Advances in Computational Mathematics (2019) 45:2191–2214
https://doi.org/10.1007/s10444-019-09675-z

A hierarchical a posteriori error estimator for the Reduced Basis Method

Stefan Hain[1] · **Mario Ohlberger**[2] · **Mladjan Radic**[1] · **Karsten Urban**[1]

Received: 7 March 2018 / Accepted: 14 February 2019 /
Published online: 27 February 2019

Abstract

In this contribution, we are concerned with tight a posteriori error estimation for projection-based model order reduction of inf-sup stable parameterized variational problems. In particular, we consider the Reduced Basis Method in a Petrov-Galerkin framework, where the reduced approximation spaces are constructed by the (weak) greedy algorithm. We propose and analyze a hierarchical a posteriori error estimator which evaluates the difference of two reduced approximations of different accuracy. Based on the a priori error analysis of the (weak) greedy algorithm, it is expected that the hierarchical error estimator is sharp with efficiency index close to one, if the Kolmogorov N-with decays fast for the underlying problem and if a suitable saturation assumption for the reduced approximation is satisfied. We investigate the tightness of the hierarchical a posteriori estimator both from a theoretical and numerical perspective. For the respective approximation with higher accuracy, we study and compare basis enrichment of Lagrange- and Taylor-type reduced bases. Numerical experiments indicate the efficiency for both, the construction of a reduced basis using the hierarchical error estimator in a greedy algorithm, and for tight online certification of reduced approximations. This is particularly relevant in cases where the inf-sup constant may become small depending on the parameter. In such cases, a standard residual-based error estimator—complemented by the successive constrained method to compute a lower bound of the parameter dependent inf-sup constant—may become infeasible.

Keywords Reduced Basis Method · A posteriori error estimator · Hierarchical error estimator

Mathematics Subject Classification (2010) 65N30 · 65N15 · 65M15

Communicated by: Anthony Nouy

✉ Karsten Urban
karsten.urban@uni-ulm.de

Extended author information available on the last page of the article.

1 Introduction

Model order reduction has become a field of great significance, both with respect to solving real-world problems and with respect to mathematical research. In this article, we consider the Reduced Basis Method (RBM), which is a well-known projection- based model order reduction technique for Parameterized Partial Differential Equations (PPDEs), for instance in multi-query and/or real-time contexts [22, 24, 35]. The key idea for the RBM is to construct a problem-specific reduced order model, e.g., in a computationally expensive offline phase, and then use this reduced model to construct an approximation in an online phase extremely fast by solving very low-dimensional Petrov-Galerkin problems.

A posteriori error estimates play an important role within the RBM, at least for the following reasons: (1) The error estimator is used in a greedy algorithm to construct the reduced model. This is, e.g., done by maximizing the error estimator over a discrete number of reduced solutions with respect to a finite training set of parameters ("sampling") and to enrich the preliminary reduced basis by the truth solution ("snapshots") that corresponds to the worst approximated reduced solution. (2) After the online computation of a reduced approximation as a linear combination of the snapshots, an error estimator yields an upper bound for the error and thus certifies the reduced numerical approximation.

This shows that such error estimators need to satisfy a number of conditions: (i) The computation of the error estimator for some given parameter has to be very fast, i.e., with a complexity that only depends on the degrees of freedom of the reduced approximation space (for the basis generation, this allows a large and representative training set; in the online phase, the certification has to be at least as efficient as the computation of the reduced approximation itself); (ii) The error estimator has to be tight in order to yield an efficient and reliable estimate of the true error.

So far, the most common approach for constructing such a posteriori RB error estimators is residual-based. This usually involves an efficient computation of (an approximation of) the residual and the inverse of the inf-sup constant. As for many problems, the inf-sup constant cannot be computed or estimated in an efficient way, the Successive Constraint Method (SCM) [10, 11, 27] is used for the calculation of a lower bound. This involves at least two drawbacks, namely the computational complexity of the SCM, in particular if a very good approximation is needed and—related—the lower bound maybe very small (and thus almost useless for the residual-based error estimator) if the inf-sup constant is small. Moreover, we have observed numerically that the SCM may not always converge (see Section 4). We also mention the ROMES method [16] using an offline learning approach involving any kind of error indicator, hence, in principle also the hierarchical error estimate.

Hierarchical error estimators use the difference of two approximations of different orders to bound the unknown error. This approach is well-known, e.g., for ordinary differential equations [34] and adaptive finite elements [3, 12, 15, 26, 42, 43], just to mention a few. Within the RBM, such an approach has been used to measure the error

of the empirical interpolation method (EIM) [4, 9, 17]. However, to the very best of our knowledge, we are not aware of an article investigating its use for a posteriori error estimation for RB approximations.

We investigate two situations: (a) A family of reduced spaces $(X_N)_{N=1,\dots,N_{\max}}$ is given. Then, we choose $N < M$ and use the difference $\|u_N - u_M\|_X$ of two RB approximations as error estimator in the online phase. We study the performance in particular in those cases, where the inf-sup constant is small or hard to access numerically. This is, e.g., the case for the Helmholtz problem, where the inf-sup constant behaves like $\mu^{-7/2}$, the wave number $\mu \in \mathbb{R}^+$ being the parameter (see [19]). Other examples (that will not be treated here) include transport and wave propagation problems, where one can may construct an optimal reduced space in a possibly costly offline stage but cannot use the residual online, since it cannot be computed efficiently [6, 21]. (b) A residual-based error estimator cannot be used at all. In this case, one would like to construct the reduced basis with the aid of the hierarchical error estimator. This, however, is not completely straightforward, since X_M needs to be constructed for given X_N. It turns out that a standard greedy procedure may not work in this case. This is the reason why we suggest to use a Taylor-type RB approach for constructing the reduced space of higher accuracy. Numerical experiments are given to demonstrate the efficiency of the resulting approach.

In both cases, (a) and (b), we investigate the effectivity of the hierarchical error estimator, both theoretically and numerically. For the latter purpose, we suggest an offline procedure to determine sharp estimates for the effectivity that can also be used in the online stage.

The remainder of this paper is organized as follows: In Section 2, we collect some preliminaries on PPDEs and RBMs. Section 3 is devoted to the introduction of the hierarchical error estimator including the analysis and realization. We report on several numerical experiments in Section 4 for the standard thermal block problem and the Helmholtz problem in a high-frequency regime, i.e., with quite small inf-sup constants. We mention that (based upon the preprint version of this paper) our RB hierarchical error estimate has recently been used in the scope of other problems [6, 20, 21].

2 Preliminaries

In this section, we collect the main facts and background material that is used in the sequel.

2.1 Parameterized Partial Differential Equations

Let $\mathcal{P} \subset \mathbb{R}^P$, $P \in \mathbb{N}$, be a compact parameter space. For suitable Hilbert (function) spaces X and Y consider the parameterized variational problem (e.g., a PDE):

$$\text{For } \mu \in \mathcal{P} \text{ find } u(\mu) \in X : \; a(u(\mu), v; \mu) = f(v; \mu) \qquad \forall v \in Y, \tag{2.1}$$

where $a : X \times Y \times \mathcal{P} \to \mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ is a continuous sesquilinear form and $f : Y \times \mathcal{P} \to \mathbb{K}$ is a given continuous linear form. For ensuring the uniform well-posedness of (2.1) for any $\mu \in \mathcal{P}$, we assume (as usually done) that

$$\forall \mu \in \mathcal{P}\colon \sup_{u\in X}\sup_{v\in Y}\frac{|a(u,v;\mu)|}{\|u\|_X\|v\|_Y} \le \gamma(\mu) \le \gamma_{\mathrm{UB}} < \infty, \quad \text{(continuity)}$$

$$\forall \mu \in \mathcal{P}\colon \inf_{u\in X}\sup_{v\in Y}\frac{|a(u,v;\mu)|}{\|u\|_X\|v\|_Y} \ge \beta(\mu) \ge \beta_{\mathrm{LB}} > 0, \quad (\inf-\sup \text{condition})$$

$$\forall \mu \in \mathcal{P}\colon \inf_{v\in Y}\sup_{u\in X}\frac{|a(u,v;\mu)|}{\|u\|_X\|v\|_Y} > 0, \quad \text{(surjectivity)}.$$

Even though these assumptions yield a *uniform* well-posedness (w.r.t. the parameter), we note that particularly β_{LB} may be fairly small, which will be crucial below.

2.2 The "Truth"

Next, we require the availability of a detailed or fine discretization in terms of suitable conforming trial and test spaces $X^{\mathcal{N}} \subset X$ and $Y^{\mathcal{N}} \subset Y$, where (just for simplicity) $\dim(X^{\mathcal{N}}) = \dim(Y^{\mathcal{N}}) = \mathcal{N} < \infty$. The discretized parameterized problem then reads for any $\mu \in \mathcal{P}$:

$$\text{Find } u^{\mathcal{N}}(\mu) \in X^{\mathcal{N}}\colon\ a^{\mathcal{N}}(u^{\mathcal{N}}(\mu), v^{\mathcal{N}};\mu) = f^{\mathcal{N}}(v^{\mathcal{N}};\mu) \quad \forall v^{\mathcal{N}} \in Y^{\mathcal{N}}, \tag{2.2}$$

where $a^{\mathcal{N}} : X^{\mathcal{N}} \times Y^{\mathcal{N}} \times \mathcal{P} \to \mathbb{K}$ and $f^{\mathcal{N}} : Y^{\mathcal{N}} \times \mathcal{P} \to \mathbb{K}$ are appropriate discrete sesquilinear and linear forms. The discrete sesquilinear and linear forms are continuous with the same constants. To ensure the uniform well-posedness of (2.2) for every $\mu \in \mathcal{P}$, we require (as usual)

$$\begin{aligned}
\forall \mu \in \mathcal{P}\colon \sup_{u^{\mathcal{N}}\in X^{\mathcal{N}}}\sup_{v^{\mathcal{N}}\in Y^{\mathcal{N}}}\frac{|a^{\mathcal{N}}(u^{\mathcal{N}}, v^{\mathcal{N}};\mu)|}{\|u^{\mathcal{N}}\|_{X^{\mathcal{N}}}\|v^{\mathcal{N}}\|_{Y^{\mathcal{N}}}} &= \gamma^{\mathcal{N}}(\mu) \le \gamma^{\mathcal{N}}_{\mathrm{UB}} < \infty,\\
\forall \mu \in \mathcal{P}\colon \inf_{u^{\mathcal{N}}\in X^{\mathcal{N}}}\sup_{v^{\mathcal{N}}\in Y^{\mathcal{N}}}\frac{|a^{\mathcal{N}}(u^{\mathcal{N}}, v^{\mathcal{N}};\mu)|}{\|u^{\mathcal{N}}\|_{X^{\mathcal{N}}}\|v^{\mathcal{N}}\|_{Y^{\mathcal{N}}}} &= \beta^{\mathcal{N}}(\mu) \ge \bar{\beta}^{\mathcal{N}}_{\mathrm{LB}} > 0.
\end{aligned} \tag{2.3}$$

Note, that the discrete surjectivity condition is ensured by (2.3) since the spaces $X^{\mathcal{N}}$ and $Y^{\mathcal{N}}$ are finite dimensional. Here, $\|\cdot\|_{X^{\mathcal{N}}}$ and $\|\cdot\|_{Y^{\mathcal{N}}}$ may be numerical approximations to $\cdot\|_X$ and $\|\cdot\|_Y$, respectively, but may also be discrete norms (such as for discontinuous Galerkin (dG) methods). Such a detailed discretization can, e.g., arise from finite element, finite volume, dG or spectral element discretizations.

It is a standard assumption that this detailed discretization is sufficiently fine so that the error $\|u(\mu) - u^{\mathcal{N}}(\mu)\|_X$ is negligible, which is the reason why $u^{\mathcal{N}}(\mu)$ is often called the "truth." In particular, we assume here that $X^{\mathcal{N}}$ and $Y^{\mathcal{N}}$ are the same for all parameters, but mention that adaptive discretizations may also be used (cf. [1, 23]).

2.3 The Reduced Basis Method

We briefly recall the main ingredients of the Reduced Basis Method (RBM) which we need here and refer, e.g., to [22, 24, 35] for more details. The aim of the RBM is

to determine a highly reduced model of size $N \ll \mathcal{N}$ in terms of reduced trial and test spaces $X_N \subset X^{\mathcal{N}}$, $Y_N \subset Y^{\mathcal{N}}$. Such a reduced model is typically determined in an offline phase, which might be computationally costly. This is done by selecting certain parameters $S_N := \{\mu_1, \ldots, \mu_N\}$, computing the corresponding (truth) snapshots $\xi_i := u^{\mathcal{N}}(\mu_i)$, $i = 1, \ldots, N$, and setting $X_N := \operatorname{span}\{\xi_1, \ldots, \xi_N\}$, $N \ll \mathcal{N}$. The basis may be orthonormalized for stability reasons.

The choice of the snapshot parameter set S_N is usually based upon an efficiently computable a posteriori error estimator $\Delta_N(\mu)$ which is then maximized in a greedy manner over a finite training set $\mathcal{P}_{\text{train}} \subset \mathcal{P}$. This approach is called here *weak greedy*.[1] Sometimes, the error is used instead of an error estimator, which is then termed as *strong greedy*. Other approaches such as nonlinear optimization of an error estimator have also been investigated (e.g., [38]).

In order to ensure well-posedness of the reduced problem, namely:

$$\text{For} \mu \in \mathcal{P} \text{ find } u_N(\mu) \in X_N\text{: } a^{\mathcal{N}}(u_N(\mu), v_N; \mu) = f^{\mathcal{N}}(v_N; \mu) \quad \forall v_N \in Y_N, \tag{2.4}$$

the spaces X_N and Y_N have to be chosen such that there exists an $0 < \beta_{\text{LB}}^{\mathcal{N}} \leq \bar{\beta}_{\text{LB}}^{\mathcal{N}}$

$$\inf_{w_N \in X_N} \sup_{v_N \in Y_N} \frac{|a^{\mathcal{N}}(w_N, v_N; \mu)|}{\|w_N\|_{X^{\mathcal{N}}} \|v_N\|_{Y^{\mathcal{N}}}} =: \beta_N^{\mathcal{N}}(\mu) \geq \beta_{\text{LB}}^{\mathcal{N}} > 0, \qquad \mu \in \mathcal{P}. \tag{2.5}$$

Let $u_N(\mu) = \sum_{i=1}^N u_{i,N}(\mu)\xi_i$ be the desired expansion of the RB approximation. It is easily seen that the unknown coefficient vector $\mathbf{u}_N(\mu) = (u_{i,N}(\mu))_{i=1}^N$ arises from solving a linear system of equations $\mathbb{A}_N(\mu)\mathbf{u}_N(\mu) = \mathbb{F}_N(\mu)$, where $(\mathbb{A}_N(\mu))_{i,j} := a^{\mathcal{N}}(\xi_i, \psi_j; \mu)$, $(\mathbb{F}_N(\mu))_j := f^{\mathcal{N}}(\psi_j; \mu)$, and $Y_N := \operatorname{span}\{\psi_1, \ldots, \psi_N\}$ is a stabilized reduced test space (see, e.g., [35]). Typically, $\mathbb{A}_N(\mu)$ is a dense matrix so that the reduced approximation can be computed with $\mathcal{O}(N^3)$ operations. This complexity is independent of the truth dimension $\mathcal{N}$, which is the reason to call it *online efficient*. In order to setup the linear system in an online efficient manner, we assume (as usual) that sesquilinear and linear forms are separable w.r.t. the parameter, i.e.,

$$a^{\mathcal{N}}(w, v; \mu) = \sum_{q=1}^{Q^a} \vartheta_q^a(\mu) a_q^{\mathcal{N}}(w, v), \quad \mu \in \mathcal{P}, w \in X^{\mathcal{N}}, v \in Y^{\mathcal{N}}, \tag{2.6}$$

$$f^{\mathcal{N}}(v; \mu) = \sum_{q=1}^{Q^f} \vartheta_q^f(\mu) f_q^{\mathcal{N}}(v), \quad \mu \in \mathcal{P}, v \in Y^{\mathcal{N}}. \tag{2.7}$$

Sometimes (2.6) is also called *affine decomposition*. If (2.6) is not satisfied, the empirical interpolation method can be used to construct an affine approximation (see, e.g., [4]). Using (2.6), one can precompute parameter-independent quantities in the offline stage allowing for an online efficient setup of the linear system. In fact, the

[1] Note, that this terminology differs from the rigorous error analysis in [5], where weak greedy algorithms use error estimators which can be proven to be equivalent to the true error with precise constants. We call an algorithm "weak" greedy if an error estimator is used for the basis construction in a greedy manner.

parameter-independent matrices $(\mathbb{A}_N^q)_{j,i} := a_q^{\mathcal{N}}(\xi_i, \psi_j)$, $i, j = 1, \ldots, N$, $q = 1, \ldots, Q^a$ and vectors $(\mathbb{F}_N^q)_j := f_q^{\mathcal{N}}(\psi_j)$, $j = 1, \ldots, N$, $q = 1, \ldots, Q^f$ can be computed offline and stored once. Then, for a given new parameter $\mu \in \mathcal{P}$

$$\mathbb{A}_N(\mu) = \sum_{q=1}^{Q^a} \vartheta_q^a(\mu)\mathbb{A}_N^q, \qquad \mathbb{F}_N(\mu) = \sum_{q=1}^{Q^f} \vartheta_q^f(\mu)\mathbb{F}_N^q,$$

which is of complexity $\mathcal{O}(Q^a N^2)$ and $\mathcal{O}(Q^f N)$, respectively. As the complexity does not depend on $\mathcal{N}$, it is *online efficient*.

The best possible rate of convergence for the error is given by the decay of the *Kolmogorov N-width*

$$d_N(\mathcal{P}) := \inf_{\dim(X_N)=N, X_N \subset X} \sup_{\mu\in\mathcal{P}} \inf_{v_N \in X_N} \|u(\mu) - v_N\|_X. \tag{2.8}$$

It is known that $d_N(\mathcal{P})$ decays fast (even exponentially) for elliptic PPDEs as $N \to \infty$ with smooth dependence of the solution on the parameter (see e.g. [5, 13, 30]).

2.4 The residual-based a posteriori error estimator

As already mentioned above, an online efficient error estimator $\Delta_N(\mu)$ is often used within a weak greedy procedure to determine the snapshot index set S_N. Moreover, such a $\Delta_N(\mu)$ is used for online certification by computing an upper bound for the error induced by the RB approximation $u_N(\mu)$. In this paper, we will consider two examples for such a $\Delta_N(\mu)$. For the subsequent analysis, we will consider

$$e_N(\mu) := \|u(\mu) - u_N(\mu)\|_X, \qquad e_N^{\mathcal{N}}(\mu) := \|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}$$

which will be termed *exact error* and *truth error*, respectively. Also other error quantities or functions of the error can be considered using adjoint methods. It is fairly standard to use the (truth) *residual* $R_N^{\mathcal{N}}(\cdot;\mu) \in (Y^{\mathcal{N}})'$ defined as

$$R_N^{\mathcal{N}}(w;\mu) := f^{\mathcal{N}}(w;\mu) - a^{\mathcal{N}}(u_N(\mu), w;\mu) = a^{\mathcal{N}}(e_N^{\mathcal{N}}(\mu), w;\mu), \quad w \in Y^{\mathcal{N}},$$

to define the residual based a posteriori RB error estimator as follows

$$\Delta_N^{\mathrm{Std}}(\mu) := \frac{\|R_N^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}}{\beta^{\mathcal{N}}(\mu)},$$

which we will call *standard RB error estimator* in the sequel. It should be noted that the (truth) residual also admits an affine decomposition and can thus in fact be computed online efficient. The involved (truth) inf-sup constant $\beta^{\mathcal{N}}(\mu)$ can only be determined exactly in very specific cases. Usually, a lower bound $\beta_{\mathrm{LB}}^{\mathcal{N}}(\mu)$ is computed for example by the *Successive Constraint Method* (SCM), [11, 25, 27]. However, even though the SCM is online efficient, the quantitative performance may be a severe problem in real-time applications, in particular if a good approximation of $\beta^{\mathcal{N}}(\mu)$ is required (which is the case, e.g., if $\beta^{\mathcal{N}}(\mu)$ is small).

The relation of the truth error and the residual is well-known and easily seen

$$\frac{1}{\gamma^{\mathcal{N}}(\mu)}\|R_N^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'} \le \|e_N^{\mathcal{N}}(\mu)\|_{X^{\mathcal{N}}} \le \frac{1}{\beta^{\mathcal{N}}(\mu)}\|R_N^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}. \tag{2.9}$$

Note, that this relation is w.r.t. the truth error, not w.r.t. the exact error [1, 31, 32, 40, 41]. Of course, one can replace $\beta^{\mathcal{N}}(\mu)$ and $\gamma^{\mathcal{N}}(\mu)$ in (2.9) by lower and upper bounds $\beta_{\mathrm{LB}} > 0$, $\gamma_{\mathrm{UB}} < \infty$, respectively, even though these bounds may be numerically infeasible. Under the assumptions of the previous sections, it has been proven that weak greedy algorithms exhibit the same rate of convergence as $d_N(\mathcal{P})$ if there exists rigorous lower and upper bounds for the error, like (2.9) (see [5, 7]). Roughly speaking, the RBM works well for a PPDE if $d_N(\mathcal{P})$ decays sufficiently fast as N grows.

3 A hierarchical error estimator

In this section, we introduce the hierarchical error estimator. To this end, let $X_N \subsetneq X_M \subset X^{\mathcal{N}}$, where $\dim(X_M) = M > N = \dim(X_N)$, and $u_N(\mu) \in X_N$, $u_M(\mu) \in X_M$, respectively. Then, we define the hierarchical error estimator by

$$\Delta_{N,M}(\mu) := \|u_M(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}, \tag{3.1}$$

3.1 Error analysis

The analysis of hierarchical error estimators is pretty standard in various applications for ODEs or PDEs. Due to the specific framework of parameter-dependent problems, we detail it here. We indicate two approaches.

3.1.1 Asymptotic analysis

Using triangle inequality, we get by (2.9) and (2.5)

$$\begin{aligned}
\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} &\leq \|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}} + \|u_M(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} \\
&= \|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}} + \Delta_{N,M}(\mu) \\
&\leq \frac{1}{\beta^{\mathcal{N}}(\mu)} \|R_M^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'} + \Delta_{N,M}(\mu) \\
&= \Delta_M^{\mathrm{Std}}(\mu) + \Delta_{N,M}(\mu).
\end{aligned}$$

Now, we recall from [5] that one can construct X_M in such a way that $\Delta_M^{\mathrm{Std}}(\mu) \to 0$ as $M \to \infty$ for every $\mu \in \mathcal{P}$ provided that the Kolmogorov M-width decays, i.e., this is a term of higher order. This means that for any N and $\varepsilon > 0$, we can choose an $M = M(\varepsilon) > N$ such that

$$\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} \leq \varepsilon + \Delta_{N,M}(\mu).$$

Alternatively, we can choose M such that $\Delta_M^{\mathrm{Std}}(\mu) \leq \varepsilon \Delta_{N,M}(\mu)$ yielding that

$$\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} \leq (1 + \varepsilon) \cdot \Delta_{N,M}(\mu).$$

If, however, the assumption $\Delta_M^{\mathrm{Std}}(\mu) \leq \varepsilon \Delta_{N,M}(\mu)$ is only satisfied on a training set $\mathcal{P}_{\mathrm{train}} \subset \mathcal{P}$, there might exist parameters $\mu \in \mathcal{P} \setminus \mathcal{P}_{\mathrm{train}}$ with $\Delta_{N,M}(\mu) = 0$ for all M,

but $\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} \neq 0$. This may happen if X_M does not converge to $X^{\mathcal{N}}$, which motivates a further assumption.

3.1.2 Saturation assumption

A way to analyze hierarchical error estimates is by showing or assuming a guaranteed error decay, typically called *saturation property* (see, e.g., [3, 26, 39]). In order to formulate it, we recall that the reduced spaces $X_N := \operatorname{span}\{\xi_1, \ldots, \xi_N\}$, $N \ll \mathcal{N}$ are formed by snapshots $\xi_i := u^{\mathcal{N}}(\mu_i)$, $i = 1, \ldots, N$. Consider now a second reduced basis space X_M with $\dim(X_M) =: M > N := \dim(X_N)$. Then, we say that X_N and X_M satisfy the *saturation property*, if there exists a constant $\Theta_{N,M}^{\mathcal{N}} \in (0, 1)$, s.t.

$$\|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}} \leq \Theta_{N,M}^{\mathcal{N}} \cdot \|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} \tag{3.2}$$

holds for all $\mu \in \mathcal{P}$. We will show a numerical procedure to validate this assumption below. At this point, we do *not* specify the particular construction of X_M (see Section 3.4 below). Then, following standard lines (see, e.g., [3, 26, 39]), we can easily prove the following estimates.

Proposition 3.1 *If (3.2) holds, then*

$$\frac{\Delta_{N,M}(\mu)}{1 + \Theta_{N,M}^{\mathcal{N}}} \leq \|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} \leq \frac{\Delta_{N,M}(\mu)}{1 - \Theta_{N,M}^{\mathcal{N}}} =: \Delta_{N,M}^{\text{Hier}}(\mu). \tag{3.3}$$

Proof For $\mu \in \mathcal{P}$ with $\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} = 0$, (3.2) yields $\Delta_{N,M}^{\text{Hier}} = 0$, so that the inequalities are obviously fulfilled. If $\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} \neq 0$, we use the reverse triangle inequality and the saturation property to obtain

$$\begin{aligned}
\frac{\|u_M(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}}{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}} &\geq \frac{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}} - \|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}}{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}} \\
&= 1 - \frac{\|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}}{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}} \\
&\geq 1 - \sup_{\mu \in \mathcal{P}} \frac{\|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}}{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}} \\
&\geq 1 - \Theta_{N,M}^{\mathcal{N}},
\end{aligned}$$

which proves the upper bound. The lower bound is proven by triangle inequality and saturation property. □

Remark 3.2 With a slight abuse of terminology, we sometimes call both $\Delta_{N,M}$ and $\Delta_{N,M}^{\text{Hier}}$ "hierarchical error estimator." Strictly speaking, only $\Delta_{N,M}^{\text{Hier}}$ is an upper bound for the error, whereas $\Delta_{N,M}$ requires the multiplicative constant $(1 - \Theta_{N,M}^{\mathcal{N}})^{-1}$ in order to be an upper bound.

For the effectivity factor

$$\eta_{N,M}^{\mathcal{N}}(\mu) := \frac{\Delta_{N,M}^{\mathrm{Hier}}(\mu)}{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}} \tag{3.4}$$

we get that by using the estimate (3.3)

$$1 \leq \eta_{N,M}^{\mathcal{N}}(\mu) \leq \frac{1 + \Theta_{N,M}^{\mathcal{N}}}{1 - \Theta_{N,M}^{\mathcal{N}}}. \tag{3.5}$$

The closer $\Theta_{N,M}^{\mathcal{N}}$ is to zero, the better is the effectivity.

Remark 3.3 Using the same arguments as in the proof of Proposition 3.1, one can obtain the estimate

$$\|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}} \leq \frac{\Theta_{N,M}^{\mathcal{N}}}{1 - \Theta_{N,M}^{\mathcal{N}}} \cdot \Delta_{N,M}(\mu), \tag{3.6}$$

provided that (3.2) holds. However, the above proof does not yield a lower bound for $\|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}$ implying that the effectivity factor cannot be bounded and the upper bound (3.6) is in general not sharp. Note that this of course heavily depends on the particular choice of X_M.

3.2 Realization

The hierarchical error estimator can be computed online efficient as we are going to show now. In fact, let

$$u_N(\mu) = \sum_{i=1}^{N} \alpha_i^N(\mu)\, \xi_i, \qquad u_M(\mu) = \sum_{i=1}^{M} \alpha_i^M(\mu)\, \xi_i,$$

be the expansions of the reduced basis approximations (in general $\alpha_i^N(\mu) \neq \alpha_i^M(\mu)$ even for $1 \leq i \leq N$). Then, setting $\alpha_i^N(\mu) := 0$ for $i = N+1, \ldots, M$, we get

$$\begin{aligned}\Delta_{N,M}(\mu)^2 &= \left\| \sum_{i=1}^{M} (\alpha_i^N(\mu) - \alpha_i^M(\mu))\, \xi_i \right\|_{X^{\mathcal{N}}}^2 \\ &= \sum_{i,j=1}^{M} (\alpha_i^N(\mu) - \alpha_i^M(\mu))(\alpha_j^N(\mu) - \alpha_j^M(\mu))\, (\xi_i, \xi_j)_{X^{\mathcal{N}}}.\end{aligned}$$

Since the values $(\xi_i, \xi_j)_{X^{\mathcal{N}}}$ (the entries of the Gramian matrix) can be precomputed and stored in the offline stage, the computation of $\Delta_{N,M}(\mu)$ requires in general $\mathcal{O}(M^2)$ operations independent of $\mathcal{N}$, i.e., online efficient. If all ξ_i are orthonormalized, the computational complexity reduces to $\mathcal{O}(M)$. Of course, we face the well-known *square root effect*, since the above reasoning yields $\Delta_{N,M}(\mu)^2$ so that we loose half of the accuracy by taking the square root. This, however, is exactly the same for the standard estimator and there are suggestions how to deal with it (see, e.g., [8]).

3.3 Offline approximation of $\Theta_{N,M}^{\mathcal{N}}$

The main challenges for using the hierarchical error estimator are (i) the choice of an appropriate X_M and (ii) the determination of the multiplicative constant ρ with $e_N^{\mathcal{N}}(\mu) \leq \rho \Delta_{N,M}(\mu)$ for all $\mu \in \mathcal{P}$. Obviously, both issues are linked. In the case using the saturation assumption, we have that $\rho = (1 - \Theta_{N,M}^{\mathcal{N}})^{-1}$, so that we start describing an offline procedure to approximate the saturation constant.

To this end, we use a result on nonlinear parametrized programming problems.

Theorem 3.4 [14] *Let $\mathcal{P} \subset \mathbb{R}^P$ be compact and connected, $f, g : \mathcal{P} \to \mathbb{R}$ continuous such that $g(\mu) > 0$ for all $\mu \in \mathcal{P}$. Setting $F(q) := \max_{\mu \in \mathcal{P}} \{f(\mu) - q \cdot g(\mu)\}$, $q \in \mathbb{R}$, it holds $q_0 := \max_{\mu \in \mathcal{P}} \frac{f(\mu)}{g(\mu)}$ if and only if $F(q_0) = 0$.*

We apply this result for the functions $f(\mu) := \|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}$ and $g(\mu) := \|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}$. Due to the requirement $g(\mu) > 0$ for all $\mu \in \mathcal{P}$, we decompose the parameter space $\mathcal{P}$ in compact subsets $\mathcal{P}_i$ in such a way, that on each subset the denominator is non-vanishing. In view of (2.9), this means here that $\|R_N^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'} \neq 0$. Then, we proceed as follows: for fixed dimension N and for $\mathcal{P}_i$, we solve the nonlinear problem

$$\Theta_{N,M,i}^{\mathcal{N}} := \arg \min_{q \in \mathbb{R}_{\geq 0}} |F_i(q)| \quad \text{with} \quad F_i(q) := \max_{\mu \in \mathcal{P}_i} \{f(\mu) - q \cdot g(\mu)\} \tag{3.7}$$

and define $\Theta_{N,M}^{\mathcal{N}} := \max_i \Theta_{N,M,i}^{\mathcal{N}}$. For each i, we construct an iteration $\theta_i^{(k)}$, $k = 0, 1, 2, \ldots$, for which we need a good starting value $\theta_i^{(0)}$. Since

$$\frac{\beta^{\mathcal{N}}(\mu)}{\gamma^{\mathcal{N}}(\mu)} \cdot \frac{\|R_M^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}}{\|R_N^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}} \leq \frac{\|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}}{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}} \leq \frac{\gamma^{\mathcal{N}}(\mu)}{\beta^{\mathcal{N}}(\mu)} \cdot \frac{\|R_M^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}}{\|R_N^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}},$$

we use the following approximation as initial guess

$$\Theta_{N,M,i}^{\mathcal{N}} := \max_{\mu \in \mathcal{P}_i} \frac{\|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}}{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}} \approx \max_{\mu \in \mathcal{P}_i} \frac{\|R_M^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}}{\|R_N^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}} =: \theta_i^{(0)},$$

which is reasonable provided that $\min_{\mu \in \mathcal{P}_i} \frac{\beta_{\mathrm{LB}}^{\mathcal{N}}(\mu)}{\gamma_{\mathrm{UB}}^{\mathcal{N}}(\mu)} \approx \max_{\mu \in \mathcal{P}_i} \frac{\gamma_{\mathrm{LB}}^{\mathcal{N}}(\mu)}{\beta_{\mathrm{UB}}^{\mathcal{N}}(\mu)}$. This results in the (offline) Algorithm 1. If this algorithm terminates with some $\Theta_{N,M}^{\mathcal{N}} < 1$, the saturation property is in fact valid.

Remark 3.5 If Algorithm 1 terminates, the output yields a rigorous error bound for *all* parameters $\mu \in \mathcal{P}$. This is due to the fact that a global optimization strategy is used (in our experiments by SQP) to solve the nonlinear optimization problems in line 4 and 8, respectively. The price to pay is that (3.7) is an NP hard problem.

The possible decomposition of $\mathcal{P}$ into subsets $\mathcal{P}_i$ is used for stability purposes. In Section 3.4, this decomposition is not required due to the use of an Hermite approach,

Algorithm 1 Computing $\Theta^{\mathcal{N}}_{N,M}$.

Choose $\texttt{tol} > 0$, fix $N \in \mathbb{N}$, choose $L \in \mathbb{N}$ compact subsets $\mathcal{P}_i$, $1 \le i \le L$
for $i = 1 : L$ **do**
  $f(\mu) := \|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}$, $g(\mu) := \|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}$
  $F_i(q) := \max_{\mu \in \mathcal{P}_i} \{f(\mu) - q \cdot g(\mu)\}$
  $k := 0$
  $\theta_i^{(0)} := \max_{\mu \in \mathcal{P}_i} \frac{\|R_M^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}}{\|R_N^{\mathcal{N}}(\cdot;\mu)\|_{(Y^{\mathcal{N}})'}}$
  while $|F_i(\theta_i^{(k)})| \ge \texttt{tol}$ **do**
    iterate nonlinear problem $F_i(q) = 0 \rightsquigarrow \theta_i^{(k+1)}$
    $k \to k + 1$
  end while
  $\Theta^{\mathcal{N}}_{N,M,i} := \theta_i^{(k)}$
end for
return $\Theta^{\mathcal{N}}_{N,M} := \max_{i=1,\dots,L} \Theta^{\mathcal{N}}_{N,M,i}$

which works as long as the solution depends smoothly on the parameter. In fact, let $\mu^* \in \mathcal{P}$ be a parameter such that $\|u^{\mathcal{N}}(\mu^*) - u_N(\mu^*)\|_{X^{\mathcal{N}}} = 0$ and $\Delta\mu \in \mathcal{P}$. Using a Taylor approximation yields for sufficient regularity $u(\mu) \in C^m(\mathcal{P}; X)$

$$\frac{\|u^{\mathcal{N}}(\mu^* + \Delta\mu) - u_M(\mu^* + \Delta\mu)\|_{X^{\mathcal{N}}}}{\|u^{\mathcal{N}}(\mu^* + \Delta\mu) - u_N(\mu^* + \Delta\mu)\|_{X^{\mathcal{N}}}} = \frac{\mathcal{O}(\|\Delta\mu\|^m)}{\mathcal{O}(\|\Delta\mu\|)} \to 0 \qquad (\|\Delta\mu\| \to 0).$$

In those experiments, where we used a Lagrangian approach with $M > N$, we used two approaches indicated in Fig. 1: In the left part, we define "security zones" around snapshot parameters and use the complement in $\mathcal{P}$ of the union of all such security zones as the single parameter set $\mathcal{P}_1$. This choice is motivated by the fact that Matlab's global optimization routine is able to handle such sets. The right part of Fig. 1 shows the subdivision of $\mathcal{P}$ along the coordinate axes given by the previously selected snapshot parameters. In both cases, we assume that $\mathcal{P}$ has tensor product structure. Of course, other subdivision strategies are possible.

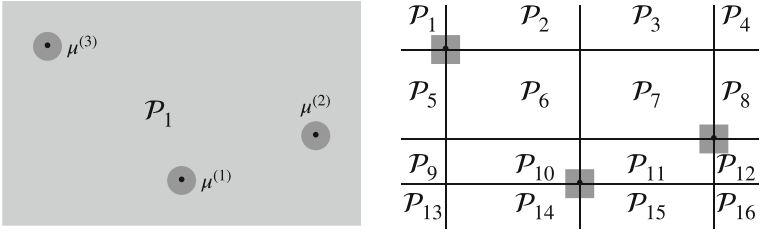


Fig. 1 Possible choices of subsets $\mathcal{P}_i$: Left: single $\mathcal{P}_1$ is the complement in $\mathcal{P}$ of the union of security zones (dark gray) around snapshot parameters (the light gray area); Right: grid parallel to the coordinate axes given by the previously selected snapshot parameters

3.4 Reduced basis generation

So far, we assumed that X_N and X_M are given, e.g., by a strong greedy method in an offline phase without using the hierarchical error estimator. One could also think of using the hierarchical part $\Delta_{N,M}(\mu)$ for this purpose. This, however, is at least not straightforward since one needs *both* N and M for the error estimator, where M has to be sufficiently large from the beginning. It would be a straightforward approach to start with $N = 1$, $M = 2$ for some parameters $\mu_1 \neq \mu_2$. Maximizing $\Delta_{1,2}(\mu)$ over a training set would yield μ_3 and we would set $N = 2$, $M = 3$, $S_3 = \{\mu_1, \mu_2, \mu_3\}$, etc. However, it can relatively easy be seen that this approach does not necessarily converge as snapshots may be selected repeatedly. Hence, we suggest a different approach.

Starting with X_N, the saturation property (3.2) is always valid as long as the Kolmogorov N-width decays and the reduced basis has been constructed with a weak greedy algorithm. However, this only means that for each RB space X_N, there exists an appropriate RB space X_M, s.t. (3.2) is satisfied—one is left with the question how to construct such a space X_M. We suggest to use the Taylor-RB method. If the solution $u(\mu)$ depends smoothly on the parameter μ, we can add derivatives of the snapshots w.r.t. the respective parameter to the basis, i.e., for $X_N = \operatorname{span}\{u(\mu_1), \dots, u(\mu_N)\}$ we set

$$X_M := \operatorname{span}\left\{ u(\mu_n), \frac{\partial^k}{\partial \mu_i^k} u(\mu_n) : k = 1, \dots, K_n,\ i = 1, \dots, P,\ n = 1, \dots, N \right\}$$

for appropriately chosen $K_n \in \mathbb{N}_0$. This means that $M = \sum_{n=1}^{N} (1 + K_n \cdot P)$. It is well-known that these *Taylor snapshots* $u_i^{(k)}(\mu) := \frac{\partial^k}{\partial \mu_i^k} u(\mu)$ can easily be computed recursively by solving the following linear variational problem (see, e.g., [35])

$$a(u_i^{(k)}(\mu), v; \mu) = \frac{\partial^k}{\partial \mu_i^k} f(v; \mu) - \sum_{m=1}^{k} \binom{k}{m} \frac{\partial^m}{\partial \mu_i^m} a(u_i^{(k-m)}(\mu), v; \mu). \tag{3.8}$$

In general, the partial derivatives appearing in (3.8) are Gâteaux derivatives. However, if the affine decomposition (2.6) holds, one just needs the derivatives of the involved functions $\theta_q^a, \theta_{q'}^f : \mathcal{P} \to \mathbb{R}$ in the classical sense. In this case, one can ensure by standard arguments that for each N there exists some $M > N$, s.t. the results of Section 3.1 hold, provided that the solution is real-analytic with respect to μ. Finally, for stability reasons, we orthonormalize the Taylor snapshots by a Gram-Schmidt procedure possibly neglecting (numerical almost) linear dependencies. The corresponding method is summarized in Algorithm 2.

Remark 3.6 It can be expected (and we have indeed confirmed this by several numerical experiments) that the saturation property (3.2) can be realized by decomposing the parameter space similar to [18] (there called "hp-RBM"). In addition to Algorithm 2, we have realized such an hp-RBM approach by modifying lines 5 to 10. We observed fast convergence.

Algorithm 2 (Weak) Greedy with hierarchical error estimator.

Choose tol > 0, $N_{\max}$, $\mathcal{P}_{\text{train}} \subset \mathcal{P}$, $\mu_1 \in \mathcal{P}$
$S_1 := \{\mu_1\}$, $\Xi_1^{(0)} := \{\xi_1 := u^{\mathcal{N}}(\mu_1)\}$
for $N = 1, \ldots, N_{\max}$ **do**
  $k = 1$
  repeat
    $\check{\Xi}_N^{(k)} := \{u_i^{(k)}(\mu_N) : i = 1, \ldots, P\}$ computed by (3.8)
    $\Xi_N^{(k)} :=$ ORTHONORMALIZE($\Xi_N^{(k-1)}$, $\check{\Xi}_N^{(k)}$)
    Set $X_N := \operatorname{span}(S_N)$, $X_M := \operatorname{span}(\Xi_N^{(k)})$ compute $\Theta_{N,M}^{\mathcal{N}}$ by Algorithm 1
    $k \leftarrow k + 1$
  until $\Theta_{N,M}^{\mathcal{N}} < 1$
  $K_N := k$
  if $\max_{\mu \in \mathcal{P}_{\text{train}}} \Delta_{N,M}(\mu) <$ tol **then**
    STOP
  else
    $\mu_{N+1} := \arg \max_{\mu \in \mathcal{P}_{\text{train}}} \Delta_{N,M}(\mu)$
    $S_{N+1} := S_N \cup \{\mu_{N+1}\}$, $\Xi_{N+1}^{(0)} := \Xi_N^{(0)} \cup \{\xi_{N+1} := u^{\mathcal{N}}(\mu_{N+1})\}$
  end if
end for
return S_N, $X_N := \operatorname{span}(S_N)$ and $X_M := \operatorname{span}(\Xi_N^{(K)})$, $\Theta_{N,M}^{\mathcal{N}}$

4 Numerical results

We investigate the quantitative performance of the RB hierarchical error estimator and focus on the sharpness and asymptotic correctness of (3.1). In particular, we want to investigate

1. How is the performance of $\Delta_{N,M}^{\text{Hier}}$ as compared to Δ_N^{Std}?
2. How does this comparison depend on the availability of a sharp lower inf-sup bound?
3. Since $\Delta_{N,M}$ is an upper bound for the error up to some multiplicative constant depending on M, what is a reasonable choice for that constant?
4. What is a good choice for X_M?

For that purpose, we report on experiments for two test problems. All experiments have been performed on iMac 2009 equipped with an Intel Core 2 Duo 3.06 GHz processor and 8 GB 1067 MHz DDR3 RAM.

The first example, the so-called thermal block from [33], is a well-known benchmark problem for the RBM. In this case, the behavior of the inf-sup/coercivity constant is known and the performance of the SCM is very good such that Δ_N^{Std} is expected to yield good results. We expect that $\Delta_{N,M}^{\text{Hier}}$ should be less sharp for general X_M and we are particularly interested in a quantitative comparison. The second example is the Helmholtz problem which has also been investigated in the RB context in [22]. In this case, it is known that the inf-sup constant has a poor behavior for

large parameters [19] and—moreover—the computation of a decent approximation using the SCM is quite costly. Hence, this should be a good benchmark test for the hierarchical error estimator.

For the basis generation, we use both the strong and the weak greedy algorithms based upon $\Delta_{N,M}$ and Δ_N^{Std} w.r.t. the same training set $\mathcal{P}_{\mathrm{train}}$. For $\Delta_{N,M}^{\mathrm{Hier}}$, we compare constructions of X_M using a Taylor and a Lagrange basis.

Remark 4.1 1. For simplicity, we compute $\Theta_{M,N}$ over a training set, i.e.,

$$\Theta_{N,M}^{\mathcal{N},\mathrm{train}} := \max_{\mu\in\mathcal{P}_{\mathrm{train}}} \frac{\|u^{\mathcal{N}}(\mu) - u_M(\mu)\|_{X^{\mathcal{N}}}}{\|u^{\mathcal{N}}(\mu) - u_N(\mu)\|_{X^{\mathcal{N}}}},$$

instead of solving the nonlinear problem (3.7).

2. Although all problems considered here are stationary, the hierarchical error estimator can also be applied to instationary problems e.g. by using a space-time formulation, [36, 37].

4.1 Thermal-block (see [33])

Let $\Omega := (0,1)^2$, divided into $B_1 \times B_2$ rectangular subblocks $\Omega_i \subset \Omega$, s.t. $\overline{\Omega} = \bigcup_{i=1}^{B_1B_2} \overline{\Omega_i}$. Let $\mu \in \mathcal{P} \subset \mathbb{R}^2$ and $\alpha(x;\mu) := \mu_j\, \chi_{\Omega_i}(x)$ for $j \in \{1,2\}$, $1 \le i \le B_1 \cdot B_2$, $\mu = (\mu_1,\mu_2) \in \mathcal{P}$, where $j = 1$ if and only if i is odd. We consider stationary heat conduction

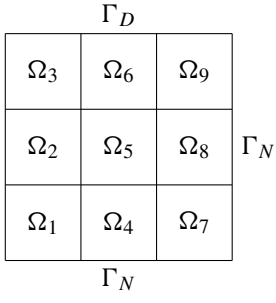
$$\begin{aligned} -\nabla\cdot(\alpha(x;\mu)\,\nabla u(x;\mu)) &= 0, && x\in\Omega,\\ u(x;\mu) &= 0, && x\in\Gamma_D := \{(x,1)^T \in \mathbb{R}^2 : 0\le x\le 1\},\\ \alpha(x;\mu)\frac{\partial u}{\partial n}(x) &= g_N(x;\mu), && x\in\Gamma_N := \partial\Omega\backslash\Gamma_D. \end{aligned}$$

Here, we choose $B_1 = B_2 = 3$ (see figure below) and set

$$g_N(x;\mu) := \begin{cases} 1, & \text{on } \{(x,0)^T \in \mathbb{R}^2 : 0\le x\le 1\},\\ 0, & \text{on } \{(0,y)^T \in \mathbb{R}^2 : 0\le y\le 1\}\cup\{(1,y)^T \in \mathbb{R}^2 : 0\le y\le 1\}. \end{cases}$$

This problem is coercive with $X = Y := H_D^1(\Omega) := \{v \in H^1(\Omega) : v|_{\Gamma_D} = 0\}$ and bilinear and linear forms defined as

$$\begin{aligned} a(u,v;\mu) &= \sum_{i=1}^{\lceil B_1B_2\rceil/2} \mu_1 \int_{\Omega_{2i-1}} \nabla u\cdot\nabla v\,dx\\ &\quad + \sum_{i=1}^{\lceil B_1B_2\rceil/2-1} \mu_2 \int_{\Omega_{2i}} \nabla u\cdot\nabla v\,dx,\\ f(v;\mu) &= \int_{\Gamma_N} v\,dx. \end{aligned}$$



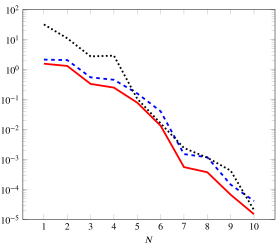
For the truth discretization, we used piecewise linear finite elements with a total number of 11.881 degrees of freedom. Further, we choose two different parameter spaces, namely

$$\mathcal{P}^{(1)} = [0.5, 1]^2, \qquad \mathcal{P}^{(2)} = [0.02, 1]^2, \qquad |\mathcal{P}^{(1)}_{\text{train}}| = |\mathcal{P}^{(2)}_{\text{train}}| = 10201 \approx 10^5.$$

For the error plots, the discrete coercivity constant (replacing the inf-sup constant) was determined as the smallest eigenvalue of a generalized eigenvalue problem. For the online CPU time for computing Δ_N^{Std}, we used the SCM.

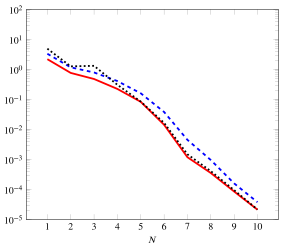
For the thermal block problem, the solution depends only mildly on the parameter since it is parametrically coercive. Hence, the SCM converges after only 3 steps to numerical precision, even on the larger parameter space $\mathcal{P}^{(2)}$. Therefore, we expect that Δ_N^{Std} is quite sharp, which is confirmed by our experiments. Starting with the smaller parameter set $\mathcal{P}^{(1)}$, we also found $\Delta_{N,M}^{\text{Hier}}$ to be quite sharp even for $M = N + 1$. We omit the corresponding figures since Δ_N^{Std} and $\Delta_{N,N+1}^{\text{Hier}}$ turned out to be almost indistinguishable. Hence, we consider the larger parameter set $\mathcal{P}^{(2)} \supset \mathcal{P}^{(1)}$. The results are displayed in Fig. 2 using the strong greedy and in Fig. 3 for the weak greedy with Δ_N^{Std} for the sampling. We do not see a significant difference between the different sampling methods to create the reduced basis spaces. In addition, we also did the parameter sampling by the hierarchical error estimator. We omit the corresponding figures since the results are quite similar to Figs. 2 and 3.

In both figures, we use 100 test parameters and plot the true (average) error in red solid lines. The dashed blue lines correspond to the average value of $\Delta_N^{\text{Std}}(\mu)$ for these 100 test parameters. Finally, the dotted black lines indicate the average values of $\Delta_{N,M}^{\text{Hier}}$ for $M \in \{N + 1, N + 2\}$ using a Taylor-based construction with $K_n = 1$ and $K_n = 2$, respectively. We see a significant improvement for $M = N + 2$ and almost no difference to Δ_N^{Std}. We comment on the max-error sequence instead of the average error in Fig. 10 below. In the tables next to the figures, we monitor the constants $\Theta_{N,M}$ for both choices. As expected, the value $\Theta_{N,M}$ significantly improves for $M = N + 2$. However, in all cases, the constant is below 1 and we can easily deduce online heuristics.



(a) Strong greedy, $M = N + 1$.

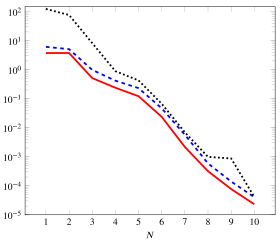
N	$\Theta_{N,N+1}$
1	0.9736
2	0.9156
3	0.9677
4	0.9141
5	0.2970
6	0.1716
7	0.8133
8	0.6927
9	0.8543
10	0.2969



(b) Strong greedy, $M = N + 2$.

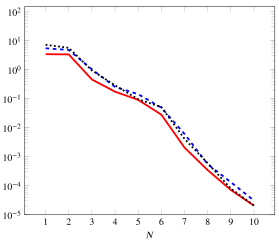
N	$\Theta_{N,N+2}$
1	0.5987
2	0.4890
3	0.6214
4	0.2715
5	0.0186
6	0.1382
7	0.1693
8	0.1413
9	0.1031
10	0.0133

Fig. 2 Thermal-Block, $\mathcal{P}^{(2)} = [0.02, 1]^2$, strong greedy sampling. Average error over test set of parameters. Red, solid: true error; blue, dashed: residual error estimator; black, dotted: hierarchical error estimator, $M \in \{N + 1, N + 2\}$. **a** Strong greedy $M = N + 1$. **b** Strong greedy $M = N + 2$



N	$\Theta_{N,N+1}$
1	0.9905
2	0.9532
3	0.9430
4	0.7848
5	0.7358
6	0.6466
7	0.7087
8	0.7163
9	0.9180
10	0.4665

(a) Weak greedy by $\Delta_N^{\mathrm{Std}}(\mu)$, $M = N + 1$.



N	$\Theta_{N,N+2}$
1	0.5350
2	0.4114
3	0.5183
4	0.4044
5	0.0469
6	0.4526
7	0.5077
8	0.4279
9	0.1000
10	0.0193

(b) Weak greedy by $\Delta_N^{\mathrm{Std}}(\mu)$, $M = N + 2$.

Fig. 3 Thermal-Block, $\mathcal{P}^{(2)} = [0.02, 1]^2$, weak greedy with standard error estimator. Average error over test set of parameters. Red, solid: true error; blue, dashed: residual error estimator; black, dotted: hierarchical error estimator, $M \in \{N + 1, N + 2\}$. **a** Weak greedy by $\Delta_N^{\mathrm{Std}}(\mu)$, $M = N + 1$. **b** Weak greedy by $\Delta_N^{\mathrm{Std}}(\mu)$ $M = N + 2$

Online effectivity As we have seen that both Δ_N^{Std} and $\Delta_{N,M}^{\mathrm{Hier}}$ (for appropriate values of M) are sharp, we investigate the online CPU time required to compute these error estimators. In order to do so, we consider the obtained effectivity η, i.e., the ratio of error estimator and true error for 100 test parameters. The results are shown in Fig. 4, where the values of η are plotted over the required online time. The circles correspond to $\Delta_{N,M}^{\mathrm{Hier}}$ for different values of M. The few circles with $\eta > 5$ correspond to quite small values of M and large parameter sets. All remaining values cluster for effectivities below 2 and online CPU times of less than 0.1 ms. As we can also see, the online CPU time is more or less independent of the choice of M. This is compared to Δ_N^{Std}. The online timings include also the SCM in this case. The crosses in Fig. 4 confirm the sharpness of the standard error estimator, but at the expense of CPU times which are about 15 times larger than for the hierarchical case.

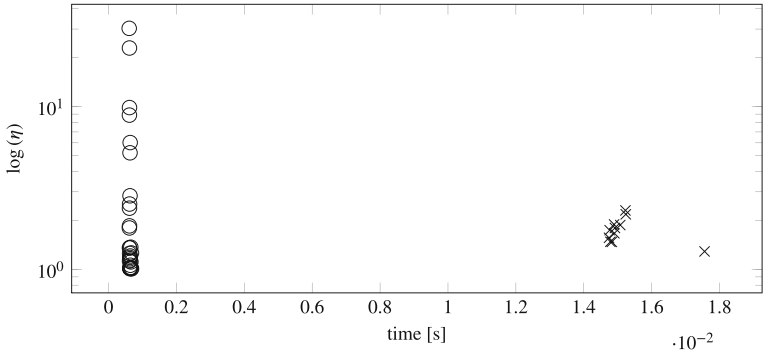


Fig. 4 Online effectivity index η over online CPU time for thermal block on $\mathcal{P}^{(2)}$, strong greedy. Circles: hierarchical error estimator for $M = N+1$, $M = N+2$ and $M = N+3$; crosses: standard error estimator (semilog scale)

4.2 Helmholtz problem

The Helmholtz equation arises from the time-dependent wave equation in the time-harmonic case (see [2, 19, 28, 29]) and references therein. Let $\Omega \subset \mathbb{R}^n$, $n \in \{1, 2, 3\}$, be a bounded Lipschitz domain with boundary $\Gamma := \partial\Omega$. For the parameter $\mu \in \mathcal{P} := [\mu_{\min}, \mu_{\max}] \subset \mathbb{R}$ with $1 \leq \mu_{\min} < \mu_{\max} < \infty$, the Helmholtz problem reads

$$\begin{aligned} -\Delta u(x) - \mu^2 u(x) &= r(x), && x \in \Omega, \\ u(x) &= 0, && x \in \Gamma_D \subset \Gamma, \\ \frac{\partial u}{\partial n}(x) + \mathrm{i}\mu u(x) &= g(x), && x \in \Gamma_R \subset \Gamma, \end{aligned} \tag{4.1}$$

where $\Gamma_D \cup \Gamma_R = \Gamma$. The parameter $\mu \in \mathcal{P}$ denotes the *wavenumber*, defined by $\mu := \frac{\omega}{c}$ (SI unit: m^{-1}), where $\omega \in \mathbb{R}$ denotes the frequency and $c \in \mathbb{R}$ the wave propagation speed, $\mathrm{i} := \sqrt{-1}$. In high-frequency problems, the wavenumber is quite large resulting in oscillations (see [19]0. We use $\mu_{\max} = 100$ here, since this suffices to show the desired effects. Test and trial spaces are again identical, $X = Y := H^1_D(\Omega; \mathbb{C}) := \{v \in H^1(\Omega; \mathbb{C}) : v|_{\Gamma_D} = 0\}$, but the sesquilinear form is no longer Hermitian, i.e.,

$$\begin{aligned} a(u, v; \mu) &= \int_\Omega \nabla u \cdot \overline{\nabla v}\, dx - \mu^2 \int_\Omega u\bar{v}\, dx + \mathrm{i}\mu \int_{\Gamma_R} u\bar{v}\, ds, \\ f(v; \mu) &= \int_\Omega r\bar{v}\, dx + \int_{\Gamma_R} g\bar{v}\, dx. \end{aligned}$$

The affine decomposition in the form (2.6) is clear. Such problems are usually analyzed using the parameter-dependent norm given by

$$\|v\|^2_{1,\mu} := \mu^2 \|v\|^2_0 + |v|^2_1, \qquad v \in H^1(\Omega; \mathbb{C}),$$

which is equivalent to $\|\cdot\|_1$, i.e., $\min\{1, \mu_{\min}\}\|v\|_1 \leq \|v\|_{1,\mu} \leq \max\{1, \mu_{\max}\}\|v\|_1$, $v \in H^1(\Omega; \mathbb{C})$, with coefficients, which depend on the parameter range, however.

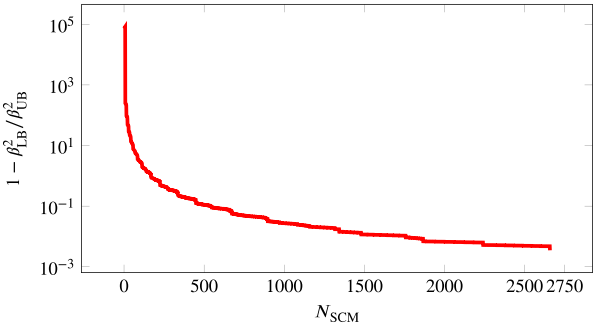
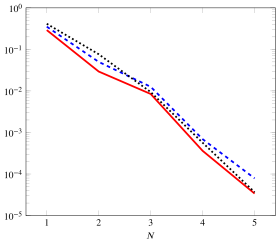
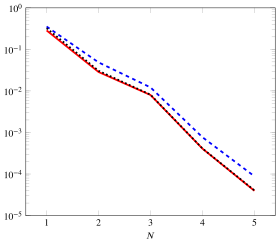


Fig. 5 SCM-convergence Helmholtz equation on $\mathcal{P}^{(2)}$



N	$\Theta_{N,N+1}$
1	0.3254
2	0.6396
3	0.1027
4	0.3714
5	0.0737

(a) Strong greedy, $M = N + 1$.



N	$\Theta_{N,N+2}$
1	0.1211
2	0.0657
3	0.0178
4	0.0103
5	0.0228

(b) Strong greedy, $M = N + 2$.

Fig. 6 Helmholtz equation, $\mathcal{P}^{(1)} = [1, 5]$, strong greedy. Average error over test set. Red, solid: true error; blue, dashed: residual error estimator; black, dotted: hierarchical error estimator for $M \in \{N + 1, M = N + 2\}$. **a** Strong greedy $M = N + 1$. **b** Strong greedy $M = N + 2$

The well-posedness is proven, e.g., in [19] by the Fredholm alternative. Moreover, there exists a constant $C_{\text{inf-sup}} > 0$ such that

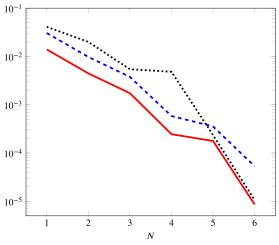
$$\inf_{w \in X} \sup_{v \in Y} \frac{|a(w, v; \mu)|}{\|w\|_{1,\mu} \|v\|_{1,\mu}} \geq \inf_{w \in X} \sup_{v \in Y} \frac{\text{Re}\{a(w, v; \mu)\}}{\|w\|_{1,\mu} \|v\|_{1,\mu}} \geq C_{\text{inf-sup}}\, \mu^{-\frac{7}{2}}. \tag{4.2}$$

For our numerical experiments, we consider three cases of parameter spaces, i.e.,

$$\mathcal{P}^{(1)} = [1, 5], \quad \mathcal{P}^{(2)} = [95, 100], \quad \mathcal{P}^{(3)} = [90, 100], \quad |\mathcal{P}^{(i)}_{\text{train}}| = 10^4 + 1, \; i = 1, 2, 3.$$

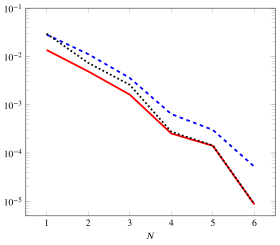
Thus, $\mathcal{P}^{(1)}$ is in the low-frequency domain so that the inf-sup constant is expected to be moderate, whereas $\mathcal{P}^{(2)}$, $\mathcal{P}^{(3)}$ will lead to oscillatory, high-frequency solutions. The latter choices allow to investigate the dependency on the *size* of the parameter set within the high-frequency regime. Our truth discretization is formed by spectral elements of degree 6 with 600 degrees of freedom for $\mathcal{P}^{(1)}$ (which turned out to be sufficient) and spectral elements of degree 16 with 16,000 degrees of freedom for $\mathcal{P}^{(2)}$ and $\mathcal{P}^{(3)}$.

In order to compare the results concerning the hierarchical estimator with the *best possible* standard one, we determined the involved discrete inf-sup constant $\beta^{\mathcal{N}}(\mu)$ by computing the smallest eigenvalue of a generalized eigenvalue problem. As this is not online efficient, we used the SCM for the online comparisons in terms of



N	$\Theta_{N,N+1}$
1	0.7075
2	0.8407
3	0.7173
4	0.9563
5	0.2672
6	0.2088

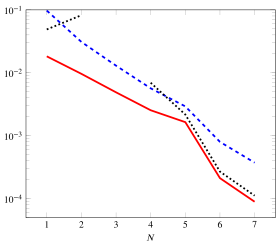
(a) Strong greedy, $M = N + 1$.



N	$\Theta_{N,N+2}$
1	0.5948
2	0.3390
3	0.3770
4	0.0610
5	0.0120
6	0.0093

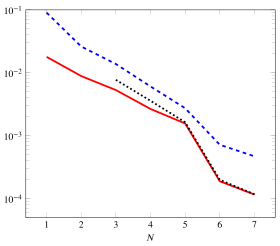
(b) Strong greedy, $M = N + 2$.

Fig. 7 Helmholtz equation, $\mathcal{P}^{(2)} = [95, 100]$, strong greedy. Average error over test set of parameters. Red, solid: true error; blue, dashed: residual error estimator; black, dotted: hierarchical error estimator with $M \in \{N + 1, N + 2\}$. **a** Strong greedy $M = N + 1$. **b** Strong greedy $M = N + 2$



N	$\Theta_{N,N+2}$
1	0.6353
2	0.8975
3	1.1299
4	0.6534
5	0.2710
6	0.1953
7	0.2259

(a) Strong greedy, $M = N + 2$.



N	$\Theta_{N,N+3}$
1	0.4499
2	1.1126
3	0.3145
4	0.2776
5	0.0422
6	0.0461
7	0.0084

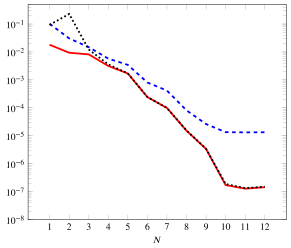
(b) Strong greedy, $M = N + 3$.

Fig. 8 Helmholtz equation, $\mathcal{P}^{(3)} = [90, 100]$, strong greedy. Average error over test set of parameters. Red, solid: true error; blue, dashed: residual error estimator; black, dotted: hierarchical error estimator, $M \in \{N + 2, N + 3\}$. **a** Strong greedy $M = N + 2$. **b** Strong greedy $M = N + 3$

CPU time. By (4.2), we expect fairly small inf-sup constants for large wavenumbers, which is expected to cause problems in Δ_N^{Std}. This fact is also mirrored by the poor convergence of the SCM shown in Fig. 5. For a good performance of Δ_N^{Std} in terms of sharpness, one needs a good online approximation of $\beta^{\mathcal{N}}(\mu)$ resulting in many SCM iterations and large CPU times.

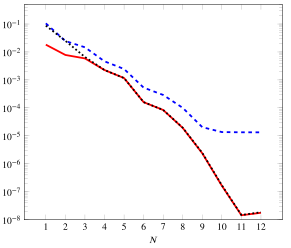
We start by describing the result for the low-frequency parameter set $\mathcal{P}^{(1)}$ and reduce ourselves to the strong greedy sampling since the results for the weak greedy with various error estimators turned out to be pretty much the same. As we can see in Fig. 6, both standard and hierarchical error estimators are quite sharp and the constants $\Theta_{N,M}$ are small—overall a similar behavior as for the thermal block.

Next, we consider the (smaller) high-frequency parameter set $\mathcal{P}^{(2)}$ and again restrict ourselves to the strong greedy sampling (the results for different versions of the weak are again quite similar). First, we note that the minimal choice of $M = N+1$ for the hierarchical error estimator is not sufficient in order to yield sharp estimates as can be seen in the left graph in Fig. 7. We have also found that the saturation property



N	$\Theta_{N,M}$
1	0.8435
2	0.9789
3	0.4281
4	0.1216
5	0.0003
6	0.0014
7	0.0017
8	0.0247
9	0.0131
10	0.0834
11	0.0391
12	0.0399

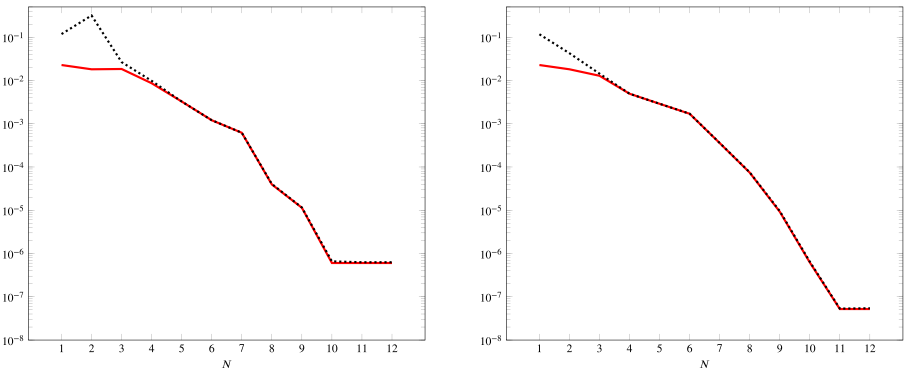
(a) Weak greedy by $\Delta_{N,M}^{\text{Hier}}(\mu)$, $K_n = 2$.



N	$\Theta_{N,M}$
1	0.81958
2	0.76875
3	0.10579
4	0.00004
5	0.00013
6	0.00292
7	0.00524
8	0.01601
9	0.03504
10	0.04012
11	0.02210
12	0.04177

(b) Weak greedy by $\Delta_{N,M}^{\text{Hier}}(\mu)$, $K_n = 3$.

Fig. 9 Helmholtz equation, $\mathcal{P}^{(3)} = [90, 100]$, weak greedy with parameter sampling via hierarchical error estimator. Average error over test set of parameters. Red, solid: true error; blue, dashed: residual error estimator; black, dotted: hierarchical error estimator, with Taylor basis, $K_n \in \{2, 3\}$ for all n. **a** Weak greedy by $\Delta_{N,M}^{\text{Hier}}(\mu)$, $K_n = 2$. **b** Weak greedy by $\Delta_{N,M}^{\text{Hier}}(\mu)$, $K_n = 3$



(a) Weak greedy by $\Delta^{\text{Hier}}_{N,M}(\mu)$, $K_n = 2$. (b) Weak greedy by $\Delta^{\text{Hier}}_{N,M}(\mu)$, $K_n = 3$.

Fig. 10 Helmholtz equation, $\mathcal{P}^{(3)} = [90, 100]$, weak greedy with parameter sampling via hierarchical error estimator. Worst-case error (max-error sequence). Red, solid: true error; black, dotted: hierarchical error estimator, with Taylor basis, $K_n \in \{2, 3\}$ for all n. **a** Weak greedy by $\Delta^{\text{Hier}}_{N,M}(\mu)$, $K_n = 2$. **b** Weak greedy by $\Delta^{\text{Hier}}_{N,M}(\mu)$, $K_n = 3$

for $M = N+2$ cannot be guaranteed numerically in this case (see Fig. 8). In the right graph, we thus use a Lagrange basis with $M = N + 2$ and obtain bounds that are even better than for the standard estimator. Recall that the blue dashed line for Δ^{Std}_N is w.r.t. to a high-fidelity approximation for the inf-sup constant, i.e., the best possible standard residual-based error bound. Also, the values for $\Theta_{N,M}$ are quite good. Thus, $\Delta^{\text{Hier}}_{N,N+2}$ is a cheap and sharp error bound even for the high-frequency case.

Finally, we consider $\mathcal{P}^{(3)}$, which is a high-frequency parameter set of doubled size as compared to $\mathcal{P}^{(2)}$. The error plots for the strong greedy sampling are shown in Fig. 8. In this case, the Lagrange-based space X_M for $M = N + 2$ only yields reasonable results for $N \geq 4$ (for smaller values, the saturation is not guaranteed), but then $\Delta^{\text{Hier}}_{N,N+2}$ outperforms Δ^{Std}_N in terms of accuracy. As we can see from the right-hand side of the figure, $M = N + 3$ gives quite sharp results for $N \geq 3$. Again, for smaller values of N, the saturation is not justified.

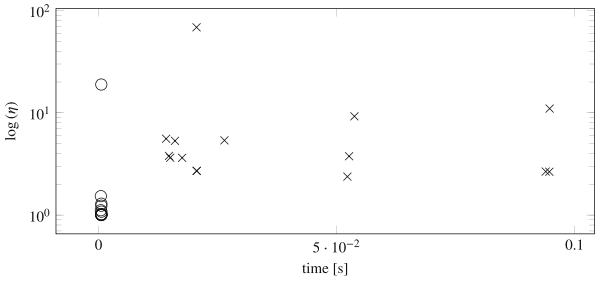


Fig. 11 Effectivity index η over online CPU time for Helmholtz problem on $\mathcal{P}^{(1)}$, $\mathcal{P}^{(2)}$; strong greedy sampling. Circles: hierarchical error estimator for different M; crosses: standard error estimator (semilog scale)

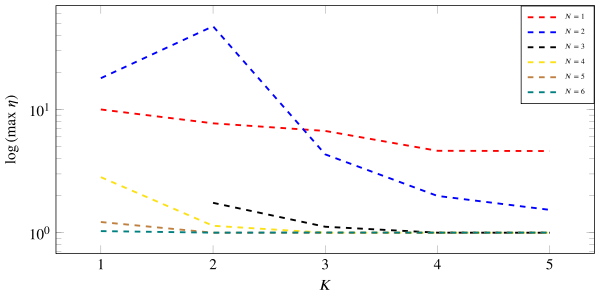


Fig. 12 Helmholtz equation, $\mathcal{P}^{(3)} = [90, 100]$. Effectivity index η over the number K of derivatives for the Taylor-RB basis (semilog scale)

Due to the lack of saturation for the Lagrange-type construction, we also tested the Taylor approach. We obtained even better results for all parameter sets. For $\mathcal{P}^{(3)}$, we display the results of a weak greedy sampling in Fig. 9. Even for $K_n = 2$ for all n, we got good results for $N \geq 5$, as can be seen by the fact that the values of $\Theta_{N,M}$ are close to zero. Moreover, $\Delta_{N,M}^{\text{Hier}}$ is quite sharp. The situation even improves for $K_n = 3$ in terms of sharpness for small N. For completeness, we show the max-error sequence in Fig. 10. Note that the slope is very close to the average error, which has been observed in all our experiments.

Online effectivity As before in Section 4.1 for the thermal block, we compare the online efficiencies of standard and hierarchical error estimators (see Fig. 11). First, note that we could not include values for the larger high-frequency parameter range $\mathcal{P}^{(3)}$ there, since the SCM required for Δ_N^{Std} did not converge, which means that the standard bound cannot be used in an online efficient manner.[2]

In Fig. 11, we show the effectivity over the online CPU time, again for Δ_N^{Std} by crosses and for $\Delta_{N,M}^{\text{Hier}}$ (for different values of M) by circles. First, we note that the values of M almost do not influence the CPU times, so that we can easily adjust the accuracy, as before. Moreover, the accuracies of both bounds are quite comparable, but the computation of $\Delta_{N,M}^{\text{Hier}}$ is much faster.

Finally, we investigate how many derivatives are required in a Taylor-RB basis in order to reach a desired effectivity (see Fig. 12). Referring to the notation in Section 3.4, we choose a fixed number $K \equiv K_n$ of additional derivatives per snapshot, so that the resulting dimension of X_M is $M = N(1 + K)$, recalling that here $P = 1$. We plot the maximal effectivity index over 1000 randomly chosen parameters for different N. For $N = 1$ and $K = 5$ ($M = 6$), we get $\eta \leq 5$; for $N = 2$, $K = 5$ ($M = 12$), we obtain $\eta < 2$; and for the larger values of N, we are even close to 1. Note, that for $N = 3$, $K = 1$, the saturation assumption is not valid. We conclude that the Taylor-RB-basis yields quite good effectivities even for moderate values of K.

[2] In addition, in our numerical experiments the SCM did not converge at all using a discontinuous Galerkin truth discretization.

4.3 Conclusions

Let us come back to the questions from the beginning of this section:

1. *How is the performance of* $\Delta_{N,M}^{\mathrm{Hier}}$ *as compared to* Δ_N^{Std}*?*
   Even for those cases that are in favor of Δ_N^{Std} (stable with precise knowledge of the inf-sup constant), $\Delta_{N,M}^{\mathrm{Hier}}$ turned out to yield a sharp error bound and to be online efficient. The potential becomes even more pronounced for problems with bad inf-sup behavior.
2. *How does this comparison depend on a sharp lower* inf-sup *bound?*
   The poorer the inf-sup estimate is, the more Δ_N^{Std} is outperformed by $\Delta_{N,M}^{\mathrm{Hier}}$ – in terms of sharpness and efficiency.
3. *What is a reasonable choice for the constant* $\Theta_{N,M}$*?*
   In all tested examples, we got very reasonable values for $\Theta_{N,M}$, provided that the saturation holds. However, even the determination via a test set requires the computation of possibly many truth solutions, the optimization problem (3.7) for the verification of the saturation and the computation of $\Theta_{M,N}^{\mathcal{N}}$ is quite costly, even though done offline. But our results show that it might be sufficient to do this on a fairly small test set since we got nice results in all case.
4. *What is a good choice for* X_M*?*
   In all investigated cases, M could be chosen quite moderate. This is due to the fact that our problems are of elliptic flavor even in the Helmholtz case. In [6, 20] for problems involving transport phenomena, X_M has to be chosen significantly larger. However, we have also seen that even for problems with very small inf-sup constant, X_M can be chosen reasonably small. Moreover, the online CPU times seem almost independent on the choice of X_M and are much smaller as for computing Δ_N^{Std} using the SCM (if the SCM converges at all).

   We compared also Lagrange- and Taylor-type approaches to construct X_M. Trying to use the Lagrange approach within parameter sampling using a weak greedy approach resulted in multiple selections of snapshots and non-guaranteed saturation. Both problems could be resolved using the Taylor approach, which, however, requires a certain regularity of u with respect to the parameter. In this case, for a fixed N, we are able to improve the effectivity by increasing the order of derivatives.

Funding information M.R. was supported by the European Union within the EU-MORNet project.

References

1. Ali, M., Steih, K., Urban, K.: Reduced basis methods with adaptive snapshot computations. Adv. Comput. Math. **43**(2), 257-294 (2016)
2. Babuška, I.M., Sauter, S.A.: Is the pollution effect of the FEM avoidable for the helmholtz equation considering high wave numbers? SIAM J. Numer. Anal. **34**(6), 2392–2423 (1997)

3. Bank, R.E., Smith, R.K.: A posteriori error estimates based on hierarchical bases. SIAM J. Numer Anal. **30**(4), 921–935 (1993)
4. 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. Math. **339**(9), 667–672 (2004)
5. Binev, P., Cohen, A., Dahmen, W., DeVore, R., Petrova, G., Wojtaszczyk, P.: Convergence rates for greedy algorithms in reduced basis methods. SIAM J. Math. Anal. **43**(3), 1457–1472 (2011)
6. Brunken, J., Smetana, K., Urban, K.: (Parametrized) first order transport equations: realization of optimally stable Petrov-Galerkin methods. SIAM J. Sci. Comput. **41**(1), A592–A621 ArXiv e-prints (2018)
7. Buffa, A., Maday, Y., Patera, A.T., Prud'homme, C., Turinici, G.: A priori convergence of the greedy algorithm for the parametrized reduced basis method. ESAIM Math. Model. Numer Anal. **46**(3), 595–603 (2012)
8. Buhr, A., Engwer, C., Ohlberger, M., Rave, S.: A numerically stable a posteriori error estimator for reduced basis approximations of elliptic equations. In: 11th World Congress on Computational Mechanics, WCCM 2014, 5th European Conference on Computational Mechanics, ECCM 2014 and 6th European Conference on Computational Fluid Dynamics, ECFD 2014, pp. 4094–4102 (2014)
9. Canuto, C., Tonn, T., Urban, K.: A posteriori error analysis of the reduced basis method for nonaffine parametrized nonlinear PDEs. SIAM J. Numer Anal. **47**(3), 2001–2022 (2009)
10. Chen, Y., Hesthaven, J.S., Maday, Y., Rodríguez, J.: A monotonic evaluation of lower bounds for inf-sup stability constants in the frame of reduced basis approximations. C.R. Acad. Sci. Math. **346**(23), 1295–1300 (2008)
11. Chen, Y., Hesthaven, J.S., Maday, Y., Rodríguez, J.: Improved successive constraint method based a posteriori error estimate for reduced basis approximation of 2D Maxwell's problem. ESAIM Math. Model. Numer Anal. **43**(6), 1099–1116 (2009)
12. Cho, J.R., Oden, J.T.: A priori modeling error estimates of hierarchical models for elasticity problems for plate- and shell-like structures. Math. Comput. Modelling **23**(10), 117–133 (1996)
13. Cohen, A., DeVore, R.: Kolmogorov widths under holomorphic mappings. IMA J. Numer. Anal. **36**(1), 1–12 (2016)
14. Dinkelbach, W.: On nonlinear fractional programming. Manag. Sci. **13**(7), 492–498 (1967)
15. Domínguez, C., Stephan, E.P., Maischak, M.: A FE-BE coupling for a fluid-structure interaction problem: hierarchical a posteriori error estimates. Numer. Methods Partial Differential Equations **28**(5), 1417–1439 (2012)
16. Drohmann, M., Carlberg, K.: The ROMES method for statistical modeling of reduced-order-model error. SIAM/ASA J. Uncertain. Quantif. **3**(1), 116–145 (2015)
17. Drohmann, M., Haasdonk, B., Ohlberger, M.: Reduced basis approximation for nonlinear parametrized evolution equations based on empirical operator interpolation. SIAM J. Sci. Comput. **34**(2), A937–A969 (2012)
18. Eftang, J.L., Patera, A.T., Rønquist, E.M.: An 'hp' certified reduced basis method for parametrized elliptic partial differential equations. SIAM J. Sci. Comput. **32**(6), 3170–3200 (2010)
19. Esterhazy, S., Melenk, J.M.: On Stability of discretizations of the Helmholtz equation, vol. 83. In: Lecture Notes in Computational Science and Engineering, pp. 285–324. Springer, Heidelberg (2012)
20. Feinauer, J., Hein, S., Rave, S., Schmidt, S., Westhoff, D., Zausch, J., Iliev, O., Latz, A., Ohlberger, M., Schmidt, V.: MULTIBAT: unified workflow for fast electrochemical 3D simulations of lithium-ion cells combining virtual stochastic microstructures, Electrochemical Degradation Models and Model Order Reduction. arXiv:1704.04139 (2017)
21. Glas, S., Patera, A., Urban, K.: Reduced basis methods for the wave equation. Unpublished manuscript (2018)
22. Haasdonk, B.: Reduced basis methods for parametrized PDEs — a tutorial. In: Benner, P., Cohen, A., Ohlberger, M., Willcox, K. (eds.) Model reduction and approximation, chapter 2, pp. 65–136. SIAM, Philadelphia (2017)
23. 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)
24. Hesthaven, J.S., Rozza, G., Stamm, B.: Certified reduced basis methods for parametrized partial differential equations. Springer, Cham (2016)
25. Hesthaven, J.S., Stamm, B., Zhang, S.: Certified reduced basis method for the electric field integral equation. SIAM J. Sci. Comput. **34**(3), A1777–A1799 (2012)

26. Huang, Y., Wei, H., Yang, W., Yi, N.: A new a posteriori error estimate for adaptive finite element methods, pp. 63–74. Springer, Berlin (2011)
27. Huynh, D.B.P., Rozza, G., Sen, S., Patera, A.T.: A successive constraint linear optimization method for lower bounds of parametric coercivity and inf-sup stability constants. C.R. Acad. Sci. Math. **345**(8), 473–478 (2007)
28. Ihlenburg, F., Babuška, I.: Finite element solution of the Helmholtz equation with high wave number part: I the h-version of the FEM. Comp. Math. Appl. **30**(9), 9–37 (1995)
29. Ihlenburg, F., Babuška, I.: Finite element solution of the Helmholtz equation with high wave number part: II the h-p version of the FEM. SIAM J. Numer. Anal. **34**(1), 315–358 (1997)
30. Ohlberger, M., Rave, S.: Reduced basis methods: success, limitations and future challenges. In: Proceedings of the Conference Algoritmy, pp. 1–12 (2016)
31. Ohlberger, M., Rave, S., Schindler, F.: True error control for the localized reduced basis method for parabolic problems. In: Model Reduction of Parametrized Systems, pp. 169-182. Springer, Cham (2017)
32. Ohlberger, M., Schindler, F.: Error control for the localized reduced basis multiscale method with adaptive on-line enrichment. SIAM J. Sci. Comput. **37**(6), A2865–A2895 (2015)
33. Patera, A., Rozza, G.: Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations. MIT, Cambridge (2006). Version 1.0
34. Prince, P.J., Dormand, J.R.: High order embedded Runge-Kutta formulae. J. Comput. Appl. Math. **7**(1), 67–75 (1981)
35. Quarteroni, A., Manzoni, A., Negri, F.: Reduced basis methods for partial differential equations: an introduction. Springer, Cham (2016)
36. Schwab, C., Stevenson, R.: Space-time adaptive wavelet methods for parabolic evolution problems. Math. Comput. **78**(267), 1293–1318 (2009)
37. Urban, K., Patera, A.T.: An improved error bound for reduced basis approximation of linear parabolic problems. Math. Comput. **83**(288), 1599–1615 (2014)
38. Urban, K., Volkwein, S., Zeeb, O.: Greedy sampling using nonlinear optimization. In: Reduced Order Methods For Modeling And Computational Reduction, pp. 137-157. Springer, Cham (2014)
39. Wohlmuth, B.I.: Hierarchical a posteriori error estimators for mortar finite element methods with lagrange multipliers. SIAM J. Numer. Anal. **36**(5), 1636–1658 (1999)
40. Yano, M.: A reduced basis method with exact-solution certificates for steady symmetric coercive equations. Comput. Methods Appl. Mech. Eng. **287**, 290–309 (2015)
41. Yano, M.: A minimum-residual mixed reduced basis method: exact residual certification and simultaneous finite-element reduced-basis refinement. ESAIM Math. Model. Numer. Anal. **50**(1), 163–185 (2016)
42. Zienkiewicz, O.C., Kelly, D.W., Gago, J., Babuška, I.: Hierarchical finite element approaches, error estimates and adaptive refinement. In: The Mathematics of Finite Elements and Applications, IV (Uxbridge, 1981), pp. 313-346. Academic Press, London (1982)
43. Zou, Q., Veeser, A., Kornhuber, R., Gräser C.: Hierarchical error estimates for the energy functional in obstacle problems. Numer Math. **117**(4), 653–677 (2011)

Affiliations

Stefan Hain[1] · Mario Ohlberger[2] · Mladjan Radic[1] · Karsten Urban[1]

Stefan Hain
stefan.hain@uni-ulm.de

Mario Ohlberger
mario.ohlberger@uni-muenster.de

Mladjan Radic
mladjan.radic@uni-ulm.de

[1] Institute for Numerical Mathematics, Ulm University, Helmholtzstr. 20, 89081 Ulm, Germany

[2] Applied Mathematics, University of Münster, Einsteinstr. 62, 48149 Münster, Germany