

Multi-fidelity error estimation accelerates greedy model reduction of complex dynamical systems

Lihong Feng ^{*}, Luigi Lombardi[†], Giulio Antonini[‡], and Peter Benner[§]

January 20, 2023

Abstract

Model order reduction usually consists of two stages: the offline stage and the online stage. The offline stage is the expensive part that sometimes takes hours till the final reduced-order model is derived, especially when the original model is very large or complex. Once the reduced-order model is obtained, the online stage of querying the reduced-order model for simulation is very fast and often real-time capable. This work concerns a strategy to significantly speed up the offline stage of model order reduction for large and complex systems. In particular, it is successful in accelerating the greedy algorithm that is often used in the offline stage for reduced-order model construction. We propose multi-fidelity error estimators and replace the high-fidelity error estimator in the greedy algorithm. Consequently, the computational complexity at each iteration of the greedy algorithm is reduced and the algorithm converges more than 3 times faster without incurring noticeable accuracy loss.

1 Introduction

Model order reduction (MOR) has achieved much success in many areas of computational science with its capability of realizing real-time simulation and providing accurate results. Different MOR methods, their applications and the promising results they produce can be found in the survey papers [2, 4, 12] and books [3, 8, 9, 10, 11, 27].

MOR needs an offline stage for constructing the ROM. For many intrusive MOR methods that are based on projection, the offline stage is usually realized via a greedy algorithm. The greedy algorithm is used to properly select *important* parameter samples that contribute most to the solution space. The offline computational time is basically the runtime of the greedy algorithm. For large-scale systems, the offline computation is expensive and the runtime is often longer than several hours even when run on a high-performance server. Sometimes, the system is not very large, for example, the number of degrees of freedom is only $O(10^5)$, but the system structure is complicated, so that the greedy algorithm still takes long time to converge.

It is known that an efficient error estimator makes the greedy algorithm successful in producing an accurate ROM without running many iterations. Therefore, many efforts have been made in this direction to develop computable error estimators for different problems [15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 28, 33, 34, 35]. However, more attention has been paid to improve the effectivity

^{*}Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, 39106 Magdeburg, Germany feng@mpi-magdeburg.mpg.de

[†]Luigi Lombardi is with Micron Semiconductor, 67051 Avezzano, Italy. luigilombardi89@gmail.com

[‡]Giulio Antonini is with the UAq EMC Laboratory, Department of Industrial and Information Engineering and Economics, University of L'Aquila, I-67100 L'Aquila, Italy. giulio.antonini@univaq.it

[§]Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany and Fakultät für Mathematik, Otto-von-Guericke-Universität Magdeburg, Germany. benner@mpi-magdeburg.mpg.de

or accuracy of the error estimator than to develop more efficient strategies to accelerate the greedy process [13, 25, 33, 34, 36]. Recently, some techniques are proposed to improve the adaptivity of the greedy algorithm [6, 7, 13, 26].

In [13, 14], we proposed a surrogate model for error estimation, and proposed an adaptive greedy algorithm by alternatively using this surrogate error estimator and the original error estimator during the greedy algorithm. The focus in [13, 14] was to make the greedy process adaptive by starting from a coarse training set of a small number of parameter samples, and adaptively update the coarse training set with the aid of a surrogate error estimator. The original error estimator is computed only over the coarse training set, while the surrogate error estimator helps to pick out candidates of *important* parameter samples from a fine training set, which are then collected in the coarse training set.

In this work, we emphasize the role of the surrogate error estimator and propose the concept of bi-fidelity error estimation and multi-fidelity error estimation. In fact, a bi-fidelity error estimation has been used in the adaptive greedy algorithm proposed in [13, 14] without being formally defined, i.e., the original (high-fidelity) error estimator, and the surrogate (low-fidelity) error estimator. To further improve the convergence speed of the greedy algorithm, we propose multi-fidelity error estimation built upon the bi-fidelity error estimation. Here, we use a more efficient high-fidelity error estimator than the two different high-fidelity error estimators used in [13, 14]. Although the proposed multi-fidelity error estimation is dependent on the original high-fidelity error estimator, the idea of using multi-fidelity error estimation is general and can be extended to develop multi-fidelity error estimation associated with other high-fidelity error estimators.

Unlike the problems considered in [13, 14], whose ROMs can be constructed by standard greedy algorithms within seconds to minutes, we consider in this work much more complicated problems. On the same computer, the standard greedy algorithm takes more than half a day to converge for such problems. By using the proposed multi-fidelity error estimator, the greedy algorithm achieves 4x speed-up and produces ROMs with little loss of accuracy. The speed-up is also higher than those reported in [13, 14] by using the bi-fidelity error estimation, which is usually 2x.

In the next section, we present the greedy algorithm in the standard form. Then we analyze some ingredients of the algorithm, which contribute most to the computational cost. Starting from those computationally expensive parts, we develop possible strategies to reduce the computational complexity in Section 3. As a consequence, it becomes clear that the resulting strategy develops a greedy algorithm with multi-fidelity error estimation. The proposed algorithm is then applied to large time-delay systems with many delays. The numerical tests on three large time-delay systems with more than 100 delays are demonstrated in Section 4. Conclusions are given in the end.

2 Standard greedy algorithm

The standard greedy algorithm was first proposed for steady systems without time evolution. Then it was extended to POD-greedy for dynamical systems, which is used to construct the ROM using snapshots in the time domain. Later the greedy algorithm found its capability in adaptively choosing interpolation points for frequency-domain MOR methods [15, 16]. The greedy algorithm for steady systems and frequency-domain MOR has the same formulation, whereas POD-greedy for time-domain MOR of time-dependent systems needs an SVD step at each greedy iteration. In this work, we focus on the greedy algorithm, though the proposed scheme can be easily extended to POD-greedy. We consider constructing a ROM for the following full-order model (FOM) using the greedy algorithm,

$$F(x(\mu), \mu) = B(\mu). \tag{1}$$

Here, $F(x(\mu), \mu) \in \mathbb{C}^{n \times n_I}$, $x(\mu) \in \mathbb{C}^{n \times n_I}$, and $B(\mu) \in \mathbb{C}^{n \times n_I}$. $\mu \in \mathcal{P}$ is a parameter in the parameter domain \mathcal{P} . The variable n is the order of the FOM, which can be the number of degrees of freedom

Algorithm 1 Standard greedy algorithm

Input: the FOM, a training set Ξ composed of parameter samples taken from the parameter domain \mathcal{P} , error tolerance $\text{tol} < 1$, $\Delta(\mu)$ to estimate the error.

Output: Projection matrix V .

- 1: Choose initial parameter $\mu^* \in \Xi$.
 - 2: $V \leftarrow \emptyset$, $\varepsilon = 1$.
 - 3: **while** $\varepsilon > \text{tol}$ **do**
 - 4: Compute the snapshot(s) $x(\mu^*)$ by solving the FOM at $\mu = \mu^*$.
 - 5: Update V by $V = \text{orth}\{V, x(\mu^*)\}$, (e.g., using the modified Gram-Schmidt process with deflation.)
 - 6: Compute μ^* such that $\mu^* = \arg \max_{\mu \in \Xi} \Delta(\mu)$.
 - 7: $\varepsilon = \Delta(\mu^*)$.
 - 8: **end while**
-

after numerical discretization of PDEs describing a physical phenomenon. The proposed algorithms are also applicable to problems with $n_I > 1$.

The ROM can be obtained via Galerkin projection using a projection matrix $V \in \mathbb{R}^{n \times r}$, $r \ll n$, as below,

$$\hat{F}(Vz(\mu), \mu) = \hat{b}(\mu), \quad (2)$$

where $\hat{F}(Vz(\mu), \mu) = V^T f(Vz(\mu), \mu) \in \mathbb{C}^{r \times n_I}$, $z(\mu) \in \mathbb{C}^{r \times n_I}$, and $\hat{B}(\mu) = V^T b(\mu) \in \mathbb{C}^{r \times n_I}$.

The standard greedy process used to compute the projection matrix V is described in Algorithm 1. Step 4 in Algorithm 1 solves the FOM at μ^* , and Step 6 computes an error estimator $\Delta(\mu)$ at all μ in Ξ . These two steps constitute the most computational expensive part of the greedy algorithm. However, Step 4 is unavoidable, since $x(\mu^*)$ is needed for the reduced basis construction. The computational complexity of Step 6 could be reduced, if the cardinality of Ξ , i.e., $|\Xi|$ is kept small, so that $\Delta(\mu)$ needs not be evaluated at many parameter samples. This is the motivation of the surrogate error estimator proposed in [13, 14]. We call the surrogate error estimator $\Delta_l(\mu)$ the low-fidelity error estimator as compared to the original error estimator $\Delta(\mu)$, since $\Delta_l(\mu)$ is only an approximation to $\Delta(\mu)$, but is much cheaper to compute.

In the next section, we present a greedy algorithm using bi-fidelity error estimation, where the low-fidelity error estimator is computed following the method in [13, 14]. Based on this, a greedy algorithm using multi-fidelity error estimation associated with a particular high-fidelity error estimator for MOR of linear parametric systems, is proposed.

3 Greedy algorithm with bi-fidelity and multi-fidelity error estimation

This section presents greedy algorithms with bi-fidelity and multi-fidelity error estimation, respectively.

3.1 Greedy algorithm with bi-fidelity error estimation

Algorithm 2 is the greedy algorithm with bi-fidelity error estimation. Its original version using a different high-fidelity error estimator was firstly proposed in [13]. The key step of Algorithm 2 is Step 8, where the low-fidelity error estimator $\Delta_l(\mu)$ is computed using values of $\Delta(\mu)$ at the samples of μ in the small parameter set Ξ_c . Basically, $\Delta_l(\mu)$ is represented by a weighted sum of radial basis

Algorithm 2 Greedy algorithm with bi-fidelity error estimation

Input: the FOM, a training set Ξ_c composed of a small number of parameter samples taken from the parameter domain \mathcal{P} , a set Ξ_f composed a large number of parameter samples of μ from \mathcal{P} , error tolerance $\text{tol} < 1$, $\Delta(\mu)$ to estimate the error.

Output: Projection matrix V .

- 1: Choose initial parameter $\mu^* \in \Xi_c$.
 - 2: $V \leftarrow \emptyset$, $\varepsilon = 1$.
 - 3: **while** $\varepsilon > \text{tol}$ **do**
 - 4: Compute the snapshot(s) $x(\mu^*)$ by solving the FOM at $\mu = \mu^*$.
 - 5: Update V by $V = \text{orth}\{V, x(\mu^*)\}$, (e.g., using the modified Gram-Schmidt process with deflation.)
 - 6: Compute μ^* such that $\mu^* = \arg \max_{\mu \in \Xi_c} \Delta(\mu)$.
 - 7: Compute μ^o such that $\mu^o = \arg \min_{\mu \in \Xi_c} \Delta(\mu)$.
 - 8: Compute the low-fidelity error estimator $\Delta_l(\mu)$ using values of $\Delta(\mu)$ corresponding to the samples of μ in Ξ_c via (3) and (4).
 - 9: Evaluate $\Delta_l(\mu)$ over Ξ_f and pick out a parameter μ_c from the large parameter set Ξ_f corresponding to the largest value of $\Delta_l(\mu)$, i.e., $\mu_c = \arg \max_{\mu \in \Xi_f} \Delta_l(\mu)$.
 - 10: Update the small parameter set Ξ_c : if $\Delta_l(\mu_c) > \text{tol}$, enrich Ξ_c with μ_c , i.e., $\Xi_c = \{\Xi_c, \mu_c\}$, if $\Delta(\mu^o) < \text{tol}$, remove μ^o from Ξ_c : $\Xi_c = \Xi_c \setminus \mu^o$.
 - 11: $\varepsilon = \Delta(\mu^*)$.
 - 12: **end while**
-

functions (RBFs), i.e.,

$$\Delta_l(\mu) = \sum_{i=1}^m w_i \Phi(\mu - \mu_i), \quad (3)$$

where $\Phi(\mu)$ are RBFs, μ_i are the samples in Ξ_c , and m is the cardinality of Ξ_c , which is small. Once w_i are known, the low-fidelity error estimator $\Delta_l(\mu)$ is known. The weights w_i are computed via enforcing $\Delta_l(\mu)$ to interpolate $\Delta(\mu)$ at $\mu_j, \forall \mu_j \in \Xi_c$, i.e., $\Delta_l(\mu_j) = \Delta(\mu_j)$. Inserting $\mu = \mu_j \in \Xi_c$ into (3), the weights w_i can be computed by solving the linear system as below,

$$\begin{bmatrix} \Phi(\mu_1 - \mu_1) & \dots & \Phi(\mu_1 - \mu_m) \\ \vdots & \vdots & \vdots \\ \Phi(\mu_m - \mu_1) & \dots & \Phi(\mu_m - \mu_m) \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} \Delta(\mu_1) \\ \vdots \\ \Delta(\mu_m) \end{bmatrix}. \quad (4)$$

Since values of $\Delta(\mu)$ at $\mu \in \Xi_c$ are available, the weights can be easily computed by solving the above small linear system with $m \times m$ being the dimension of the coefficient matrix. As this is a rather small system, we usually do not observe ill conditioning. Otherwise, we can use a regularized version of (4) [13, 14].

At each iteration of the bi-fidelity greedy algorithm, the linear system is solved for once (Step 8), then the low-fidelity error estimator $\Delta_l(\mu)$ is evaluated over a larger parameter set Ξ_f using the weighted sum in (3) (Step 9). This process of computing the weights and evaluating $\Delta_l(\mu)$ is much faster than evaluating the high-fidelity error estimator over a training set Ξ whose cardinality $|\Xi|$ is much larger than $|\Xi_c|$. This is usually the case for the standard greedy algorithm, where $|\Xi| > |\Xi_c|$. Finally, at each iteration of the bi-fidelity algorithm, the total computational cost of Steps 6-8: computing the high-fidelity error estimator $\Delta(\mu)$ over Ξ_c , solving the linear system (4) and evaluating the low-fidelity error estimator $\Delta_l(\mu)$ over Ξ_f is still much cheaper than computing the

high-fidelity error estimator $\Delta(\mu)$ over a training set Ξ , whose cardinality is, e.g., twice that of $|\Xi_c|$, as shown in the numerical tests.

Besides computing μ^* corresponding to the maximal value of the error estimator $\Delta(\mu)$ over Ξ_c , the minimal value of $\Delta(\mu)$ is also computed in Step 7. The corresponding parameter μ^o could be deleted from Ξ_c if $\Delta(\mu^o)$ is already below the tolerance tol , see Step 10. In this way, the cardinality of the training set Ξ_c remains almost constant, and can further save computations as compared with enriching Ξ_c only. We will show in the numerical results that adding and removing samples to and from Ξ_c gets ROMs with similar accuracy (even smaller) as only adding samples to Ξ_c , but leads to even faster convergence of the greedy algorithm.

Remark 3.1 *In Step 9, it is also possible to choose more than one parameter from Ξ_f by modifying Step 9 as: choose n_{add} samples from Ξ_f corresponding to n_{add} largest values of $\Delta_l(\mu)$. Similarly, In Step 7, one can also choose $n_{del} > 1$ parameter samples corresponding to n_{del} smallest values of $\Delta(\mu)$ from Ξ_c . However, this will more or less increase the computational time at each iteration, since more computations are needed to choose those samples. Furthermore, to make sure that only samples at which $\Delta_l(\mu)$ is larger than the tolerance tol are added to Ξ_c , and only samples at which $\Delta(\mu)$ is smaller than tol are removed, additional calculations are necessary to check if all the selected samples meet the above criteria and should be finally selected or removed (see Step 10). Therefore, adding/removing at most one parameter sample each time should be more efficient. In the numerical tests, we also show results when $n_{add} = n_{del} = 2$ and $n_{add} = n_{del} = 5$ at each iteration of Algorithm 2.*

The bi-fidelity error estimation is general and can be applied to any high-fidelity error estimators. For example, the high-fidelity error estimator in [13] estimates the error of the ROM for nonlinear time-dependent parametric systems in the time domain, while the high-fidelity error estimator in [14] estimates the error of the ROM in the frequency-domain for linear parametric systems.

3.2 Greedy algorithm with multi-fidelity error estimation

The multi-fidelity error estimation we are going to introduce depends on the formulation of the high-fidelity error estimator $\Delta(\mu)$. To illustrate the basic concept, we use an error estimator proposed in [16] as the high-fidelity error estimator and discuss how to further reduce the computational load by using multi-fidelity error estimation.

3.2.1 An error estimator for linear parametric systems

The error estimator is applicable to estimating the output error of the ROM for FOMs in the following linear parametric form,

$$\begin{aligned} M(\mu)x(\mu) &= B(\mu), \\ y(\mu) &= C(\mu)x(\mu). \end{aligned} \tag{5}$$

Here, $M(\mu) \in \mathbb{R}^{n \times n}$, $B(\mu) \in \mathbb{R}^{n \times n_I}$, $C(\mu) \in \mathbb{R}^{n_O \times n}$, $x(\mu) \in \mathbb{R}^n$, $y(\mu) \in \mathbb{R}^{n_O \times n_I}$. We consider the general case that both $B(\mu)$ and $C(\mu)$ are matrices, i.e. systems with multiple inputs and multiple outputs. The ROM of the above linear parametric system can be derived via Galerkin projection using a projection matrix V composed of the reduced basis. That is,

$$\begin{aligned} \hat{M}(\mu)z(\mu) &= \hat{B}(\mu), \\ \hat{y}(\mu) &= \hat{C}(\mu)z(\mu), \end{aligned} \tag{6}$$

where $\hat{M}(\mu) = V^T M(\mu) V$, $\hat{B}(\mu) = V^T B(\mu)$, $\hat{C}(\mu) = C(\mu) V$.

For the general situation when both $B(\mu)$ and $C(\mu)$ are matrices, the error of the i, j -th entry of the output matrix $\hat{y}(\mu)$ is

$$\begin{aligned}
& |y_{ij}(\mu) - \hat{y}_{ij}(\mu)| \\
&= |C_i(\mu)(M^{-1}(\mu)B(\mu) - V\hat{M}^{-1}(\mu)\hat{B}_j(\mu))| \\
&= |C_i(\mu)M^{-1}(\mu)(B_j(\mu) - \underbrace{M(\mu)V\hat{M}^{-1}(\mu)\hat{B}_j(\mu)}_{\hat{x}_j(\mu):=Vz_j(\mu)})| \\
&= |C_i(\mu)M^{-1}(\mu)r_j(\mu)|,
\end{aligned} \tag{7}$$

where $C_i(\mu)$ is the i -th row of $C(\mu)$ and $B_j(\mu)$ is the j -th column of $B(\mu)$. Here, we have defined: $z_j(\mu) = \hat{M}^{-1}(\mu)\hat{B}_j(\mu)$, i.e., $\hat{M}(\mu)z_j(\mu) = \hat{B}_j(\mu)$, $\hat{x}_j(\mu) := Vz_j(\mu)$ and $r_j(\mu) := B_j(\mu) - M(\mu)\hat{x}_j(\mu)$. It is clear that

$$\hat{M}(\mu)z_j(\mu) = \hat{B}_j(\mu)$$

is a reduced-order model of

$$M(\mu)x_j(\mu) = B_j(\mu), \tag{8}$$

and $\hat{x}_j(\mu) \approx x_j(\mu)$, the j -th column of $x(\mu)$. Finally, $r_j(\mu)$ is the residual induced by $\hat{x}_j(\mu)$.

From the last equation in (7), it is clear that to compute the absolute error of \hat{y}_{ij} , we need to solve a residual system:

$$M(\mu)x_{r_j}(\mu) = r_j(\mu). \tag{9}$$

Instead, we construct a ROM of it:

$$V_r^T M(\mu)V_r z_{r_j}(\mu) = V_r^T r_j(\mu), \tag{10}$$

so that $x_{r_j}(\mu) \approx \hat{x}_{r_j}(\mu) = V_r z_{r_j}(\mu)$. Finally,

$$|y_{ij}(\mu) - \hat{y}_{ij}(\mu)| \approx |C_i(\mu)\hat{x}_{r_j}(\mu)|.$$

Note that $\hat{x}_{r_j}(\mu)$ depends on $B_j(\mu)$, since $r_j(\mu)$ depends on $B_j(\mu)$. Each column $B_j(\mu)$ is associated with a $\hat{x}_{r_j}(\mu)$. The overall error of $\hat{y}(\mu)$ as a matrix can be estimated as:

$$\|y(\mu) - \hat{y}(\mu)\|_{\max} := \max_{i,j} |y_{ij}(\mu) - \hat{y}_{ij}(\mu)| \approx \max_{i,j} |C_i(\mu)\hat{x}_{r_j}(\mu)| =: \tilde{\Delta}(\mu). \tag{11}$$

$\tilde{\Delta}(\mu)$ defined in (11) is one of the error estimators proposed in [16], where the proposed error estimators were shown to outperform other existing error estimators in the literature [34, 15] in terms of both accuracy and computational efficiency. Furthermore, it has been discussed in [16] that $\tilde{\Delta}(\mu)$ is even more accurate but has less computational complexity than other proposed estimators, including the one used in [14]. Even with this error estimator, the greedy algorithm could take several hours to converge for some complex systems, for example, the time-delay systems we consider in this work. For such systems, although the standard greedy algorithm can already be accelerated by the bi-fidelity greedy algorithm, we suggest a possibility to further improve the bi-fidelity greedy algorithm by introducing multi-fidelity error estimation.

We notice that in order to compute $\tilde{\Delta}(\mu)$, an extra projection matrix V_r has to be constructed for $\hat{x}_{r_j}(\mu)$. Although $\hat{x}_{r_j}(\mu)$ is dependent on the individual column of $B(\mu)$, the matrix V_r can be uniformly constructed based on the whole matrix $B(\mu)$. Then V_r is valid for any column of $B(\mu)$. It is proved in [16] that taking $V_r = V$ leads to $\tilde{\Delta}(\mu)$ identically zero for all μ . Therefore, V_r should be additionally computed.

3.2.2 Standard greedy algorithm using $\tilde{\Delta}(\mu)$

For easy understanding of the multi-fidelity error estimation, we first present Algorithm 3, the standard greedy algorithm using $\tilde{\Delta}(\mu)$ in (11) as the error estimator. There, some additional steps are added to compute V_r , see Step 5, Steps 7-8. In Step 7 of Algorithm 3, V_r is not only updated by $x(\mu^r)$, but also by V . This is due to the fact that the solution $x_{r_j}(\mu)$ to the residual system in (9) can be written as

$$\begin{aligned}
 x_{r_j}(\mu) &= M(\mu)^{-1}r_j(\mu) \\
 &= M(\mu)^{-1}(B_j(\mu) - M(\mu)\hat{x}_j(\mu)) \\
 &= M(\mu)^{-1}B_j(\mu) - Vz_j(\mu) \\
 &\approx \tilde{V}_r z_{r_j} - Vz_j(\mu).
 \end{aligned} \tag{12}$$

It is clear that $x_{r_j}(\mu)$ is a linear combination of $(M(\mu))^{-1}B_j(\mu)$ and the columns of V . Therefore, V contributes to the subspace approximating the solution space of $x_{r_j}(\mu)$ and cannot be neglected. It is also noticed that $(M(\mu))^{-1}B_j(\mu)$ is in fact the solution $x_j(\mu)$ in (8), while $Vz_j(\mu)$ is $\hat{x}_j(\mu)$ that approximates $x_j(\mu)$. This means $x_{r_j}(\mu)$ is the difference between $x_j(\mu)$ and $\hat{x}_j(\mu)$, which is a nonzero vector. Therefore, we should compute another matrix \tilde{V}_r , so that $x_j(\mu) \approx \tilde{V}_r z_{r_j}(\mu)$, but $\tilde{V}_r z_{r_j}(\mu) \neq \hat{x}_j(\mu) = Vz_j(\mu)$. Finally, x_{r_j} is approximated by the difference between $\tilde{V}_r z_{r_j}(\mu)$ and $Vz_j(\mu)$. In other words, it is approximately represented as the linear combination of the columns of both V_r and V . This approximation also explains Step 5 and Step 7 of Algorithm 3: Step 5 computes the reduced basis vectors contributing to \tilde{V}_r , Step 7 computes the complete reduced basis vectors contributing to V_r . New reduced basis vectors for both V and V_r are computed at each iteration of the greedy algorithm. Step 8 and Step 9 compute the new *important* parameter samples for V_r and V , respectively. In general, μ^r should be different from μ^* , since $\tilde{\Delta}(\mu) \neq \max_{j=1, \dots, n_I} \|r_j(\mu) - M(\mu)\hat{x}_{r_j}(\mu)\|$. Here, $r_j(\mu) - M(\mu)\hat{x}_{r_j}(\mu)$ is nothing but the residual induced by the approximate solution ($\hat{x}_{r_j}(\mu)$) to the residual system (9).

Algorithm 3 Standard greedy algorithm using $\tilde{\Delta}(\mu)$ for linear parametric systems.

Input: the FOM, a training set Ξ composed of parameter samples taken from the parameter domain $\mu \in \mathcal{P}$, error tolerance $\text{tol} < 1$.

Output: Projection matrix V .

- 1: Choose initial parameter $\mu^* \in \Xi$ for V , and initial parameter $\mu^r \neq \mu^* \in \Xi$ for V_r .
 - 2: $V \leftarrow \emptyset$, $V_r \leftarrow \emptyset$, $\varepsilon = 1$.
 - 3: **while** $\varepsilon > \text{tol}$ **do**
 - 4: Compute the snapshot(s) $x(\mu^*)$ by solving the FOM, i.e. $x(\mu^*) = (M(\mu^*))^{-1}B(\mu^*)$.
 - 5: Compute the snapshot(s) $x(\mu^r)$ by solving the FOM, i.e. $x(\mu^r) = (M(\mu^r))^{-1}B(\mu^r)$.
 - 6: Update V by $V = \text{orth}\{V, x(\mu^*)\}$, (e.g., using the modified Gram-Schmidt process with deflation.)
 - 7: Update V_r by $V_r = \text{orth}\{V, V_r, x(\mu^r)\}$.
 - 8: Compute μ^r such that $\mu^r = \arg \max_{\mu \in \Xi} \max_{j=1, \dots, n_I} \|r_j(\mu) - M(\mu)\hat{x}_{r_j}(\mu)\|$, (n_I is the total number of columns of $B(\mu)$).
 - 9: Compute μ^* such that $\mu^* = \arg \max_{\mu \in \Xi} \tilde{\Delta}(\mu)$.
 - 10: $\varepsilon = \tilde{\Delta}(\mu^*)$.
 - 11: **end while**
-

3.2.3 Greedy algorithm with multi-fidelity error estimation

The computational complexity of Algorithm 3 using the error estimator $\tilde{\Delta}(\mu)$ comes from Steps 4-9. Efficiency of Step 9 can be improved by using the bi-fidelity error estimation as shown in Algorithm 2. The computations in Step 4, 6 are unavoidable, since V is used to compute the ROM of the original FOM and should be updated till acceptable error tolerance is satisfied. In contrast, V_r in Step 7 needs not be updated at every iteration. This implies that the ROM of the residual system does not have to be very accurate, since it is not the ROM that we seek, but an auxiliary ROM aiding the computation of $\tilde{\Delta}(\mu)$.

An immediate consequence of Theorem 4.2 in [16] for single-input and single-output systems is the following Lemma for systems with multiple inputs and multiple outputs:

Lemma 3.1 *The error of the output $\hat{y}(\mu)$ of the ROM (6) can be bounded as*

$$\tilde{\Delta}(\mu) - \delta(\mu) \leq \|y(\mu) - \hat{y}(\mu)\|_{\max} \leq \tilde{\Delta}(\mu) + \delta(\mu), \quad (13)$$

where $\delta(\mu) := \max_{i,j} |C_i(\mu)(x_{r_j}(\mu) - \hat{x}_{r_j}(\mu))| \geq 0$.

Proof From (7), we know

$$|y_{ij}(\mu) - \hat{y}_{ij}(\mu)| = |C_i(\mu)x_{r_j}(\mu)| \approx |C_i(\mu)\hat{x}_{r_j}(\mu)|.$$

Then

$$\begin{aligned} |y_{ij}(\mu) - \hat{y}_{ij}(\mu)| &= |C_i(\mu)x_{r_j}(\mu)| + |C_i(\mu)\hat{x}_{r_j}(\mu)| - |C_i(\mu)\hat{x}_{r_j}(\mu)| \\ &\leq |C_i(\mu)\hat{x}_{r_j}(\mu)| + \underbrace{|C_i(\mu)x_{r_j}(\mu) - C_i(\mu)\hat{x}_{r_j}(\mu)|}_{\delta_{ij}(\mu)}. \end{aligned} \quad (14)$$

On the other hand,

$$\begin{aligned} |C_i(\mu)\hat{x}_{r_j}(\mu)| &= |C_i(\mu)\hat{x}_{r_j}(\mu)| + |C_i(\mu)x_{r_j}(\mu)| - |C_i(\mu)x_{r_j}(\mu)| \\ &\leq |C_i(\mu)x_{r_j}(\mu)| + \delta_{ij}(\mu). \end{aligned} \quad (15)$$

From (11), (15) and the definition of $\delta(\mu)$, we have

$$\tilde{\Delta}(\mu) = \max_{i,j} |C_i(\mu)\hat{x}_{r_j}(\mu)| \leq \|y(\mu) - \hat{y}(\mu)\|_{\max} + \delta(\mu).$$

Similarly, from (14), we get

$$\|y(\mu) - \hat{y}(\mu)\|_{\max} \leq \tilde{\Delta}(\mu) + \delta(\mu).$$

This completes the proof. ■

From the definition of $\delta(\mu)$, it is seen that the more accurate the ROM of the residual system, the smaller $\delta(\mu)$ is. As a result, $\tilde{\Delta}(\mu)$ should measure the true error more accurately so that the *important* parameters it selects are closer to those selected by the true error, given the same training set Ξ . On the contrary, if the ROM of the residual system is less accurate, $\tilde{\Delta}(\mu)$ will be less accurate, too. However, at a certain stage, when $\tilde{\Delta}(\mu)$ is already small, the right-hand side of the residual system $r_j(\mu)$ will also be small, so that it can be expected that both $x_{r_j}(\mu)$ and $\hat{x}_{r_j}(\mu)$ are close to zero. This leads to a small $\delta(\mu)$. Variation of a small $\delta(\mu)$ will not cause big variation of the difference between $\tilde{\Delta}(\mu)$ and the true error $\|y(\mu) - \hat{y}(\mu)\|_{\max}$. The trend, though not the exact route, of error decay could still be anticipated so that *important* parameters corresponding to the error peaks can also be detected. The above analyses are also justified by the numerical results in the next section, see, e.g., Figure 3 and Figure 5.

This motivates the multi-fidelity error estimation. We set a second tolerance $\epsilon > \text{tol}$, and when $\tilde{\Delta}(\mu) < \epsilon < 1$, we stop updating the ROM of the residual system, i.e., stop implementing Step 5, Step 7 and Step 8 of Algorithm 3. The error estimator $\tilde{\Delta}(\mu)$ after this stage may not be as accurate as it would be when keep updating the ROM of the residual system. However, the difference should be small as $\tilde{\Delta}(\mu)$ is already below a small value ϵ . Without implementing Step 5, we have saved computations of simulating the FOM. For large and complex systems, solving the FOM even once is not fast. The computation in Step 7 is relatively cheap if the system is not very large. The computational cost in Step 8 is not low for certain complex problems, though some μ -independent parts of $r_j(\mu)$ and $M(\mu)$ can be pre-computed. For example, this is the case for the time-delay systems in the next section.

Stop updating the ROM of the residual system gives rise to a low-fidelity error estimator at later iteration steps of the greedy algorithm. When this low-fidelity error estimator is combined with $\Delta_l(\mu)$ in Algorithm 2, we obtain the multi-fidelity error estimation. This is detailed in Algorithm 4. Compared with the standard greedy algorithm, the overall saving in computational costs is noticeable, which can be seen from the numerical results in the next section.

The concept of multi-fidelity error estimation could also be applied to other high-fidelity error estimators. For example, Step 15 could be modified as ‘‘Stop updating partial information of $\Delta(\mu)$ ’’, if some parts of the high-fidelity error estimator $\Delta(\mu)$ are not ‘‘essential’’ for computing $\Delta(\mu)$.

Algorithm 4 Greedy algorithm with multi-fidelity error estimation

Input: the FOM, a training set Ξ_c composed of a small number of parameter samples taken from the parameter domain $\mu \in \mathcal{P}$, a set Ξ_f composed of a large number of parameter samples of μ from \mathcal{P} , error tolerance $\text{tol} < 1$.

Output: Projection matrix V .

- 1: Choose initial parameter $\mu^* \in \Xi_c$ for V , and initial parameter $\mu^r \neq \mu^* \in \Xi_c$ for V_r .
 - 2: $V \leftarrow \emptyset$, $V_r \leftarrow \emptyset$, $\epsilon = 1$.
 - 3: **while** $\epsilon > \text{tol}$ **do**
 - 4: Compute the snapshot(s) $x(\mu^*)$ by solving the FOM, i.e. $x(\mu^*) = (M(\mu^*))^{-1}B(\mu^*)$.
 - 5: Compute the snapshot(s) $x(\mu^r)$ by solving the FOM, i.e. $x(\mu^r) = (M(\mu^r))^{-1}B(\mu^r)$.
 - 6: Update V by $V = \text{orth}\{V, x(\mu^*)\}$ (e.g., using the modified Gram-Schmidt process with deflation).
 - 7: Update V_r by $V_r = \text{orth}\{V, V_r, x(\mu^r)\}$.
 - 8: Compute μ^* such that $\mu^* = \arg \max_{\mu \in \Xi_c} \tilde{\Delta}(\mu)$.
 - 9: Compute μ^o such that $\mu^o = \arg \min_{\mu \in \Xi_c} \tilde{\Delta}(\mu)$.
 - 10: Compute μ^r such that $\mu^r = \arg \max_{\mu \in \Xi_c} \max_{j=1, \dots, n_I} \|r_j(\mu) - M(\mu)\hat{x}_{r_j}(\mu)\|$, % n_I is the total number of columns of $B(\mu)$.
 - 11: Compute the low-fidelity error estimator $\tilde{\Delta}_l(\mu)$ using values of $\tilde{\Delta}(\mu)$ corresponding to the samples of μ in Ξ_c via (3) and (4).
 - 12: Evaluate $\tilde{\Delta}_l(\mu)$ over Ξ_f and pick out a parameter μ_c from the large parameter set Ξ_f corresponding to the largest value of $\tilde{\Delta}_l(\mu)$, i.e., $\mu_c = \arg \max_{\mu \in \Xi_f} \tilde{\Delta}_l(\mu)$.
 - 13: Update the small parameter set Ξ_c : if $\Delta_l(\mu_c) > \text{tol}$, enrich Ξ_c with μ_c , i.e., $\Xi_c = \{\Xi_c, \mu_c\}$, if $\Delta(\mu^o) < \text{tol}$, remove μ^o from Ξ_c : $\Xi_c = \Xi_c \setminus \mu^o$.
 - 14: $\epsilon = \tilde{\Delta}(\mu^*)$.
 - 15: **if** $\epsilon < \epsilon$ **then**
 - 16: Stop performing Step 5, Step 7 and Step 10. % stop updating the ROM of the residual system.
 - 17: **end if**
 - 18: **end while**
-

3.3 Application to MOR for time-delay systems

In this section, we consider applying Algorithm 2, the greedy algorithm with bi-fidelity error estimation, Algorithm 3, the standard greedy algorithm and Algorithm 4, the greedy algorithm with multi-fidelity error estimation to large-scale time-delay systems with many delays. The time-delay systems are defined as:

$$\begin{aligned} \sum_{j=0}^d E_j \dot{x}(t - \tau_j) &= \sum_{j=0}^d A_j x(t - \tau_j) + Bu(t), & \forall t \geq 0 \\ y(t) &= Cx(t), \end{aligned} \quad (16)$$

with an initial condition $x(t) = \Phi(t) \in \mathbb{C}^n, \forall t \in [-\tau_d, 0]$. Here, $E_0, \dots, E_d, A_0, \dots, A_d \in \mathbb{C}^{n \times n}, B \in \mathbb{C}^{n \times n_I}, C \in \mathbb{C}^{n_O \times n}, 0 = \tau_0 < \tau_1 < \dots < \tau_d$ and n is called the order of the delay system. The transfer function of the delay system is defined as:

$$H(s) = CK^{-1}(s)B, \quad (17)$$

where $\mathcal{K}(s) = s \sum_{j=0}^d E_j e^{-s\tau_j} - \sum_{j=0}^d A_j e^{-s\tau_j}$, $s = 2\pi f j$ is the variable in the frequency domain, f is the ordinary frequency with unit Hz and j is the imaginary unit.

A ROM of the delay system, which has the same delays as the original system, can be obtained via Galerkin projection using a projection matrix $V \in \mathbb{R}^{n \times r}, r \ll n$, i.e.,

$$\begin{aligned} \sum_{j=0}^d \hat{E}_j \dot{z}(t - \tau_j) &= \sum_{j=0}^d \hat{A}_j z(t - \tau_j) + \hat{B}u(t), & \forall t \geq 0, \\ \hat{y}(t) &= \hat{C}z(t), \end{aligned} \quad (18)$$

where $\hat{E}_j = V^T E_j V \in \mathbb{R}^{r \times r}, \hat{A}_j = V^T A_j V \in \mathbb{R}^{r \times r}, \hat{B} = V^T B \in \mathbb{R}^{r \times n_I}, \hat{C} = CV \in \mathbb{R}^{n_O \times r}$, with $r \ll n$ being the order of the ROM. The original state vector $x(t)$ in (16) can be recovered by the approximation: $x(t) \approx Vz(t)$. The transfer function of the ROM is

$$\hat{H}(s) = \hat{C}\hat{\mathcal{K}}^{-1}(s)\hat{B},$$

where $\hat{\mathcal{K}}(s) = s \sum_{j=0}^d \hat{E}_j e^{-s\tau_j} - \sum_{j=0}^d \hat{A}_j e^{-s\tau_j}$. The projection matrix V can be constructed via approximating $H(s)$ [5] as follows. Note that $H(s)$ is nothing but the output $y(\mu)$ of the linear parametric system in (5), with $M(\mu) = \mathcal{K}(s), B(\mu) = B$ and $\mu = s$, i.e.,

$$\begin{aligned} \mathcal{K}(s)x(s) &= B, \\ H(s) &= C(s)x(s). \end{aligned} \quad (19)$$

The reduced transfer function $\hat{H}(s)$ is the output $\hat{y}(\mu)$ of the ROM in (6) with $\hat{M}(\mu) = \hat{\mathcal{K}}(s)$ and $\hat{B}(\mu) = \hat{B}$.

It is easy to see that the projection matrix V that is used to construct the ROM (18) in the time domain is exactly the same matrix to obtain the reduced transfer function $\hat{H}(s)$. Therefore, V can be obtained by constructing a ROM of system (19) in the frequency domain, i.e., by approximating the transfer function $H(s)$. This can be done by the standard greedy Algorithm 3 with the error estimator $\hat{\Delta}(s)$, where V is iteratively computed by choosing proper samples of s [5, 1]. In fact, the reduced transfer function $\hat{H}(s)$ interpolates the original transfer function $H(s)$ at the selected samples of s [1]. The matrix $M(\mu)$ in Steps 4-5 of Algorithm 3 is now replaced by $\mathcal{K}(s)$. The difference of the coefficient matrix $\mathcal{K}(s)$ from a single matrix $M(\mu)$ in the usual case is its high complexity. To solve the system in (19) is much more expensive than solving the system in (5) where $M(\mu)$ is a single matrix.

On the one hand, the matrices constituting $\mathcal{K}(s)$ must be assembled to get $\mathcal{K}(s)$. On the other hand, the finally assembled matrix has some dense blocks, though each single matrix contributing to $\mathcal{K}(s)$ is sparse.

To further improve the efficiency of the standard greedy algorithm, we propose to apply Algorithm 2 and Algorithm 4 to time-delay systems. The application is straightforward by simply replacing the FOM in (5) in both algorithms with the system in (19), i.e., the matrix $M(\mu)$ is replaced by $\mathcal{K}(s)$, the input matrix $B(\mu)$ and the output matrix $C(\mu)$ are replaced by B and C in (19), respectively.

4 Numerical tests

We consider three time-delay systems obtained from partial element equivalent circuit (PEEC) modelling and simulation, which transfer problems from the electromagnetic domain to the circuit domain [29, 32, 30, 31]. When the propagation delays are explicitly kept for both partial inductances and coefficients of potential, time-delay systems can be derived [17]. Numerical tests are done with MATLAB R2016b on a computer server with 4 Intel Xeon E7-8837 CPUs running at 2.67 GHz, 1TB main memory, split into four 256 GB partitions.

We test the standard greedy Algorithm 3, the bi-fidelity greedy Algorithm 2 and the multi-fidelity Algorithm 4 on three time-delay systems. To run the algorithms, we need to initialize the algorithms by doing the following:

- The samples in the training set Ξ , the small set Ξ_c and the large set Ξ_f are taken from the prescribed frequency domain and are generated using the MATLAB function `linspace`: `linspace(fl, fh, cardi)`. Here, f_l is the lowest frequency, f_h is the highest frequency used in `linspace`, `cardi` is the corresponding cardinality of each set. The samples of s are then computed using the relation: $s = 2\pi j f$.
- For the multi-fidelity error estimation, we set $\epsilon = 0.1$ in Step 15 of Algorithm 4.
- To compute the low-fidelity error estimator, we choose the inverse multiquadratic RBF (IMQ) $\Phi = \frac{1}{1+(a\|\mu-\mu_i\|)^2}$ with the shape parameter $a = 30$.

We also need to define some variables uniformly used in all the tables and figures:

- The error $\|H(s) - \hat{H}(s)\|_{\max}$ of the transfer function $\hat{H}(s)$ of the ROM is finally computed over 1000 samples of s drawn independently of the training sets, resulting in the validated error: Valid.err in Tables 1-8.
- Runtime, the walltime of each algorithm till convergence.
- Iter., the total number of iterations of each algorithm.
- r , the order of the ROM.
- The high-fidelity error estimator at each iteration of Algorithm 3 is defined as $\max_{\mu \in \Xi} \tilde{\Delta}(\mu)$.
- The bi-fidelity error estimator at each iteration of Algorithm 2 is defined as $\max_{\mu \in \Xi_c} \tilde{\Delta}(\mu)$.
- The multi-fidelity error estimator at each iteration of Algorithm 4 is defined as $\max_{\mu \in \Xi_c} \tilde{\Delta}(\mu)$. Here $\tilde{\Delta}(\mu)$ will be different from the bi-fidelity error estimator once Step 15 of the algorithm takes action.

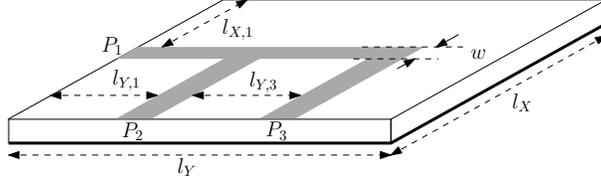


Figure 1: The three-port microstrip power-divider circuit.

- The true error at each iteration of Algorithm 3 is defined as $\max_{\mu \in \Xi} \|H(s) - \hat{H}(s)\|_{\max}$.
- The true error at each iteration of Algorithm 2 or Algorithm 4 is defined as $\max_{\mu \in \Xi_c} \|H(s) - \hat{H}(s)\|_{\max}$.

Note that Ξ_c could be enriched only by adding samples from Ξ_f to Ξ_c . As the high-fidelity error estimator $\tilde{\Delta}(\mu)$ needs to be computed at every sample in Ξ_c at each iteration, samples in Ξ_c whose corresponding error is already smaller than tol can also be removed from Ξ_c to keep the cardinality of Ξ_c constant, so that more computations can be saved. We consider both cases separately and compare their efficiency with respect to both runtime and accuracy.

4.1 Test 1: results for a model of three-port divider

The model structure of a three-port microstrip power-divider circuit is shown in Fig. 1 (P_1, P_2 and P_3 denote the ports). The dimensions of the circuit are $[20, 20, 0.5]$ mm in the $[x, y, z]$ directions and the width of the microstrips is set as 0.8 mm. Furthermore, the dimensions l_{X1}, l_{Y1} , and l_{Y3} are 9, 7.2 and 7.2 mm, respectively. The relative dielectric constant is $\epsilon_r = 2.2$. All the ports are terminated on 50Ω resistances. The order of the FOM is $n = 10,626$, and it has $d = 93$ delays. The interesting frequency band is $[0, 20]$ GHz.

For this model, we use $f_l = 1 \times 10^6$, $f_h = 2 \times 10^{10}$ in the function `linspace`. $|\Xi| = 30$ or $|\Xi| = 40$ for the standard greedy Algorithm 3. For Algorithm 2 and Algorithm 4, $|\Xi_c| = 15$ or $|\Xi_c| = 20$ and $|\Xi_f| = 100$. The set Ξ_c is then updated during the iteration of the greedy algorithm. The 1000 samples for validating the ROM accuracy are created using the MATLAB function `logspace`, i.e., `logspace(log10(fl), log10(fh), 1000)`. $f_{l_1} = 1 \times 10^4$.

In Table 1, we list the results of the three algorithms. The standard greedy algorithm is the slowest. The other algorithms are all much faster and take at least 2 hours less than the standard algorithm. The bi-fidelity greedy algorithm by enriching Ξ_c only is slower than other bi-(multi-)fidelity algorithms, this is in agreement with our theoretical analysis in Section 3. The multi-fidelity algorithm by adding and removing samples to and from Ξ_c performs the best in terms of runtime and accuracy. Compared to the standard algorithm, it has reduced the offline runtime from 5.6 hours to 1.8 hours, and almost 4 hours have been saved. Finally, a speed-up factor 3.1 is achieved. Except for the bi-fidelity algorithm by adding and removing samples, the other algorithms have produced ROMs with validated errors below the tolerance. The bi-fidelity algorithms perform similarly as the standard algorithm. All three algorithms converge in 14 iterations, and produce ROMs smaller than the others.

It is worth pointing out that if using fewer samples in Ξ for the standard greedy algorithm, the ROM has a validated error that is slightly larger than the tolerance, as shown in Table 2, where $|\Xi| = 30$. Also, the bi-fidelity greedy algorithms are less accurate if using fewer samples in Ξ_c , as shown in Table 2. There, the same Ξ_c used for the multi-fidelity greedy algorithms are used, but less accurate ROMs are obtained.

In Table 3, we show the results of the bi-fidelity greedy algorithm and the multi-fidelity greedy algorithm when $n_{\text{add}} = n_{\text{del}} > 1$ samples are added or removed from the small training set Ξ_c at each

Table 1: Three-port divider: $n = 10,626$, $d = 93$ delays, $\text{tol}=0.001$, adding/removing a single sample at each iteration.

Method	Iter.	Runtime (h)	r	Valid.err
Alg. 3 (standard, $ \Xi = 40$)	14	5.6	84	9.2×10^{-4}
Alg. 2 (bi-fidelity, add only, $ \Xi_c = 20$)	14	3.6	84	6×10^{-4}
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 20$)	14	2.7	84	0.0022
Alg. 4 (multi-fidelity, add only, $ \Xi_c = 15$)	15	2.4	90	6.2×10^{-4}
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 15$)	15	1.8	90	6.2×10^{-4}

Table 2: Three-port divider: $n = 10,626$, $d = 93$ delays, $\text{tol}=0.001$, smaller $|\Xi|$ and $|\Xi_c|$, adding/removing a single sample at each iteration.

Method	Iter.	Runtime (h)	r	Valid.err
Alg. 3 (standard, $ \Xi = 30$)	14	4.2	84	0.0017
Alg. 2 (bi-fidelity, add only, $ \Xi_c = 15$)	13	2.5	78	0.0026
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 15$)	13	1.9	78	0.0088

Table 3: Three-port divider: $n = 10,626$, $d = 93$ delays, $\text{tol}=0.001$, adding/removing $n_{\text{add}} = n_{\text{del}} > 1$ samples at each iteration.

Method	Iter.	Runtime (h)	r	Valid.err
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 15$, $n_{\text{add}} = 2$)	14	2.0	84	0.0022
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 20$, $n_{\text{add}} = 2$)	14	2.7	84	0.0022
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 20$, $n_{\text{add}} = 5$)	14	2.7	84	0.0022
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 15$, $n_{\text{add}} = 2$)	14	1.7	84	0.0039
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 15$, $n_{\text{add}} = 5$)	14	1.7	84	0.0039

iteration of the algorithm. In general, they produce similar results as those in Table 1 and Table 2 given the same Ξ_c . For $|\Xi_c| = 15$, the bi-fidelity greedy algorithm with $n_{\text{add}} = n_{\text{del}} = 2$ converges in 14 iterations, running one more iteration than with $n_{\text{add}} = n_{\text{del}} = 1$ as shown in Table 2, and generates a ROM with slightly higher accuracy. On the contrary, given $|\Xi_c| = 15$, the multi-fidelity greedy algorithm with either $n_{\text{add}} = n_{\text{del}} = 2$ or $n_{\text{add}} = n_{\text{del}} = 5$ runs one iteration less than in the case of adding/removing a single sample as shown in Table 1, and constructs ROMs with lower accuracy. Furthermore, it is seen that increasing $n_{\text{add}} = n_{\text{del}}$ from 2 to 5 did not change the results for both algorithms. In general, adding/removing a single sample keeps the algorithms simple but efficient.

To illustrate the behavior of the error estimators further, we plot the decay of error estimators and their corresponding true errors during the greedy iterations. Since different μ^* are chosen according to different error estimators, the projection matrix V is updated with different snapshots, leading to ROMs with different accuracy. Consequently, the true errors of the ROMs are expected to be different.

Figures 2-3 are the results of the algorithms in Table 1. The left part of Figure 2 shows the error of the high-fidelity error estimator at each iteration of Algorithm 3 and the decay of the corresponding true error. The error estimator almost exactly matches the true error at all the iterations. The right part of Figure 2 plots the decay of the bi-fidelity error estimator with respect to the true error. The bi-fidelity error estimator in both of the two cases: only adding (add-only) samples to Ξ_c , adding and removing (add-remove) samples to and from Ξ_c , can accurately catch the true error. Both cases converge in 14 iterations, but the case “add-only” is more accurate as can be seen from Table 1.

Figure 3 plots the decay of the multi-fidelity error estimator and the corresponding true error decay. For clarity, the two cases “add-only” and “add-remove” are plotted in two separate figures.

The multi-fidelity error estimator is not as accurate as the bi-fidelity error estimator. This is indicated by the error decay from the 10-th iteration to the end in both figures. From the 10-th iteration, the error estimator is below $\epsilon = 0.1$, the multi-fidelity error estimation at Step 15 of Algorithm 4 begins to be implemented. For this example, the multi-fidelity error estimator overestimates the true error more often than the bi-fidelity error estimator, it did not choose the interpolation points that lead to error decay as fast as those chosen by the bi-fidelity error estimator. Finally, it uses more iteration steps to converge. Whereas, they still produce ROMs with best accuracy.

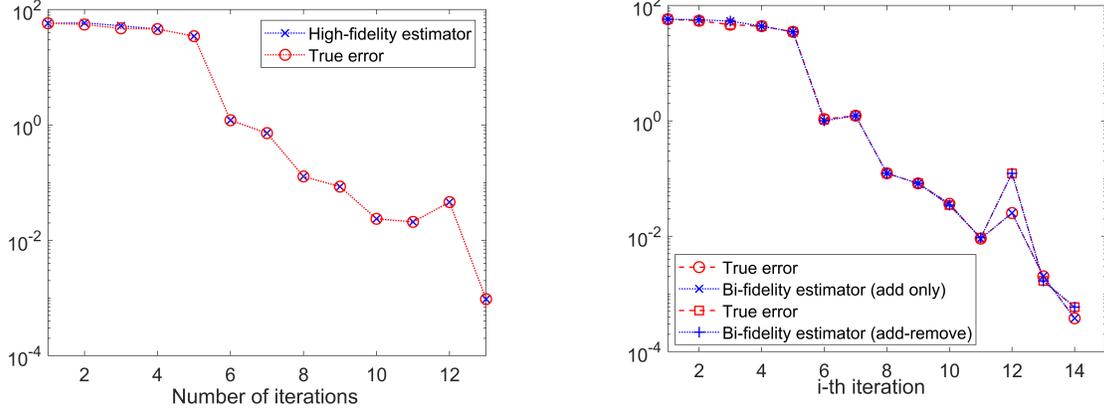


Figure 2: Error decay. Left: true error vs high-fidelity error estimator. Right: true error vs bi-fidelity error estimators.

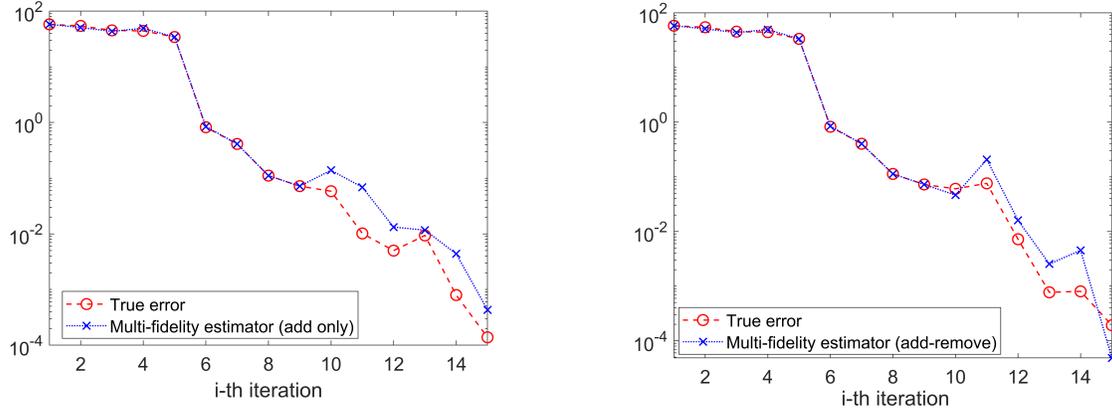


Figure 3: Error decay. Left: true error vs multi-fidelity error estimator by only adding samples to Ξ_c . Right: true error vs multi-fidelity error estimator by adding and deleting samples to and from Ξ_c .

4.2 Test 2: results for a model of coplanar microstrips

The second example is a model of a three coplanar microstrips structure shown in Fig. 4. The width of the metal strips is $m_w = 0.178$ mm, the thickness of metal strips and ground plane is $m_t = 0.035$ mm while the left and right wing of the microstrips are $w_d = 3$ mm. Finally, the length of each strip is $\ell = 5$ cm, the thickness of the dielectric is $d_t = 0.8$ mm, and the spacing between 2 strips is $s = 0.3$ mm. The relative dielectric constant is set to be $\epsilon_r = 4$ and the conductivity of the metal is assumed to be $\sigma = 5.8^7$ S/m. The six ports, located between the ends of each strip and the ground

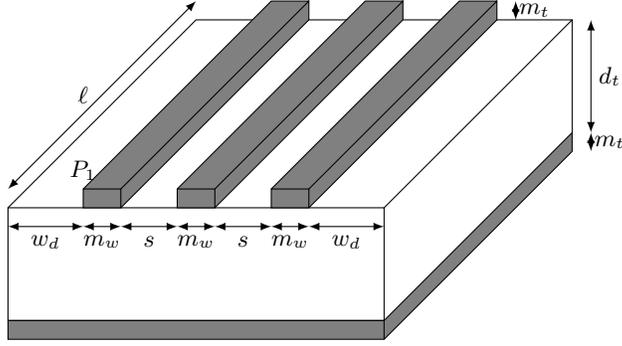


Figure 4: Three coplanar microstrips

plane below, are terminated on load resistors $R_{load} = 50 \Omega$. The order of the FOM is $n = 16,644$, and there are $d = 168$ delays. The frequency band of interest is $[0, 10]$ GHz.

For this model, we take $f_l = 1 \times 10^6$, $f_h = 1 \times 10^{10}$. We set 30 samples for Ξ in the standard greedy Algorithm 3, i.e., $|\Xi| = 30$. For Algorithm 2 and Algorithm 4, $|\Xi_c| = 10$ or $|\Xi_c| = 15$, and $|\Xi_f| = 100$. The 1000 samples used for validating the ROM accuracy are generated using the MATLAB function `linspace`, with $f_l = 100$ and the given f_h .

The results of the three algorithms are listed in Table 4. The standard greedy Algorithm 3 takes 19 hours, resulting in a ROM of order $r = 132$ with validated error below the tolerance tol . During the greedy iteration, if the small parameter set Ξ_c is enriched only (add only), the greedy algorithm with bi-fidelity error estimation and that with multi-fidelity error estimation converge within the same number of iterations, producing ROMs with the same sizes and validated errors. But the greedy algorithm with multi-fidelity error estimation is almost one hour faster. Similar phenomenon happens to the case “add-remove”. The greedy algorithm with bi-fidelity error estimation and that with multi-fidelity error estimation also converge within the same number of iterations and construct ROMs with the same sizes and accuracy. The runtimes of both algorithms are much less as compared to their “add only” versions. Finally, the greedy algorithm with multi-fidelity error estimation by adding and deleting samples to and from Ξ_c (“add-remove”) is most efficient in terms of both runtime and accuracy. It is more than 3 times faster than the standard greedy algorithm resulting in a speed-up of 4.2x, and produces a ROM with even the smallest validated error.

We note that using $|\Xi_c| = 10$, the ROMs constructed by the bi-fidelity greedy algorithm and the multi-fidelity greedy algorithm with adding the samples only have validated errors larger than the tolerance. If we increase $|\Xi_c|$ from 10 to 15, both algorithms generate ROMs with improved accuracy. The results are presented in Tabel 5. However, the computational time also increases a lot. Again, the multi-fidelity greedy algorithm outperforms the bi-fidelity one w.r.t. both accuracy and runtime. In contrast to the results in Tables 1-2 for the divider model, the results for the coplanar microstrips model in both Tables 4-5 show that the bi-fidelity greedy algorithm (“add-remove”) is more accurate than its “add-only” version.

Table 6 shows the results of the bi-fidelity greedy algorithm and the multi-fidelity greedy algorithm based on adding/removing multiple samples at each iteration. For both cases, i.e., $n_{add} = n_{del} = 2$ and $n_{add} = n_{del} = 5$, the algorithms using $|\Xi_c| = 10$, converge in 10 iterations, one less iteration than they did with $n_{add} = n_{del} = 1$ in Table 4, resulting in ROMs with smaller order r but with larger validated errors. If we increase $|\Xi_c|$ to 15, then the multi-fidelity greedy algorithm generates a ROM with reduced error, but takes longer time to converge. The bi-fidelity greedy algorithm behaves similarly and its results for $|\Xi_c| = 15$ is not presented to avoid repetition. This example again shows that adding/removing a single parameter at each iteration outperforms the cases with $n_{add} = n_{del} > 1$, and produces ROMs with desired accuracy.

Table 4: Three coplanar microstrips: $n = 16,644$, $d = 168$ delays, $\text{tol}=0.001$, adding/removing a single sample at each iteration.

Method	Iter.	Runtime (h)	r	Valid.err
Alg. 3 (standard, $ \Xi = 30$)	11	15	132	8.5×10^{-4}
Alg. 2 (bi-fidelity, add only, $ \Xi_c = 10$)	11	6.2	132	0.0033
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 10$)	11	5.3	132	8.2×10^{-4}
Alg. 4 (multi-fidelity, add only, $ \Xi_c = 10$)	11	5.3	132	0.0033
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 10$)	11	4.5	132	8.2×10^{-4}

Table 5: Three coplanar microstrips: $n = 16,644$, $d = 168$ delays, $\text{tol}=0.001$, larger $|\Xi_c|$, adding/removing a single sample at each iteration.

Method	Iter.	Runtime (h)	r	Valid.err
Alg. 2 (bi-fidelity, add only, $ \Xi_c = 15$)	11	10	132	0.0011
Alg. 4 (multi-fidelity, add only, $ \Xi_c = 15$)	12	9.3	144	4.4×10^{-4}

Table 6: Three coplanar microstrips: $n = 16,644$, $d = 168$ delays, $\text{tol}=0.001$, adding/removing $n_{\text{add}} = n_{\text{del}} > 1$ samples at each iteration.

Method	Iter.	Runtime (h)	r	Valid.err
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 10$, $n_{\text{add}} = 2$)	10	4.7	120	0.019
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 10$, $n_{\text{add}} = 5$)	10	4.7	120	0.019
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 10$, $n_{\text{add}} = 2$)	10	4.2	120	0.019
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 10$, $n_{\text{add}} = 5$)	10	4.3	120	0.019
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 15$, $n_{\text{add}} = 2$)	13	7.6	156	0.0027

In Figure 5, we show the *important* frequency samples of f selected by the greedy algorithms in Table 4. For the case “add-remove”, we find that the greedy algorithm with bi-fidelity error estimation and the one with multi-fidelity error estimation select the same *important* frequency samples. Therefore we only plot one group of samples for both algorithms, see the plot “bi-(multi-) add-remove” in the figure. For the case “add-only”, both algorithms also select the same *important* frequency samples, see the plot “bi-(multi-) add-only” in the figure. This is in agreement with the results given in Table 4 where both algorithms for either case produce the same results. The *important* frequency samples selected by the high-fidelity error estimator are mostly different from those selected by the other algorithms. It is seen that the *important* frequency samples selected by the (bi-)multi-fidelity estimator could be different from those selected by the high-fidelity estimator. However, both can derive ROMs with good accuracy.

The left part of Figure 6 gives the error-peak frequencies detected by the multi-fidelity error estimator and the true error, respectively, at each iteration of the greedy algorithm. Those frequencies correspond to the largest values of the error estimator/true error. The error-peak frequency detected by the error estimator at the i -th iteration is then selected as the *important* frequency sample at the next iteration to update the reduced basis space. From iteration 5, the error-peak frequencies detected by the error estimator are exactly the same as those selected by the true error. This can be explained by the error decay in the right part of the figure. From the 5-th iteration, the error estimator tightly catches the true error. Although it is less tight at the first 4 iterations, it still follows the overall trend of the error decay and therefore, can still detect reasonable error-peak frequencies. This example, once again, supports our theoretical analysis and demonstrates the efficacy of the proposed greedy algorithms with bi-(multi-) fidelity error estimation.

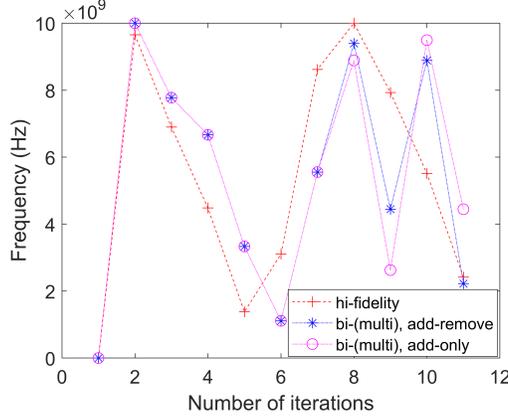


Figure 5: *Important* parameters selected by the greedy algorithms.

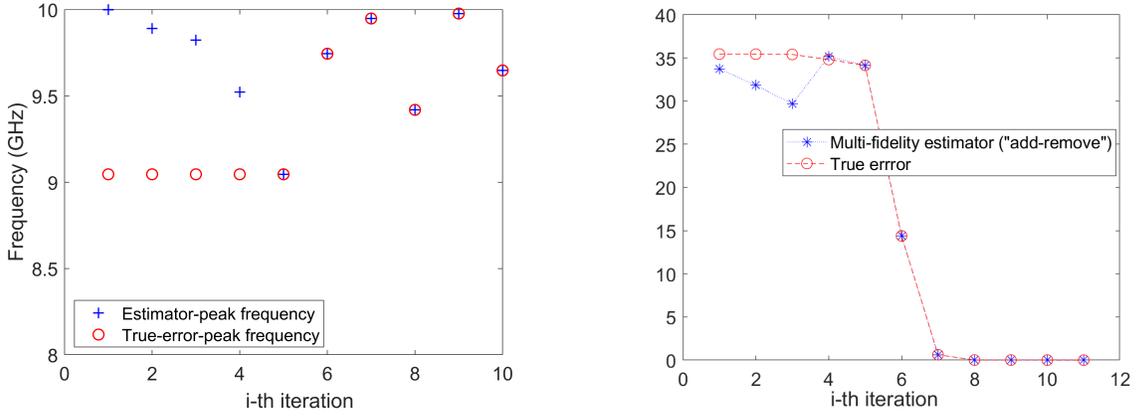


Figure 6: Left: Frequencies causing error/estimator peaks. Right: true error vs multi-fidelity error estimator.

4.3 Test 3: results for a model of microstrip filter

The third example is a model of a microstrip filter. The 3D structure of a microstrip filter is depicted in Fig. 7. The physical dimensions for the geometry of the 3D structure are: $w_{z_l} = 0.5$ mm, $w_{z_0} = 1.125$ mm, $w_{z_C} = 4$ mm, $\ell_{z_l} = 18.3$ mm, $\ell_{z_0} = 1$ mm, $\ell_{z_C} = 14.1$ mm, $w = 2.4$ cm, $\ell = 2\ell_{z_l} + 2\ell_{z_0} + \ell_{z_C}$, $t_m = 100$ μm , $t_s = 100$ μm , $t_d = 508$ μm . The two ends of the microstrip are terminated on 50Ω resistors. The order of the FOM is $n = 12,132$, and there are $d = 190$ delays. The interesting frequency band is $[0, 5]\text{GHz}$.

We take $f_l = 1 \times 10^5$, $f_h = 5 \times 10^9$ to generate frequency samples in Ξ_c and Ξ . We use $|\Xi| = 30$ for the standard greedy Algorithm 3. For Algorithm 2 and Algorithm 4, $|\Xi_c| = 10$, and $|\Xi_f| = 100$. The 1000 samples used for computing the validated error are generated using the MATLAB function `logspace`, with $f_l = 10$ and the given f_h .

The results of the high-fidelity greedy algorithm, and the bi-(multi)-fidelity greedy algorithms by adding/removing a single sample at each iteration, are listed in Table 7. All the bi-(multi)-fidelity greedy algorithms produce similar results. The runtime of each is around 1 hour, 3 hours faster than the high-fidelity greedy algorithm. All the ROMs have similar accuracy, with validated errors below the tolerance.

Table 8 further shows the performance of the bi-(multi)-fidelity greedy algorithms by adding and removing multiple samples at each iteration. For this model, all these algorithms behave similarly as

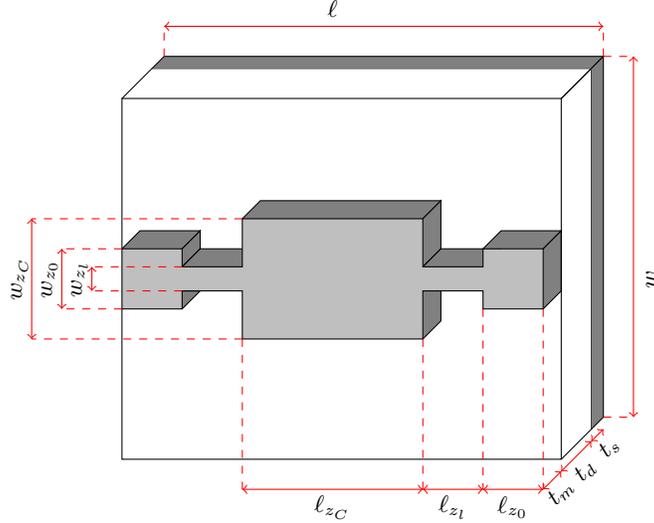


Figure 7: Microstrip filter.

they did by adding/removing a single sample at each iteration. The multi-fidelity greedy algorithm produces ROMs with slightly larger sizes. The ROMs also have larger validated errors, but still fulfill the accuracy requirement. All algorithms converge within 8 iterations, much faster than for the first two examples. This may be due to the much smaller frequency band of interest $[0, 5]$ GHz making the problem much easier to solve and leading to the most efficient performance of all algorithms.

In summary, for all the tested examples, the multi-fidelity algorithm by adding/removing a single sample at each iteration behaves the best w.r.t. both runtime and accuracy.

Table 7: Microstrip filter: $n = 12, 132$, $d = 190$ delays, $\text{tol}=0.001$, adding/removing a single sample at each iteration.

Method	Iter.	Runtime (h)	r	Valid.err
Alg. 3 (standard, $ \Xi = 30$)	8	2.5	32	5.6×10^{-4}
Alg. 2 (bi-fidelity, add only, $ \Xi_c = 10$)	7	1.1	28	4.6×10^{-4}
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 10$)	7	1.1	28	4.6×10^{-4}
Alg. 4 (multi-fidelity, add only, $ \Xi_c = 10$)	7	1.0	28	4.6×10^{-4}
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 10$)	8	1.1	32	5.7×10^{-4}

Table 8: Microstrip filter: $n = 12, 132$, $d = 190$ delays, $\text{tol}=0.001$, adding/removing $n_{\text{add}} = n_{\text{del}} > 1$ samples at each iteration.

Method	Iter.	Runtime (h)	r	Valid.err
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 10$, $n_{\text{add}} = 2$)	7	1.1	28	4.6×10^{-4}
Alg. 2 (bi-fidelity, add-remove, $ \Xi_c = 10$, $n_{\text{add}} = 5$)	7	1.1	28	4.6×10^{-4}
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 10$, $n_{\text{add}} = 2$)	8	1.1	32	9.1×10^{-4}
Alg. 4 (multi-fidelity, add-remove, $ \Xi_c = 10$, $n_{\text{add}} = 5$)	8	1.1	32	9.1×10^{-4}

5 Conclusions

Concepts of bi-fidelity error estimation and multi-fidelity error estimation are proposed in this work. The concept of bi-fidelity error estimation is general and can be applied to any high-fidelity estimator. Although the multi-fidelity error estimation is dependent on the high-fidelity error estimation in consideration, the framework is general to a certain extent and could also be combined with other high-fidelity error estimators. The robustness of the proposed greedy algorithms with bi-fidelity and multi-fidelity error estimation is tested on three large time-delay systems with many delays. Although the standard greedy algorithm converges in a few iterations, the computational complexity in each iteration is high. As a consequence, the runtime is long for such systems. The proposed (bi-)multi-fidelity greedy processes have significantly accelerated the standard greedy algorithm with no loss of accuracy in most cases.

References

- [1] D. Alfke, L. Feng, L. Lombardi, G. Antonini, and P. Benner. Model order reduction for delay systems by iterative interpolation. *Internat. J. Numer. Methods Engrg.*, 122(3):684–706, 2021.
- [2] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*, volume 6 of *Adv. Des. Control*. SIAM Publications, Philadelphia, PA, 2005.
- [3] A. C. Antoulas, C. A. Beattie, and S. Gugercin. *Interpolatory Methods for Model Reduction*. Computational Science & Engineering. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2020.
- [4] U. Baur, P. Benner, and L. Feng. Model order reduction for linear and nonlinear systems: A system-theoretic perspective. *Arch. Comput. Methods Eng.*, 21(4):331–358, 2014.
- [5] C. A. Beattie and S. Gugercin. Interpolatory projection methods for structure-preserving model reduction. *Systems Control Lett.*, 58(3):225–232, 2009.
- [6] A. Benaceur, V. Ehrlacher, A. Ern, and S. Meunier. Simultaneous empirical interpolation and reduced basis method for non-linear problems. *C. R. Acad. Sci. Paris*, 353(12):1105–1109, 2015.
- [7] A. Benaceur, V. Ehrlacher, A. Ern, and S. Meunier. A progressive reduced basis/empirical interpolation method for nonlinear parabolic problems. *SIAM J. Sci. Comput.*, 40(5):A2930–A2955, 2018.
- [8] P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors. *Model Reduction and Approximation: Theory and Algorithms*. Computational Science & Engineering. SIAM Publications, Philadelphia, PA, 2017.
- [9] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. Schilders, and L. M. Silveira, editors. *Model Order Reduction, Volume 1: System- and Data-Driven Methods and Algorithms*. De Gruyter, 2021.
- [10] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. Schilders, and L. M. Silveira, editors. *Model Order Reduction, Volume 2: Snapshot-Based Methods and Algorithms*. De Gruyter, 2021.
- [11] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. Schilders, and L. M. Silveira, editors. *Model Order Reduction, Volume 3: Applications*. De Gruyter, 2021.

- [12] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.*, 57(4):483–531, 2015.
- [13] S. Chellappa, L. Feng, and P. Benner. An adaptive sampling approach for the reduced basis method. In *Realization and Model Reduction of Dynamical Systems - A Festschrift in Honor of the 70th Birthday of Thanos Antoulas*, pages 137–155. Springer, Cham, 2022.
- [14] S. Chellappa, L. Feng, V. de la Rubia, and P. Benner. Adaptive interpolatory MOR by learning the error estimator in the parameter domain. In *Model Reduction of Complex Dynamical Systems*, volume 171 of *International Series of Numerical Mathematics*, pages 97–117. Birkhäuser, Cham, 2021.
- [15] L. Feng, A. C. Antoulas, and P. Benner. Some a posteriori error bounds for reduced order modelling of (non-)parametrized linear systems. *ESAIM: M2AN*, 51(6):2127–2158, 2017.
- [16] L. Feng and P. Benner. On error estimation for reduced-order modeling of linear non-parametric and parametric systems. *ESAIM: Math. Model. Numer. Anal.*, 55(2):561–594, 2021.
- [17] C. Gianfagna, L. Lombardi, and G. Antonini. Marching-on-in-time solution of delayed PEEC models of conductive and dielectric objects. *IET Microwaves, Antennas Propagation*, 13(1):42–47, 2019.
- [18] M. Grepl. *Reduced-basis approximation a posteriori error estimation for parabolic partial differential equations*. PhD thesis, Massachusetts Institute of Technology (MIT), Cambridge, USA, 2005.
- [19] M. A. Grepl. Certified reduced basis methods for nonaffine linear time-varying and nonlinear parabolic partial differential equations. *Math. Models Methods Appl. Sci.*, 22(3), 2012.
- [20] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM: Math. Model. Numer. Anal.*, 41(3):575–605, 2007.
- [21] M. A. Grepl and A. T. Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *M2AN Math. Model. Numer. Anal.*, 39:157–181, 2005.
- [22] D. Grunert, J. Fehr, and B. Haasdonk. Well-scaled, a-posteriori error estimation for model order reduction of large second-order mechanical systems. *Z. Angew. Math. Mech.*, 100(8):1–43, 2019.
- [23] B. Haasdonk and M. Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Math. Model. Numer. Anal.*, 42(2):277–302, 2008.
- [24] B. Haasdonk and M. Ohlberger. Efficient reduced models and a-posteriori error estimation for parametrized dynamical systems by offline/online decomposition. *Math. Comput. Model. Dyn. Syst.*, 17(2):145–161, 2011.
- [25] S. Hain, M. Ohlberger, M. Radic, and K. Urban. A hierarchical a-posteriori error estimator for the reduced basis method. *Advances in Computational Mathematics*, 45(2):2191–221, 2019.
- [26] A. Paul-Dubois-Taine and D. Amsallem. An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models. *Internat. J. Numer. Methods Engrg.*, 102(12):1262–1292, 2015.

- [27] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations*, volume 92 of *La Matematica per il 3+2*. Springer International Publishing, 2016. ISBN: 978-3-319-15430-5.
- [28] D. V. Rovas. *Reduced-Basis Output Bound Methods for Parametrized Partial Differential Equations*. PhD thesis, Massachusetts Institute of Technology (MIT), Cambridge, USA, 2003.
- [29] A. E. Ruehli. Inductance calculations in a complex integrated circuit environment. *IBM Journal of Research and Development*, 16(5):470–481, Sept. 1972.
- [30] A. E. Ruehli. Equivalent circuit models for three dimensional multiconductor systems. *IEEE Transactions on Microwave Theory and Techniques*, MTT-22(3):216–221, Mar. 1974.
- [31] A. E. Ruehli, G. Antonini, and L. Jiang. *Circuit Oriented Electromagnetic Modeling Using the PEEC Techniques*. Wiley-IEEE Press, 2017.
- [32] A. E. Ruehli and P. A. Brennan. Efficient capacitance calculations for three-dimensional multiconductor systems. *IEEE Transactions on Microwave Theory and Techniques*, 21(2):76–82, 1973.
- [33] A. Schmidt and B. Wittwar, D. Haasdonk. Rigorous and effective a-posteriori error bounds for nonlinear problems—application to RB methods. *Adv. Comput. Math.*, 46(32):30 pages, 2020.
- [34] K. Smetana, O. Zahm, and A. T. Patera. Randomized residual-based error estimators for parametrized equations. *SIAM J. Sci. Comput.*, 41(2):A900–A926, 2019.
- [35] K. Veroy, C. Prud’Homme, D. V. Rovas, and A. T. Patera. A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. In *16th AIAA Computational Fluid Dynamics Conference*, Orlando, United States, 2003.
- [36] Y. Zhang, L. Feng, S. Li, and P. Benner. An efficient output error estimation for model order reduction of parametrized evolution equations. *SIAM J. Sci. Comput.*, 37(6):B910–B936, 2015.