-magdeburg.mpg.de/morwiki/index.php/Thermal_Model

Algorithm	Runtime (seconds)
Fixed training set (Algorithm 3)	74.09
Adaptive training set (Algorithm 4)	41.52

(a) Convection-diffusion example.

Algorithm	Runtime (seconds)
Fixed training set (Algorithm 3)	151.16
Adaptive training set (Algorithm 4)	38.76

(b) Thermal example.

Table 2: Runtime comparison between Algorithm 3 and Algorithm 4.

coarse training set includes only $N_c = 10$ randomly chosen samples. To generate these random samples, we first generate a uniform sampling of each parameter $h_j, j = 1, 2, 3$, given as, $h_{ij} = 10^{\frac{i}{N_c/8}}, i = 1, 2, \dots, N_c$, which leads to a matrix $(h_{ij}) \in \mathbb{R}^{N_c \times 3}$. Then, the rows of the matrix are permuted using the `rng` and `randperm` commands. The seeds and random number generators used are 100 (`twister`), 120 (`combRecursive`) and 600 (`combRecursive`), respectively. In this way, we get random samples which spread through the 3-D parameter domain. The fine training set Ξ_f for Algorithm 4 is made up of 16^3 equidistant samples, while the test set Ξ_{test} consists of 8^3 logarithmically equidistant samples. For the RBF interpolation, Thin-plate spline kernel is used. No shape parameter needs to be determined. The tolerance for this model is set as `tol` = 10^{-3} . In Fig. 5a, we show error decay ε_{\max} over Ξ_{test} at each iteration of Algorithm 3 and Algorithm 4, respectively. Algorithm 3 takes 23 iterations to converge with the fixed training set Ξ . However, the maximum error ε_{\max} (Eq. (7)) over the test set Ξ_{test} is still above the tolerance. Algorithm 4 converges in 24 iterations. The maximum error ε_{\max} over Ξ_{test} is below the tolerance upon convergence. We note that, in the first few iterations, Algorithm 3 has a faster convergence in comparison to Algorithm 4. However, since the training set is fixed, the convergence eventually saturates. In Fig. 5b, we plot the error $\varepsilon(\mu)$ of the final ROM over Ξ_{test} , computed by the two algorithms respectively. Algorithm 4 once again outperforms Algorithm 3.

In Table 2b, we provide the runtime comparison between Algorithm 3 and Algorithm 4 for the thermal model, where an obvious speed-up by Algorithm 4 is observed. More importantly, with the reduced runtime, Algorithm 4 produces a ROM with sufficient accuracy, whereas, the ROM computed by Algorithm 3 still does not meet the accuracy requirement. This further justifies the motivation of using adaptive sampling for models with two or more parameters, especially when the parameter domain is very large.

5 Conclusion

Using a fixed training set in the POD-Greedy algorithm may either entail high computational costs or lead to large errors if the parameter domain is not properly sampled. In this work, we introduce a method to construct the training set for the RB

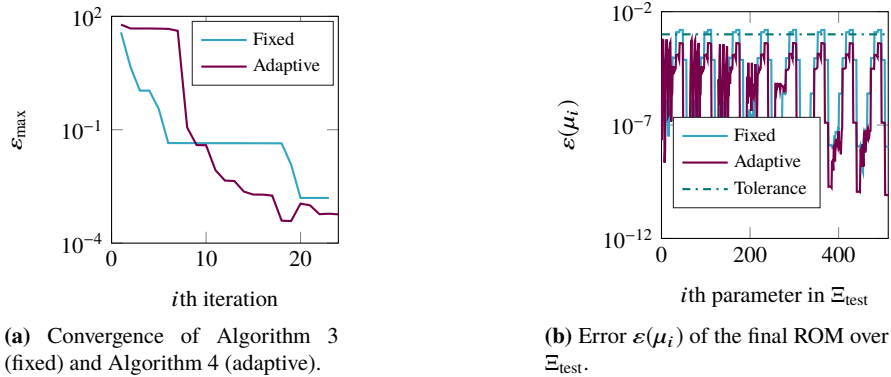


Fig. 5: Results for the thermal model.

method in an adaptive manner. To achieve this, we make use of an error surrogate model based on RBF interpolation.

The use of an error surrogate model enables sufficient exploration of the parameter space, since it avoids the need to solve the ROM at every parameter in the training set, as required by the error estimator. Furthermore, we implement a cross validation strategy in order to choose good shape parameters for the RBF kernels. The proposed algorithm is tested on several examples and it is shown to be effective in constructing a ROM with required accuracy. Furthermore, the adaptive parameter sampling method is integrated with the adaptive POD-Greedy-(D)EIM algorithm in [5] to achieve a fully adaptive scheme for the RB method.

As future work, we propose to extend the adaptive sampling approach to the frequency domain model reduction methods, such as multi-moment matching [8], in order to adaptively sample the interpolation points.

References

1. Banks, H.T., Kunisch, K.: Estimation Techniques for Distributed Parameter Systems, *Systems & Control: Foundations & Applications*, vol. 6. Birkhäuser, Boston (1989). DOI 10.1007/978-1-4612-3700-6
2. Barrault, M., Maday, Y., Nguyen, N.C., Patera, A.T.: An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations. *C.R. Acad. Sci. Paris* **339**(9), 667–672 (2004)
3. Buhmann, M.D.: Radial basis functions: theory and implementations, *Cambridge Monographs on Applied and Computational Mathematics*, vol. 12. Cambridge University Press, Cambridge (2003)
4. Chaturantabut, S., Sorensen, D.C.: Nonlinear model reduction via discrete empirical interpolation. *SIAM J. Sci. Comput.* **32**(5), 2737–2764 (2010). DOI 10.1137/090766498
5. Chellappa, S., Feng, L., Benner, P.: Adaptive basis construction and improved error estimation for parametric nonlinear dynamical systems. In preparation (2019)

6. Eftang, J.L., Knezevic, D.J., Patera, A.T.: An hp certified reduced basis method for parametrized parabolic partial differential equations. *Math. Comput. Model. Dyn. Syst.* **17**(4), 395–422 (2011)
7. Fasshauer, G.E., Zhang, J.G.: On choosing “optimal” shape parameters for RBF approximation. *Numer. Algorithms* **45**(1), 345–368 (2007)
8. Feng, L., Benner, P.: A robust algorithm for parametric model order reduction based on implicit moment matching. In: A. Quarteroni, G. Rozza (eds.) *Reduced Order Methods for modeling and computational reduction, MS & A*, vol. 9, pp. 159–186. Springer-Verlag, Berlin, Heidelberg, New York (2014). DOI 10.1007/978-3-319-02090-7_6
9. Grepl, M.: Reduced-basis approximation a posteriori error estimation for parabolic partial differential equations. Ph.D. thesis, Massachusetts Institute of Technology (MIT), Cambridge, USA (2005)
10. Haasdonk, B., Dihlmann, M., Ohlberger, M.: A training set and multiple bases generation approach for parameterized model reduction based on adaptive grids in parameter space. *Math. Comput. Model. Dyn. Syst.* **17**(4), 423–442 (2011)
11. Haasdonk, B., Ohlberger, M.: Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Math. Model. Numer. Anal.* **42**(2), 277 – 302 (2008)
12. Hesthaven, J.S., Stamm, B., Zhang, S.: Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. *ESAIM: Math. Model. Numer. Anal.* **48**(1), 259–283 (2014)
13. James, G., Witten, D., Hastie, T., Tibshirani, R.: An introduction to statistical learning : With applications in R, *Springer Texts in Statistics*, vol. 103. Springer, New York (2013)
14. Maday, Y., Stamm, B.: Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces. *SIAM J. Sci. Comput.* **35**(6), A2417–A2441 (2013)
15. Paul-Dubois-Taine, A., Amsallem, D.: An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models. *Internat. J. Numer. Methods Engrg.* **102**(5), 1262–1292 (2015)
16. Quarteroni, A., Manzoni, A., Negri, F.: Reduced Basis Methods for Partial Differential Equations, *La Matematica per il 3+2*, vol. 92. Springer International Publishing (2016). ISBN: 978-3-319-15430-5
17. Rippa, S.: An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *Adv. Comput. Math.* **11**(2), 193–210 (1999)
18. Sen, S.: Reduced-basis approximation and a posteriori error estimation for many-parameter heat conduction problems. *Numerical Heat Transfer, Part B: Fundamentals* **54**(5), 369–389 (2008)
19. van Stein, B., Wang, H., Kowalczyk, W., Emmerich, M.T.M., Bäck, T.: Cluster-based Kriging approximation algorithms for complexity reduction (2017). URL <http://arxiv.org/abs/1702.01313>
20. Wendland, H.: Scattered data approximation, *Cambridge Monographs on Applied and Computational Mathematics*, vol. 17. Cambridge University Press, Cambridge (2005)
21. Zhang, Y., Feng, L., Li, S., Benner, P.: An efficient output error estimation for model order reduction of parametrized evolution equations. *SIAM J. Sci. Comput.* **37**(6), B910–B936 (2015)