Adv Comput Math (2020) 46: 32
https://doi.org/10.1007/s10444-020-09741-x

Rigorous and effective a-posteriori error bounds for nonlinear problems—application to RB methods

Andreas Schmidt[1] · Dominik Wittwar[1] · Bernard Haasdonk[1]

Published online: 23 March 2020

Abstract

Quantifying the error that is induced by numerical approximation techniques is an important task in many fields of applied mathematics. Two characteristic properties of error bounds that are desirable are reliability and efficiency. In this article, we present an error estimation procedure for general nonlinear problems and, in particular, for parameter-dependent problems. With the presented auxiliary linear problem (ALP)-based error bounds and corresponding theoretical results, we can prove large improvements in the accuracy of the error predictions compared with existing error bounds. The application of the procedure in parametric model order reduction setting provides a particularly interesting setup, which is why we focus on the application in the reduced basis framework. Several numerical examples illustrate the performance and accuracy of the proposed method.

Keywords Error estimation · Model order reduction · Reduced basis method

Mathematics subject classification (2010) 65J15 · 65L70

Communicated by: Anthony Nouy

This article belongs to the Topical Collection: *Model Reduction of Parametrized Systems*
Guest Editors: Anthony Nouy, Peter Benner, Mario Ohlberger, Gianluigi Rozza, Karsten Urban and Karen Willcox

✉ Dominik Wittwar
dominik.wittwar@mathematik.uni-stuttgart.de

Andreas Schmidt
andreas.schmidt@mathematik.uni-stuttgart.de

Bernard Haasdonk
haasdonk@mathematik.uni-stuttgart.de

[1] Universitat Stuttgart, Pfaffenwaldring 57, 70569, Stuttgart, Germany

1 Introduction

A-posteriori error estimates are important tools in many disciplines of applied mathematics. For example, they are required for assessing the quality of numerical approximations and to guarantee their feasibility in the corresponding scenario. Popular examples for the application of error bounds are adaptive refinement strategies, where error estimates are used to judge whether the spatial or temporal discretization should be refined to improve the quality of the approximation, see for example [1, 9, 17]. Two desirable properties of such bounds are rigorosity, i.e. the error bound should be a valid upper bound, and effectivity, i.e. the factor of overestimation should be computable. In the context of the finite element method (FEM), those properties are also referred to as reliability and efficiency.

An area where error bounds are of utmost importance is reduced order modeling. Faced with the computational complexity involved with solving high-dimensional systems of equations arising for example for highly accurate discretizations of partial differential equations (PDEs), several techniques have emerged that try to tackle this challenge by reducing the dimension of the problem. This is typically done by a projection of the problem onto a low-dimensional subspace that contains enough information about the solution to the problem. The projection then yields a problem of low-dimension which can be solved with low computational complexity and which yields an approximation to the high-dimensional solution. The important question that should then be answered is how far the approximation is from the true solution. To this end, one typically employs a-posteriori error estimates which in the ideal case deliver a rigorous upper bound that does not deviate too much from the true error. Giving a complete overview over the available methods and corresponding results for the error estimation is out of scope of this paper. Instead, we refer to [3] and [4] for recent overviews of model (order) reduction in the parameteric and nonparametric cases. Based on these techniques, approximate solutions can be calculated cheaply and in a computationally efficient manner. One framework that is particularly suitable for parametric problems is the reduced basis (RB) method. The essential idea of RB methods is to identify low-dimensional subspaces in the high-dimensional solution spaces by exploring the parameter domain with so-called greedy algorithms. In this article, we will demonstrate that classical error bounds which are well-established within RB methods can be significantly improved by introducing an auxiliary linear problem and corresponding RB approximation. By the proposed procedure, we are able to reach optimal effectivities of almost 1 in many examples. Additionally, the necessity for good lower bounds on the inf–sup constant can be alleviated, which allows the use of rougher (and thus computationally less expensive) lower bounds. Furthermore, the quality of the error bound can be tuned according to the application requirements.

In this paper, we improve a residual-based error estimation technique that has been used frequently during the last decades. We illustrate the essential idea of the improvement that enables highly accurate error estimates by the following simple example: Consider the vector-valued equation $Ax = f$ for $A = \left(\begin{smallmatrix} 1 & 0 \\ 10 & 1 \end{smallmatrix}\right)$ and $f = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$. This equation has the unique solution $x^* = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$. Assume we have a numerical scheme that is able to produce the approximate solution $\hat{x} \in \mathbb{R}^2$ with say $\hat{x} = 1.01x^*$. This results in a very low error (in the Euclidean norm) of only $\|\hat{x} - x^*\|_2 = 0.01$. Usually the true solution x^* is not available for the evaluation of the error, which is why we are interested in finding rigorous upper bounds to the norm of the true error $e := \hat{x} - x^*$. The straightforward procedure for doing this is to define the residual $r := A\hat{x} - f$ and to derive the equation $Ae = r$ for the error e. It then directly follows $\|e\|_2 \leq \|A^{-1}\|_2\|r\|_2 \approx 10.1 \cdot 0.01 \approx 0.101$, which is an overestimation of factor ≈ 10, which is already quite large in this small example. To obtain more accurate error bounds in this linear setting, we start again with the equation for the error $Ae = r$. It is an interesting observation that by solving this equation exactly, the error can be calculated exactly, i.e. with no overestimation. Unfortunately this is often too expensive in applications since this essentially adds the complexity of solving the original problem again (at least in the linear case). The central idea now is to not solve the error equation $Ae = r$ exactly but by another computationally efficient method that produces approximate solutions. If we assume to have a numerical scheme that is able to calculate an approximate error $\widehat{e}$ rapidly, we can make use of the triangle inequality and deduce the upper bound $\|e\|_2 \leq \|\widehat{e}\|_2 + \|\widehat{e} - e\|_2$. The second term can be estimated similarly to the first error bound by introducing a second residual $R := A\widehat{e} - r$, from which we then obtain the final bound

$$\|e\|_2 \leq \|\widehat{e}\|_2 + \|A^{-1}\|_2\|R\|_2. \tag{1}$$

Returning to the toy example and assuming an approximation $\widehat{e} = 1.01e$, we can evaluate equation (1) and obtain $\|e\|_2 \leq 0.01111$, which gives an overestimation factor of approximately 1.111. Hence, the error estimate is improved by a factor of about 10.

In this paper, we show how the idea behind this very simple example can be generalized to a large class of linear and nonlinear problems, especially in the context of RB methods. To this end, we introduce a generic nonlinear error estimate in Section 2. We discuss the application in the RB context in Section 3 and provide several numerical examples in Section 4. Finally, we present a conclusion and an outlook in Section 5.

2 Rigorous and effective error bounds

We first clarify the setting used throughout this article. In what follows, we always assume X and Y to be Banach spaces with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. The set of all bounded linear operators from X to Y is denoted as $\mathscr{L}(X, Y)$. For $A \in \mathscr{L}(X, Y)$, we define the operator norm $\|A\|_{\mathscr{L}(X,Y)} := \sup_{0 \neq x \in X} \frac{\|Ax\|_Y}{\|x\|_X}$.

Throughout this article, we consider continuously differentiable mappings $G \in C^1(X, Y)$ and are interested in solving the problem

$$\text{Find } x \in X \text{ such that } G(x) = 0. \tag{P}$$

An element $x^* \in X$ is called (true) solution to the problem (P), if $G(x^*) = 0$.

In the remainder of this article, we always assume that at least one solution exists. We are interested in estimating the error $e := \hat{x} - x^*$ between a true solution and a suitable approximation $\hat{x} \in X$ by means of reliable a-posteriori error bounds, which can be represented by functions $\Delta : X \to \mathbb{R}$ with the property

$$\|\hat{x} - x^*\|_X \leq \Delta(\hat{x}). \tag{2}$$

A general framework for providing such error estimates can be found in [7]. However, the results presented in that reference often lead to quite large overestimations of the true error. The quality of the upper bound Δ can be quantified in terms of the so-called effectivity, which is defined as

$$\mathrm{eff}(\hat{x}) := \frac{\Delta(\hat{x})}{\|x^* - \hat{x}\|_X}. \tag{3}$$

By its definition, it is clear that for reliable (i.e. rigorous) error estimates, it always holds $\mathrm{eff}(\hat{x}) \geq 1$. Ideally we aim for error bounds that provide effectivities close to one as we then get almost exact error predictions.

2.1 Rigorous, effective, and computable a-posteriori error estimates with effectivity bounds

In this section, we refine the results derived in [7] and show how significant improvements can be achieved. We want to emphasize that these derivations are independent of model order reduction but apply to any kind of approximation procedure. To this end, let us assume that the Fréchet-derivative $\mathrm{DG}|_{\hat{x}}$ of G at the approximate solution $\hat{x}$ defines an invertible linear operator from X to Y. Based on this derivative, we then define the following three quantities, where $\overline{B_\alpha}(\hat{x}) = \{x \in X | \; \|x - \hat{x}\|_X \leq \alpha\}$ denotes the closed ball in X with radius α around $\hat{x}$

$$\begin{aligned} \epsilon(\hat{x}) &:= \| \,\mathrm{DG}|_{\hat{x}}^{-1} \left(G(\hat{x})\right) \|_X, && \text{(nonsplit residual)} \\ \gamma(\hat{x}) &:= \| \,\mathrm{DG}|_{\hat{x}}^{-1} \|_{\mathscr{L}(Y,X)}, && \text{(stability constant)} \\ L(\alpha) &:= \sup_{x \in \overline{B_\alpha}(\hat{x})} \| \,\mathrm{DG}|_x - \mathrm{DG}|_{\hat{x}} \|_{\mathscr{L}(X,Y)}, && \text{(local nonlinearity indicator).} \end{aligned}$$

Due to the assumption on G, $\mathrm{DG}|_{\hat{x}}$ is bounded and thus $\mathrm{DG}|_{\hat{x}}^{-1}$ is also bounded according to the bounded inverse theorem and the above quantities are well-defined. Based on these quantities, we are able to prove the following fundamental error estimate.

Theorem 1 (**Rigorous a-posteriori error estimation**) *Let $\hat{x} \in X$ be an approximate solution and assume that $\mathrm{DG}|_{\hat{x}} : X \to Y$ is invertible. Let the validity criterion*

$$\tau(\hat{x}) := 2\gamma(\hat{x})L(2\epsilon(\hat{x})) \leq 1$$

holds. Then the problem $G(x) = 0$ has a unique solution $x^* \in X$ in the closed ball $\overline{B_{2\epsilon(\hat{x})}}(\hat{x})$ and the following upper bound for the error $e = \hat{x} - x^* \in X$ holds

$$\|e\|_X = \|\hat{x} - x^*\|_X \leq \Delta(\hat{x}) := \frac{1}{1 - \tau(\hat{x})/2}\epsilon(\hat{x}) \leq 2\epsilon(\hat{x}). \tag{4}$$

Proof We present the full proof of the theorem in Appendix A. □

Remark 1 Note that similar bounds have been derived by various authors [20, 21, 23]. However, in the bounds in literature known to us, the nonsplit residual $\epsilon(\hat{x})$ is replaced by the upper bound

$$\epsilon_{\text{split}}(\hat{x}) := \gamma(\hat{x})\|G(\hat{x})\|_Y \geq \| \operatorname{DG}|_{\hat{x}}^{-1} (G(\hat{x}))\|_X = \epsilon(\hat{x}), \tag{5}$$

which we call split residual for obvious reasons. As we have seen in the introduction and as we will see in the numerical results, this splitting can induce a very large overestimation. This is not the case when the quantity $\epsilon(\hat{x})$ or other, more accurate approximations to it, are used. Hence, the results in this article improve all the aforementioned existing results.

The quantity $L(\alpha)$ can be seen as a measure for the nonlinearity of the problem in the vicinity of the approximate solution. In particular, for (affine) linear problems, we immediately get $L(\alpha) = 0$ (and $\tau(\hat{x}) = 0$, i.e. unconditional validity) and hence even exact error predictions as stated in the following corollary.

Corollary 1 (**Exact error prediction for linear problems**) *Let G be affine linear in x. Then it holds*

$$\|e\|_X = \|\hat{x} - x^*\|_X = \Delta(\hat{x}), \quad \textit{and} \quad \operatorname{eff}(\hat{x}) = 1.$$

Proof Since G is affine linear in x it can be written as $G(x) = Ax + g$ for some $A \in \mathscr{L}(X, Y)$ and $g \in Y$. We then obtain $G(\hat{x}) = G(\hat{x}) - G(x^*) = A(\hat{x} - x^*) = Ae$ or equivalently $e = A^{-1}(G(\hat{x}))$. We further infer

$$\|e\|_X = \|A^{-1}(G(\hat{x}))\|_X = \| \operatorname{DG}|_{\hat{x}}^{-1} (G(\hat{x}))\|_X = \epsilon(\hat{x}) = \Delta(\hat{x}),$$

since $\operatorname{DG}|_x = A$ for all $x \in X$ and $\tau(\hat{x}) = 0$. □

If the local nonlinearity indicator $L(\alpha)$ does not vanish but satisfies $L(\alpha) \leq C\alpha$ for some constant $C > 0$, the constant in front of the nonsplit residual in (4) can be improved as follows:

Lemma 1 *Let the assumptions of Theorem 1 hold and let $L(\alpha) \leq C\alpha$ for some $C > 0$. If the modified validity criterion*

$$\hat{\tau} := 4\gamma(\hat{x})C\epsilon(\hat{x}) \leq 1$$

is satisfied, then the problem $G(x) = 0$ has a unique solution $x^* \in X$ and the error $e = \hat{x} - x^*$ is bounded by

$$\|e\|_X \leq \frac{1 - \sqrt{1 - \hat{\tau}}}{2\gamma(\hat{x})C} = \frac{2}{1 + \sqrt{1 - \hat{\tau}}}\epsilon(\hat{x}). \tag{6}$$

Proof We present the full proof of the lemma in Appendix B. □

In (6), we recover the multiplicative structure of error bounds for RB methods for quadratic nonlinearities, e.g. [23] except for different leading factors. In particular, we omit the stability factor which arises when the split ϵ_{split} is used as a bound to the nonsplit residual ϵ.

As it was motivated in the introduction, the key quantity to assess the quality of the error bound is the effectivity $\text{eff}(\hat{x})$. In order to make quantitative statements of the effectivity for the error bound in the general nonlinear case, we assume that the function $\text{DG}|_{\hat{x}}^{-1} G(\cdot)$ is locally Lipschitz-continuous around $\hat{x}$. By this, we mean that there exists a constant $C_G(\hat{x}) \geq 0$ such that it holds

$$\| \text{DG}|_{\hat{x}}^{-1} (G(x)) - \text{DG}|_{\hat{x}}^{-1} (G(\hat{x}))\|_X \leq C_G(\hat{x})\|x - \hat{x}\|_X, \quad \forall x \in \overline{B_{2\epsilon(\hat{x})}}(\hat{x}). \quad (7)$$

Based on this property, we are able to prove an estimate for the effectivity for locally Lipschitz-continuous problems.

Lemma 2 (**Effectivity estimate**) *Let* $\text{DG}|_{\hat{x}}^{-1} (G(\cdot))$ *be locally Lipschitz-continuous around* $\hat{x}$ *with constant* $C_G(\hat{x})$ *and let the error estimate from Theorem 1 holds true. Then it holds*

$$\text{eff}(\hat{x}) \leq \frac{C_G(\hat{x})}{1 - \tau(\hat{x})/2}.$$

Proof The proof follows directly from the fact that $G(x^*) = 0$ and

$$\begin{aligned}\Delta(\hat{x}) &= \frac{1}{1 - \tau(\hat{x})/2} \| \text{DG}|_{\hat{x}}^{-1} (G(\hat{x}))\|_X \\ &= \frac{1}{1 - \tau(\hat{x})/2} \| \text{DG}|_{\hat{x}}^{-1} (G(\hat{x}) - G(x^*))\|_X \\ &\leq \frac{C_G(\hat{x})}{1 - \tau(\hat{x})/2} \|\hat{x} - x^*\|_X.\end{aligned}$$

□

Note that the effectivity estimate agrees with the result stated in Corollary 1. Indeed, for linear problems, we immediately observe $\tau(\hat{x}) = 0$ and $C_G(\hat{x}) = 1$ which results in $\text{eff}(\hat{x}) = 1$.

It is noteworthy that the local Lipschitz-continuity assumption is satisfied for a large class of problems. One class that is of particular interest in applications are quadratic problems such as the Navier-Stokes equation, Burgers equation, the algebraic Riccati equation (ARE), or nonlinear reaction-diffusion equations. The following proposition provides a bound on $C_G(\hat{x})$ in this case.

Proposition 1 (**Local Lipschitz-continuity for quadratic problems**) *Let G be quadratic, i.e. there exists a* $y_0 \in Y$, $A \in \mathscr{L}(X, Y)$ *and a continuous bilinear mapping* $B : X \times X \to Y$ *such that*

$$G(x) = y_0 + Ax + \frac{1}{2}B(x, x).$$

Then $\text{DG}|_{\hat{x}}^{-1} (G(\cdot)) : \overline{B_{2\epsilon(\hat{x})}}(\hat{x}) :\to X$ *is locally Lipschitz-continuous around* $\hat{x}$ *with*

$$C_G(\hat{x}) \leq 1 + \gamma(\hat{x})c_B\epsilon(\hat{x}),$$

where $c_B := \sup_{x,x' \in X\setminus\{0\}} \frac{\|B(x,x')\|_Y}{\|x\|_X\|x'\|_X}$ *is the continuity constant of B.*

Proof It holds

$$\mathrm{DG}|_{\hat{x}}(x) = Ax + \frac{1}{2}\left(B(x,\hat{x}) + B(\hat{x},x)\right) \quad \text{and} \quad \mathrm{D}^2G\Big|_{\hat{x}}(x,x') = \frac{1}{2}(B(x,x') + B(x',x)).$$

By direct computation, we get the Taylor-like expansion

$$G(x) = G(\hat{x}) + \mathrm{DG}|_{\hat{x}}(x-\hat{x}) + \frac{1}{2}\,\mathrm{D}^2G\Big|_{\hat{x}}(x-\hat{x},x-\hat{x})$$

and therefore,

$$\begin{aligned}
\mathrm{DG}|_{\hat{x}}^{-1}(G(x)) - \mathrm{DG}|_{\hat{x}}^{-1}(G(\hat{x})) &= \mathrm{DG}|_{\hat{x}}^{-1}(G(x) - G(\hat{x}))\\
&= \mathrm{DG}|_{\hat{x}}^{-1}\left(\mathrm{DG}|_{\hat{x}}(x-\hat{x}) + \frac{1}{2}\,\mathrm{D}^2G\Big|_{\hat{x}}(x-\hat{x},x-\hat{x})\right)\\
&= x - \hat{x} + \frac{1}{2}\,\mathrm{DG}|_{\hat{x}}^{-1}(\mathrm{D}^2G\Big|_{\hat{x}}(x-\hat{x},x-\hat{x}))\\
&= x - \hat{x} + \frac{1}{2}\,\mathrm{DG}|_{\hat{x}}^{-1}(B(x-\hat{x},x-\hat{x}).
\end{aligned}$$

Taking the norm on both sides, applying the definition of the continuity constant c_B and using the triangle inequality, we get

$$\|\,\mathrm{DG}|_{\hat{x}}^{-1}(G(x)) - \mathrm{DG}|_{\hat{x}}^{-1}(G(\hat{x}))\|_Y \leq \|x-\hat{x}\|_X + \tfrac{1}{2}\gamma(\hat{x})c_B\|x-\hat{x}\|_X^2.$$

Finally, applying the bound given in (4) gives the desired result. □

In the infinite-dimensional settings, the calculation of the involved quantities is often not possible while in the finite-dimensional case, it can be computationally demanding or even infeasible. This is particularly true for very high-dimensional settings arising for example from semi-discretized PDEs. Instead, one often only has computable upper bounds to the quantities, i.e.

$$\epsilon(\hat{x}) \leq \epsilon_{\mathrm{ub}}(\hat{x}), \quad \gamma(\hat{x}) \leq \gamma_{\mathrm{ub}}(\hat{x}), \quad L(\alpha) \leq L_{\mathrm{ub}}(\alpha). \tag{8}$$

In this case, Theorem 1 remains valid with the replaced quantities:

Theorem 2 (**Computable error bound and effectivity estimate**) *Let $\hat{x} \in X$ be an approximate solution and assume that $\mathrm{DG}|_{\hat{x}}$ is invertible. Let the validity criterion*

$$\tau_{ub}(\hat{x}) := 2\gamma_{ub}(\hat{x})L_{ub}(2\epsilon_{ub}(\hat{x})) \leq 1. \tag{9}$$

holds. Then the problem $G(x) = 0$ has a unique solution $x^* \in X$ in the ball $\overline{B_{2\epsilon_{ub}(\hat{x})}}(\hat{x})$ and the upper bound holds

$$\|e\|_X = \|\hat{x} - x^*\|_X \leq \Delta_{ub}(\hat{x}) := \frac{1}{1 - \tau_{ub}(\hat{x})/2}\epsilon_{ub}(\hat{x}) \leq 2\epsilon_{ub}(\hat{x}). \tag{10}$$

Furthermore, if there exists a constant $C_\epsilon(\hat{x}) > 0$ such that $\epsilon_{ub}(\hat{x}) \leq C_\epsilon(\hat{x})\epsilon(\hat{x})$ and $\mathrm{DG}|_{\hat{x}}^{-1}(G(\cdot))$ is locally Lipschitz-continuous around $\hat{x}$ with constant $C_G(\hat{x})$, the *effectivity estimate*

$$\operatorname{eff}_{ub}(\hat{x}) \leq \frac{C_\epsilon(\hat{x}) C_G(\hat{x})}{1 - \tau_{ub}(\hat{x})/2},$$

where $\operatorname{eff}_{ub}(\hat{x})$ *is the effectivity for the error bound* $\Delta_{ub}(\hat{x})$ *holds.*

Proof The first statement (10) follows identical to Theorem 1. For proving the additional effectivity estimate, we infer

$$\Delta_{\text{ub}}(\hat{x}) = \frac{\epsilon_{\text{ub}}(\hat{x})}{1 - \tau_{\text{ub}}(\hat{x})/2} \leq \frac{C_\epsilon(\hat{x})\epsilon(\hat{x})}{1 - \tau_{\text{ub}}(\hat{x})/2},$$

from which we can proceed similar to Lemma 2. □

Similar to Lemma 1, the error bound and effectivity bound can be improved if $L_{\text{ub}}(\alpha) \leq C\alpha$ for some $C > 0$ and if the modified validity criterion $4\gamma_{\text{ub}}(\hat{x}) C \epsilon_{\text{ub}}(\hat{x}) \leq 1$ holds.

2.2 Reaching high effectivities through auxiliary linear problems

In this section, we will see how a very sharp bound for $\epsilon(\hat{x})$ can be obtained with low additional computational overhead. As it was motivated in the introduction by a simple two-dimensional linear problem, the effectivity of the a-posteriori error bound deteriorates by a large factor if the calculation of $\epsilon(\hat{x})$ is split according to (5). Thus, the key towards highly effective (i.e. $\operatorname{eff}(\hat{x}) \approx 1$) error bounds lies in finding highly effective approximations or bounds to $\epsilon(\hat{x})$.

We first observe that the value of $\epsilon(\hat{x})$ can be calculated exactly by solving the following linear system

$$\mathrm{DG}|_{\hat{x}}\,(E(\hat{x})) = G(\hat{x}) \tag{11}$$

for $E(\hat{x}) \in X$, which then gives $\epsilon(\hat{x}) = \|E(\hat{x})\|_X$ by definition. While in the linear example, i.e. for linear G, this equation calculates the exact error, this is no longer the case for nonlinear G. We therefore refer to (11) as the auxiliary linear problem (ALP) and will consequently denote the here presented error bounds as ALP-based error bounds. Although linear problems of the form (11) are often relatively easy to solve, it can be prohibitive to do so in high-dimensional or multi-query scenarios. To obtain a computationally efficient scheme, instead of requiring the true solution $E(\hat{x})$, we assume to have a suitable method that can be used to calculate an approximate solution $\widehat{E}(\hat{x}) \in X$.

Remark 2 One technique that is often applied to solve nonlinear problems of the form $G(x) = 0$ is the Newton-iteration. Based on an initial guess $x_0 \in X$, the following procedure is performed in an iterative manner:

$$x_{n+1} = x_n + \Delta x_n, \quad \text{with} \quad \mathrm{DG}|_{x_n}\,(\Delta x_n) = -G(x_n), \quad n \geq 0.$$

Hence, we can see that the computation of $E(\hat{x})$ is equivalent to performing one step in the Newton-iteration and $\widehat{E}(\hat{x})$ can be considered as a quasi Newton update. Thus, the additional computational effort can either be used for improved error quantification—as in our case—or improved approximation quality by setting $\hat{\hat{x}} = \hat{x} - E(\hat{x})$ or $\hat{\hat{x}} = \hat{x} - \widehat{E}(\hat{x})$, respectively.

The strength of the proposed method lies in the fact that we can easily derive a rigorous bound for the quantitiy $\epsilon(\hat{x})$, when an approximation $\widehat{E}(\hat{x})$ is available:

Lemma 3 (Upper bound for $\epsilon(\hat{x})$) *Let $\widehat{E}(\hat{x}) \in X$ be an approximate solution to the ALP (11) and define the ALP residual $R(\hat{x}) := \mathrm{DG}|_{\hat{x}}\,(\widehat{E}(\hat{x})) - G(\hat{x})$. Then the upper bound*

$$\epsilon(\hat{x}) \leq \epsilon_{ub}(\hat{x}) := \|\widehat{E}(\hat{x})\|_X + \gamma_{ub}(\hat{x})\|R(\hat{x})\|_Y. \tag{12}$$

holds true.

Proof The proof is a straightforward application of the triangle inequality. It holds

$$\epsilon(\hat{x}) = \|E(\hat{x})\|_X = \|E(\hat{x}) + \widehat{E}(\hat{x}) - \widehat{E}(\hat{x})\|_X \leq \|\widehat{E}(\hat{x})\|_X + \|\widehat{E}(\hat{x}) - E(\hat{x})\|_X.$$

For the difference $\widehat{E}(\hat{x}) - E(\hat{x})$, we make use of the linearity of the ALP and obtain the relation $\mathrm{DG}|_{\hat{x}}\,(\widehat{E}(\hat{x}) - E(\hat{x})) = R(\hat{x})$, from which we get

$$\|\widehat{E}(\hat{x}) - E(\hat{x})\|_X = \|\,\mathrm{DG}|_{\hat{x}}^{-1}\,(R(\hat{x}))\|_X \leq \gamma(\hat{x})\|R(\hat{x})\|_Y \leq \gamma_{\mathrm{ub}}(\hat{x})\|R(\hat{x})\|_Y. \tag{13}$$

□

Provided that an efficient scheme for the approximation of $E(\hat{x})$ exists, the computational overhead for the calculation of ϵ_{ub} is not very large as it only requires the calculation of $\|\widehat{E}(\hat{x})\|_X$ and $\|R(\hat{x})\|_Y$. Many iterative solvers for large-scale linear systems provide the residual of the equation as an abortion criterion which can be directly used for the calculation of the residual norm $\|R(\hat{x})\|_Y$. Furthermore, no additional quantities are required: In particular, $\gamma_{\mathrm{ub}}(\hat{x})$ has to be calculated anyway for the evaluation of the error bound.

At this point, we want to emphasize that our numerical examples reveal very accurate error predictions when using $\Delta_{\mathrm{ub}}(\hat{x})$ from Theorem 2 with the choice $\epsilon_{\mathrm{ub}}(\hat{x})$ according to Lemma 3. One possible explanation for this observation can be deduced from

$$\|\hat{x} - x^*\|_X \leq \Delta(\hat{x}) \leq 2\epsilon(\hat{x}) \leq 2\epsilon_{\mathrm{ub}}(\hat{x}),$$

and the fact that $\epsilon_{\mathrm{ub}}(\hat{x})$ is a very accurate estimate of $\epsilon(\hat{x})$. In contrast to the original splitting of $\epsilon(\hat{x})$ in (5), the splitting in (13) does not deteriorate the bound $\epsilon_{\mathrm{ub}}(\hat{x})$ significantly since $\|R(\hat{x})\|_Y$ is often much smaller than $\|\widehat{E}(\hat{x})\|_X$. To quantify this observation rigorously, we use the following lemma.

Lemma 4 (Relation of $\epsilon(\hat{x})$ and $\epsilon_{\mathrm{ub}}(\hat{x})$) *Assume*

$$\frac{2\gamma_{ub}(\hat{x})\|R(\hat{x})\|_Y}{\|\widehat{E}(\hat{x})\|_X} \leq 1.$$

Then the following inequality holds true for ϵ_{ub} chosen as in (12).

$$\epsilon_{ub}(\hat{x}) \leq C_\epsilon(\hat{x})\epsilon(\hat{x}), \quad \textit{with} \quad C_\epsilon(\hat{x}) := \left(1 + 4\frac{\gamma_{ub}(\hat{x})\|R(\hat{x})\|_X}{\|\widehat{E}(\hat{x})\|_X}\right) \leq 3.$$

Proof Note that the following proof is similar to a proof for the effectivity of relative RB error bounds [18]: The proof follows with $E(\hat{x}) = \mathrm{DG}|_{\hat{x}}^{-1}(G(\hat{x}))$ and $\|E(\hat{x})\|_X \neq 0$

$$\begin{aligned}\epsilon_{\mathrm{ub}}(\hat{x}) &= \|\widehat{E}(\hat{x})\|_X + \gamma_{\mathrm{ub}}(\hat{x})\|R(\hat{x})\|_Y \\ &\leq \|E(\hat{x})\|_X + \|\widehat{E}(\hat{x}) - E(\hat{x})\|_X + \gamma_{\mathrm{ub}}(\hat{x})\|R(\hat{x})\|_Y \\ &= \left(1 + \frac{\|\widehat{E}(\hat{x}) - E(\hat{x})\|_X}{\|E(\hat{x})\|_X} + \frac{\gamma_{\mathrm{ub}}(\hat{x})\|R(\hat{x})\|_Y}{\|E(\hat{x})\|_X}\right)\|E(\hat{x})\|_X. \qquad (14)\end{aligned}$$

From the triangle inequality and (13), we infer

$$\left|\frac{\|E(\hat{x})\|_X - \|\widehat{E}(\hat{x})\|_X}{\|\widehat{E}(\hat{x})\|_X}\right| \leq \frac{\|\widehat{E}(\hat{x}) - E(\hat{x})\|_X}{\|\widehat{E}(\hat{x})\|_X} \leq \frac{\gamma_{\mathrm{ub}}(\hat{x})\|R(\hat{x})\|_Y}{\|\widehat{E}(\hat{x})\|_X} \leq \frac{1}{2}.$$

If $\|\widehat{E}(\hat{x})\|_X > \|E(\hat{x})\|_X$, we thus get $\|\widehat{E}(\hat{x})\|_X - \|E(\hat{x})\|_X \leq \frac{1}{2}\|\widehat{E}(\hat{x})\|_X$, and hence $\frac{1}{2}\|\widehat{E}(\hat{x})\|_X \leq \|E(\hat{x})\|_X$. In the other case, i.e. $\|\widehat{E}(\hat{x})\|_X \leq \|E(\hat{x})\|_X$, the inequality $\frac{1}{2}\|\widehat{E}(\hat{x})\|_X \leq \|E(\hat{x})\|_X$ follows trivially. Hence, in total, we obtain

$$\frac{\|\widehat{E}(\hat{x}) - E(\hat{x})\|_X}{\|E(\hat{x})\|_X} \leq \frac{\gamma_{\mathrm{ub}}(\hat{x})\|R(\hat{x})\|_Y}{\|E(\hat{x})\|_X} \leq 2\frac{\gamma_{\mathrm{ub}}(\hat{x})\|R(\hat{x})\|_Y}{\|\widehat{E}(\hat{x})\|_X}.$$

Inserting this twice into (14) yields the final result. □

3 Highly accurate error bounds in the reduced basis context

In this section, we apply the proposed error bound within the RB framework. In particular, we explain how the a-posteriori error bound derived in Section 2 can be applied to parametric and nonlinear problems within the RB context.

3.1 Parametric nonlinear problems and the reduced basis method

In the following, we consider parametric problems. To this end, let $\mu \in \mathscr{P}$ be a parameter vector where $\mathscr{P} \subset \mathbb{R}^P$ for $P \in \mathbb{N}$ is a compact set of admissible parameters. The problems that we are interested in take the form

$$\text{For } \mu \in \mathscr{P} \text{ find } x^*(\mu) \in X \,:\, G(x^*(\mu); \mu) = 0, \qquad (P(\mu))$$

for the parameter-dependent operator $G(\cdot; \mu) : X \to Y$. In the following, we always assume that for every parameter $\mu \in \mathscr{P}$ at least one solution exists.

The idea behind RB methods is to determine a low-dimensional subspace $X_N \subset X$ with $N = \dim(X_N) \ll \dim(X) = d \leq \infty$ and to find approximate solutions in this subspace by solving an N-dimensional so-called reduced problem. To illustrate the procedure, we equip the approximation space X_N with a reduced basis

$\{\phi_1, \ldots, \phi_N\} \subset X$, of linearly independent basis elements $\phi_i \in X$. We then define the approximation $\hat{x}(\mu) \in X_N$ via

$$\hat{x}(\mu) := \sum_{i=1}^{N} x_{N,i}(\mu)\,\phi_i = \Phi x_N(\mu),$$

where the coefficient functions $x_{N,i} : \mathscr{P} \to \mathbb{R}$ are called reduced coordinates of the reduced coordinate vector $x_N = (x_{N,i})_{i=1}^N \in \mathbb{R}^N$ and where we introduce $\Phi := (\phi_1, \ldots, \phi_N)$ as the row vector of basis functions. By restricting the set of possible solutions of the problem ($P(\mu)$) to the subspace X_N and by projecting the residual $G(\hat{x}(\mu); \mu))$ to another low-dimensional subspace Y_N of dimension N, we arrive at the reduced problem:

$$\text{For } \mu \in \mathscr{P} \text{ find } \hat{x}(\mu) = \Phi x_N(\mu) \in X_N \;:\; G_N(\hat{x}(\mu); \mu) = 0, \qquad (P_N(\mu))$$

where the reduced problem is given as

$$G_N(\cdot; \mu) : X_N \to Y_N, \qquad G_N(\cdot; \mu) := \Pi_{Y_N}\left(G(\cdot; \mu)|_{X_N}\right).$$

Here, $\Pi_{Y_N} : Y \to Y_N$ denotes a projection onto the subspace Y_N, which we equip with a basis $\{\psi_1, \ldots, \psi_N\} \subset Y$. This procedure is commonly referred to as Petrov-Galerkin projection and it is widely used for projection-based model order reduction (MOR) methods.

The solvability of ($P_N(\mu)$) is typically ensured by a careful construction of the spaces X_N and Y_N. In the following, we always assume that all problems are solvable, i.e. in particular we can compute true solutions $x^*(\mu) \in X$ and approximations $\hat{x}(\mu) \in X_N$ for any parameter $\mu \in \mathscr{P}$. But as mentioned above, we do not require uniqueness.

3.2 Effective error prediction for the RB method

Given an approximate solution $\hat{x}(\mu) \in X_N$ to a true solution $x^*(\mu) \in X$, the fundamental question arises whether the norm of the error $e(\mu) := \hat{x}(\mu) - x^*(\mu)$ can be quantified rigorously and with good effectivity. To give a positive answer to this question, we apply Theorem 1. In the parametric setting, we are challenged with the requirement of calculating the following parameter-dependent quantities efficiently, where we often omit the explicit dependency on $\hat{x}(\mu)$ for the sake of readability:

$$\begin{aligned}
\gamma(\mu) &:= \| \mathrm{D}G(\cdot; \mu)|_{\hat{x}(\mu)}^{-1} \|_{\mathscr{L}(Y,X)},\\
\epsilon(\mu) &:= \| [\mathrm{D}G(\cdot; \mu)|_{\hat{x}(\mu)}^{-1}](G(\hat{x}(\mu); \mu)) \|_X,\\
L(\alpha; \mu) &:= \sup_{x \in \overline{B_\alpha}(\hat{x}(\mu))} \| \mathrm{D}G(\cdot; \mu)|_{\hat{x}(\mu)} - \mathrm{D}G(\cdot; \mu)|_x \|_{\mathscr{L}(X,Y)}.
\end{aligned}$$

Since a direct calculation of these quantities is often too expensive, we employ rapidly computable and (in the ideal case) rigorous upper bounds similar to the nonparametric case

$$\gamma(\mu) \le \gamma_{\mathrm{ub}}(\mu), \quad \epsilon(\mu) \le \epsilon_{\mathrm{ub}}(\mu), \quad L(\alpha; \mu) \le L_{\mathrm{ub}}(\alpha; \mu).$$

Although all quantities are important, we will primarily focus on the efficient calculation of $\epsilon_{\text{ub}}(\mu)$ and provide comments about the role of the other quantities in the subsequent section. We recall that $\epsilon(\mu)$ can be calculated explicitly by solving the following parametric linear equation for $E(\mu) \in X$

$$\text{For } \mu \in \mathscr{P} \text{ find } E(\mu) \in X \ : \ [\mathrm{D}G(\cdot;\mu)|_{\hat{x}(\mu)}](E(\mu)) = G(\hat{x}(\mu);\mu). \qquad (P^E(\mu))$$

and by computing $\epsilon(\mu) = \|E(\mu)\|_X$. Lemma 3 shows how an upper bound for $\epsilon(\mu)$ can be calculated based on an approximation $\widehat{E}(\mu) \in X$ of the solution $E(\mu)$. The idea to obtain such approximations in the context of RB methods is to employ another Petrov-Galerkin projection of the parametric ALP ($P^E(\mu)$) for a different pair of subspaces $X_M^E \subset X$, $Y_M^E \subset Y$ with $\dim(X_M^E) = \dim(Y_M^E) = M \ll d = \dim(X)$. We equip both subspaces with bases $\{\phi_1^E, \dots, \phi_M^E\} \subset X$ and $\{\psi_1^E, \dots, \psi_M^E\} \subset Y$ consisting of linearly independent basis functions and define the ansatz

$$\widehat{E}(\mu) := \sum_{i=1}^{M} E_{M,i}(\mu)\,\phi_i^E \in X_M^E, \quad \text{with} \quad E_M(\mu) := [E_{M,1}(\mu), \dots, E_{M,M}(\mu)]^T \in \mathbb{R}^M,$$

and project the ALP ($P^E(\mu)$) analogously to the original problem

$$\Pi_{Y_M^E}\left([\mathrm{D}G(\cdot;\mu)|_{\hat{x}(\mu)}](\widehat{E}(\mu))\right) = \Pi_{Y_M^E}\left(G(\hat{x}(\mu);\mu)\right). \qquad (P_M^E(\mu))$$

Note that this equation is of dimension M and can be solved efficiently, provided M is sufficiently small. To be able to state a rigorous upper bound $\epsilon_{\text{ub}}(\mu)$ for $\epsilon(\mu)$, we define the residual $R(\mu) \in Y$ of the approximation of the ALP as

$$R(\mu) := [\mathrm{D}G(\cdot;\mu)|_{\hat{x}(\mu)}](\widehat{E}(\mu)) - G(\hat{x}(\mu);\mu).$$

We then get from Lemma 3 the upper bound

$$\epsilon(\mu) \le \epsilon_{\text{ub}}(\mu) = \|\widehat{E}(\mu)\|_X + \gamma_{\text{ub}}(\mu)\|R(\mu)\|_Y. \qquad (15)$$

Based on this, we denote as $\Delta_{\text{ub}}(\mu)$ the parametric computable error bound stemming from Theorem 2 (10) where we use $\epsilon_{\text{ub}}(\mu)$ given by (15).

Remark 3 We want to note that $\Delta_{\text{ub}}(\mu)$ identifies snapshot reproduction. To see this, we first assume that the true solution for some parameter $\mu \in \mathscr{P}$ lives in our approximation space, i.e. $x^*(\mu) = \hat{x}(\mu) \in X_N$. It immediately follows, that $G(\hat{x}(\mu);\mu) = 0$ and thus the reduced ALP ($P_M^E(\mu)$) has the solution $\widehat{E}(\mu) = 0$ which leads to $R(\mu) = 0$ and $\epsilon_{\text{ub}}(\mu) = 0$. Hence, in total, $\Delta_{\text{ub}}(\mu) = 0$. Conversely, if we have $\Delta_{\text{ub}}(\mu) = 0$ for some parameter $\mu \in \mathscr{P}$ then also $\epsilon_{\text{ub}}(\mu) = 0$. We can now conclude that both $\widehat{E}(\mu) = 0$ and $R(\mu) = 0$. The latter again implies that $G(\hat{x}(\mu);\mu) = 0$, since $\mathrm{D}G(\cdot;\mu)|_{\hat{x}(\mu)}$ is invertible by assumption, which means that $x^*(\mu) = \hat{x}(\mu) \in X_N$.

3.3 Improvement of classical RB bounds for linear elliptic problems

The RB method is classically applied in the context of parametric PDEs. In this section, we recall the basic error estimation results for linear elliptic problems and relate them to the bound presented in the previous section.

Let X be suitable Hilbert (function) space and consider the following weak formulation of a parameterized PDE:

$$\text{For } \mu \in \mathscr{P} \text{ find } u(\mu) \in X \ : \ a(u(\mu), v; \mu) = f(v; \mu), \quad \forall v \in X. \tag{16}$$

We assume $a(\cdot, \cdot; \mu) : X \times X \to \mathbb{R}$ to be a continuous bilinear form and $f(\cdot; \mu) \in X'$, where X' denotes the dual space of X. We further assume the following essential properties that ensure the well-posedness of (16) for any $\mu \in \mathscr{P}$

$$\sup_{u \in X} \sup_{v \in X} \frac{|a(u, v; \mu)|}{\|u\|_X \|v\|_Y} =: c(\mu) \leq c_{\text{ub}}(\mu) < \infty, \quad \text{(continuity)},$$
$$\inf_{u \in X} \sup_{v \in X} \frac{|a(u, v; \mu)|}{\|u\|_X \|v\|_Y} =: \beta(\mu) \geq \beta_{\text{lb}}(\mu) > 0, \quad \text{(inf-sup stability)},$$

and for each $0 \neq v \in X$, there exists a $u \in X$ such that $a(u, v; \mu) \neq 0$. Provided these assumptions hold true, it is a well-known result that there exists a unique solution $u^*(\mu) \in X$ to the problem (16) (cf. [6]).

This problem fits in the general framework by setting $Y := X'$ and $G(\cdot; \mu) : X \to Y$ via

$$G(u; \mu)(v) := a(u, v; \mu) - f(v; \mu), \qquad \forall v \in X.$$

Let us now assume that an RB approximation $\hat{x}(\mu) \in X_N$ for some suitable subspace X_N with $N = \dim(X_N) \ll d$ is given. Then, the classical relation between the error $e(\mu) := \hat{x}(\mu) - x^*(\mu) \in X$ and the residual of the approximation is established via the the norm of the residual $G(\hat{x}(\mu); \mu)$ and reads as follows

$$\|\hat{x}(\mu) - x^*(\mu)\|_X \leq \Delta_{\text{RB}}(\mu) := \frac{\|G(\hat{x}(\mu); \mu)\|_{Y'}}{\beta(\mu)} \leq \frac{\|G(\hat{x}(\mu); \mu)\|_{Y'}}{\beta_{\text{lb}}(\mu)}.$$

In this setting, the equation $\frac{1}{\beta(\mu)} = \gamma(\mu)$ relates the inf-sup constant to the stability constant in the abstract formulation in this paper. Recall that for linear problems, we have

$$\|e(\mu)\|_X = \epsilon(\mu) \leq \epsilon_{\text{split}}(\mu) = \gamma(\mu)\|G(\hat{x}(\mu); \mu)\|_{X'} = \Delta_{\text{RB}}(\mu).$$

Hence, by not splitting the calculation of the residual and by directly applying an approximation scheme to $\epsilon(\mu)$, we can expect more accurate error predictions.

To apply the improved error estimation technique, we setup the ALP, whose weak form in the linear case is given via

$$a(e(\mu), v; \mu) = a(\hat{x}(\mu), v; \mu) - f(v; \mu), \quad \forall v \in X. \tag{17}$$

Since this equation is as expensive as the original problem, we perform the additional RB approximation of the ALP according to the framework described in the previous section. To this end, we assume to have another subspace $X_M^E \subset X$ with $\dim(X_M^E) = M \ll d$, which leads to the reduced ALP

$$a(\widehat{e}(\mu), v_M; \mu) = a(\hat{x}(\mu), v_M; \mu) - f(v_M; \mu), \quad \forall v_M \in X_M^E.$$

Recall that this is an M-dimensional equation that can be solved rapidly, similar to the RB approximation of the main problem. Based on the approximate solution, we then get according to Lemma 3 the error bound

$$\Delta_{\text{ub}}(\mu) = \|\widehat{e}(\mu)\|_X + \frac{1}{\beta_{\text{lb}}(\mu)}\|R(\mu)\|_Y,$$

where $R(\mu) \in Y$ is the Riesz-representative of the residual of (17), when the approximation $\widehat{e}(\mu)$ replaces the true error $e(\mu)$.

Often, the calculation of the inf-sup constant $\beta(\mu)$ poses many difficulties when it comes to an efficient implementation. As a remedy, one often employs pessimistic lower bounds $\beta_{\text{lb}}(\mu)$ to $\beta(\mu)$ that can be calculated rapidly. Such lower bounds can, for example, be computed by employing standard estimation techniques in the RB framework such as the min–θ scheme or the successive constraint method (SCM) (cf. [12, 18]). However, they are often either not applicable, computationally involved or deliver highly imprecise results that render the classical RB error bounds useless. The following example demonstrates the influence of $\beta(\mu)$ onto the classical and improved error bound and shows the benefit of using the results presented in this article. By using the upper bound $\Delta_{\text{ub}}(\mu)$ with $\epsilon_{\text{ub}}(\mu)$ and a lower bound of the inf-sup constant $\beta_{\text{lb}}(\mu) := \frac{\beta(\mu)}{\lambda}$ with a parameter $\lambda \geq 1$, we get the following estimates when using the classical RB bound and the improved version presented in this paper.

$$\Delta_{\text{RB}}(\mu) \leq \lambda \cdot \frac{\|G(\hat{x}(\mu); \mu)\|_Y}{\beta(\mu)} \quad \text{and} \quad \Delta_{\text{ub}}(\mu) \leq \|\widehat{E}(\mu)\|_X + \lambda \cdot \frac{\|R(\mu)\|_Y}{\beta(\mu)}.$$

Since we expect that $\|R(\mu)\|_Y \ll \|G(\hat{x}(\mu); \mu)\|_Y$, severe underestimations, i.e. assuming large λ, of the inf-sup constant have less impact in the nonsplit bound $\Delta_{\text{ub}}(\mu)$. In particular, this property might be useful in cases for which a (probably pessimistic) lower bound $\beta(\mu) \geq \underline{\beta} > 0$ for all $\mu \in \mathscr{P}$ is available, making expensive estimation techniques for those stability constants superfluous. We demonstrate the effect of this behavior in our numerical examples in Section 4.

3.4 Basis generation and offline/online efficient implementation

The essential idea of RB methods is to split the calculation into a potentially expensive offline phase where precomputations are performed which then allow a rapid online phase. During the offline step, the first task is to construct the subspaces, which in our case means finding a suitable basis for X_N, Y_N. To avoid technical difficulties and to ease the following, we only construct the ansatz space X_N and set $Y_N = (X_N)'$.

While there are many ways to determine suitable subspaces (cf. [10]), we focus on snapshot-based techniques. In this case, the subspace is contained in the span of several true solutions, i.e. $X_N \subset \operatorname{span}(\{x(\mu_1), \ldots, x(\mu_N)\})$ for suitable $\mu_1, \ldots, \mu_N \in \mathscr{P}$. Two popular techniques with this respect are the proper orthogonal decomposition (POD) method [25] and the class of greedy algorithms [24]. The first extracts the relevant information from a given set of solutions based on an eigenvalue decomposition of the empirical correlation operator of a set of snapshots $S := \{x(\mu_1), \ldots, x(\mu_N)\}$ (cf. [10]). Greedy procedures, on the other hand, determine the parameter whose solution should be used to enhance the space based on its current approximation quality. In this section, we focus on the latter; however, we also make use of the first in our numerical section. The general structure of the greedy algorithm is as follows and the pseudocode is given in Algorithm 1: Starting from an initial subspace $X_0 \subset X$ and a finite training set $\mathscr{P}_{\text{train}} \subset \mathscr{P}$, the maximum approximation error is sought by evaluating an error indicator $\delta(\cdot; \mu) : X \to \mathbb{R}_{\geq 0}$ for all reduced solutions with parameters in the training set $\mathscr{P}_{\text{train}}$. The subspace is then extended with the element that delivers the maximum error estimate and the loop continues until a prescribed tolerance is met. By this, we get an iterative scheme where the subspace is extended in each iteration.

Algorithm 1: Greedy algorithm($\mathscr{P}_{\text{train}}$, ρ, δ, X_0).

```
Data: Training set $\mathscr{P}_{\text{train}}$, greedy tolerance $\rho$, error indicator $\delta$, initial
      subspace $X_0$
Result: Subspace $X_N$.
while $\max_{\mu \in \mathscr{P}_{train}} \delta(x_N^*(\mu); \mu) > \rho$ do
    Set $\mu^* := \arg\max_{\mu \in \mathscr{P}_{\text{train}}} \delta(x_N^*(\mu); \mu)$;
    Solve full problem $G(x; \mu^*) = 0$ for $x^*(\mu^*) \in X$;
    Extend subspace $X_{N+1} := X_N \bigoplus \operatorname{span}(x^*(\mu^*))$;
    Increment $N := N + 1$;
end
```

For the approximation of the ALP, we have to identify another pair of subspaces X_M^E and Y_M^E. To this end, we proceed in an analogous fashion, i.e. we also restrict ourself to the case $Y_M^E = (X_M^E)'$; however, instead of solving the full problem ($P(\mu)$), we now solve the ALP ($P^E(\mu)$) in each greedy iteration. Additionally, the error indicator and tolerance have to be chosen in a sensible manner. In our case, we use the error indicator $\delta(\hat{E}(\mu)) := \Delta_{RB}^E(\mu) = \frac{\|R_E(\mu)\|}{\beta_{\mathrm{lb}}(\mu)}$ which is a suitable (and even rigorous) choice. Ideally, one would like to build an error space X_M^E in each iteration of Algorithm 1 to use the here proposed improved error estimators in the basis building process, i.e. one would need to perform a (nested) double greedy algorithm. However, this proves to be highly computationally expensive since the error space X_M^E depends on the approximation space X_N. Thus, we defer to a sequential computation of the spaces which makes the construction of both spaces computationally feasable. The pseudocode for this sequential double greedy algorithm is given in Algorithm 2.

Furthermore, this dependency of the error space X_M^E on the approximation space X_N limits the usefulness of our proposed bound as an error indicator for the greedy algorithm since the space X_M^E has to be rebuild in every iteration.

Algorithm 2: Sequential Double Greedy algorithm($\mathscr{P}_{\text{train}}$, $\mathscr{P}^E_{\text{train}}$, ρ, ρ_E, δ, δ_E, X_0, X_0^E).

```
Data: Training sets $\mathscr{P}_{\text{train}}$, $\mathscr{P}^E_{\text{train}}$, greedy tolerances $\rho$, $\rho^E$, error indicators
      $\delta$, $\delta^E$, initial subspaces $X_0$, $X_0^E$
Result: Subspaces $X_N$ and $X_M^E$.
while $\max_{\mu \in \mathscr{P}_{train}} \delta(x_N^*(\mu); \mu) > \rho$ do
    Set $\mu^* := \arg\max_{\mu \in \mathscr{P}_{\text{train}}} \delta(x_N^*(\mu); \mu)$;
    Solve full problem $G(x; \mu^*) = 0$ for $x^*(\mu^*) \in X$;
    Extend subspace $X_{N+1} := X_N \bigoplus \text{span}(x^*(\mu^*))$;
    Increment $N := N + 1$;
end
while $\max_{\mu \in \mathscr{P}^E_{train}} \delta_E(E_M(\mu); \mu) > \rho_E$ do
    Set $\mu^* := \arg\max_{\mu \in \mathscr{P}^E_{\text{train}}} \delta_E(E_M(\mu); \mu)$;
    Solve full problem $\text{DG}|_{\hat{x}(\mu^*)} (E; \mu^*) = G(\hat{x}(\mu^*); \mu^*)$ for $E(\mu^*) \in X$;
    Extend subspace $X_{M+1}^E := X_M^E \bigoplus \text{span}(E(\mu^*))$;
    Increment $M := M + 1$;
end
```

Remark 4 1. We want to note that in contrast to our improved bound which is less sensitive to underestimation in the inf-sup constant $\beta_{\text{lb}}(\mu)$, the chosen coarse error indicator δ might result in a less efficient approximation space.

2. The computational efficiency of our a-posteriori estimator is directly influenced by $\dim(X_M^E)$. Thus, if one wants to achieve a fast online phase including error estimation, one might have to make sacrifices in the quality of the error space such that the computational overhead is comparable with the computational demands required for solving the reduced problem ($P_N(\mu)$).

Finally, we shortly address how an efficient online phase can be achieved: The classical assumption that is made in this respect is the parameter separability of the problem. Given the parametric problem $G(\cdot; \mu) = 0$, we assume that it can be decomposed into an expansion of the form

$$G(\cdot; \mu) = \sum_{q=1}^{Q} \Theta_q(\mu) G_q(\cdot),$$

i.e. it consists of parameter-dependent coefficient functions $\Theta_q : \mathscr{P} \to \mathbb{R}$ and parameter-independent operators $G_q : X \to Y$. In cases where no such decomposition is present, the (discrete) empirical interpolation method can be employed (cf. [8, 16]). This property carries over to G_N as

$$G_N(\cdot;\mu) = \Pi_{Y_N}\left(G(\cdot;\mu)|_{X_N}\right) = \sum_{q=1}^{Q}\Theta_q(\mu)\Pi_{Y_N}\left(G_q(\cdot)\big|_{X_N}\right) =: \sum_{q=1}^{Q}\Theta_q(\mu)G_{N,q}(\cdot).$$

Thus, one can easily assemble the reduced system by precomputing $G_{N,q}$ during the offline-phase. In the same way, the above property is inherited by $\mathrm{D}G$ and thus the reduced error problem can be handled analogously.

4 Numerical examples

In this section, we evaluate the proposed a-posteriori error estimation theory in the context of the RB method. The first example is a well-known thermal-block test case, modeling a parametric heat conduction problem on the unit square. Here we will see that by making use of the proposed method, we are able to reach excellent effectivities in any norm that we consider. The second example shows the application of the framework to a nonlinear finite-dimension problem that stems from a semidiscretized parametric PDE with nonvariational finite difference (FD) discretization. Finally, in the last example, we consider a parametric ARE, i.e. a parametric nonlinear matrix-valued equation. All examples are implemented in the toolbox RBmatlab[1] and were run on a machine with an Intel Core i7-6700 CPU with 16GB RAM in MATLAB 2017a.

4.1 Standard linear test case: thermal block model

The thermal block example is a well-known test example in the RB community (cf. [10, 18]). It consists of a steady linear heat equation on the unit square $\Omega = (0,1)^2$, which is divided into $B := B_1 \cdot B_2$ subblocks, where $B_1, B_2 \in \mathbb{N}$ denote the number of subblocks per dimension. We denote the subblocks by Ω_i for $i = 1, \dots B$, counted rowwise starting from the left bottom. We prescribe a unit flux into the domain on the bottom boundary, which is denoted as $\Gamma_{N,1}$ with unit outward normal $n(\xi)$, where $\xi \in \Omega$ indicates the spatial variable. The left and right boundary part $\Gamma_{N,0}$ is insulated, which is modeled by a zero Neumann boundary condition and the top Dirichlet boundary Γ_D has constant 0 temperature. A schematic drawing of the domain is provided in Fig. 1a.

The parametric PDE for the temperature field $u(\cdot;\mu) : \Omega \to \mathbb{R}$ for this example is given as

$$\begin{aligned} -\nabla \cdot (\kappa(\xi;\mu)\nabla u(\xi;\mu)) &= 0, & \xi &\in \Omega, \\ u(\xi;\mu) &= 0, & \xi &\in \Gamma_D, \\ (\kappa(\xi;\mu)\nabla u(\xi;\mu)) \cdot n(\xi) &= i, & \xi &\in \Gamma_{N,i}, i = 0, 1, \end{aligned}$$

[1] http://www.morepas.org

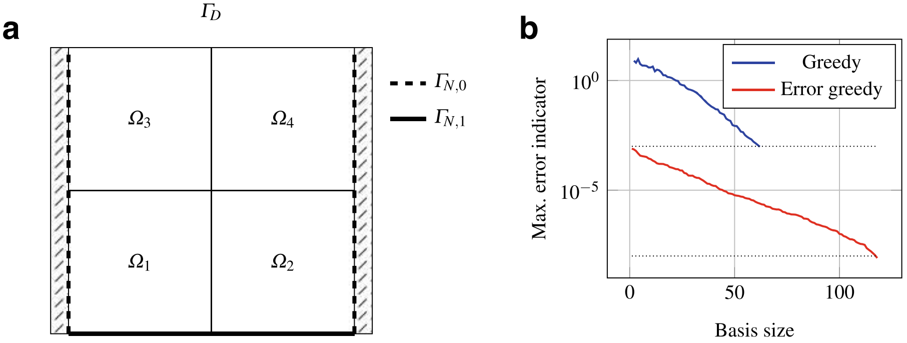


Fig. 1 Test 1: **a** Illustration of the thermal block setting used in the examples. **b** Decay of error indicator for the primal greedy and the greedy for the ALP

where we define the heat conductivity function

$$\kappa(\cdot;\mu) : \Omega \to 0, \quad \kappa(\xi;\mu) := \sum_{i=1}^{B} \mu_i \chi_{\Omega_i}(\xi),$$

using the indicator function χ_A for sets $A \subset \Omega$. The parametric domain for this problem is given as $\mathscr{P} := [1/\mu_{\max}, \mu_{\max}]^B$ for some $\mu_{\max} > 1$. With the function space $X = H^1_D(\Omega) := \{v \in H^1(\Omega) \mid v|_{\Gamma_D} = 0\}$ and its dual X', we define the problem $G(\cdot;\mu) : X \to X'$ via

$$G(u;\mu)(v) := \int_{\Omega} \kappa(\xi;\mu)\nabla u(\xi) \cdot \nabla v(\xi)\,\mathrm{d}\xi - \int_{\Gamma_{N,1}} v(\xi)\mathrm{d}\xi, \quad \forall v \in X.$$

We equip the space X with the norm $\|x\|_X = \|\nabla x\|_{L^2(\Omega)}$. It is a well-known fact that for every $\mu \in \mathscr{P}$, this problem possesses a unique solution $u^*(\mu) \in X$.

For the first tests, we pick $B_1 = B_2 = 3$, i.e. we have in total 9 parameters and $\mu_{\max} = 10$. For the truth-approximation, we apply a finite-element approximation of the PDE with piecewise linear elements resulting in a $d = 3\,721$ dimensional problem. The generation of the basis for the subspace X_N for the RB-approximation of the problem is performed with a standard greedy procedure, see also Section 3.4. For this, we define the residual-based error estimator $\delta(u_N(\mu)) := \Delta_{RB}(\mu) = \frac{\|G(u_N(\mu);\mu)\|_{X'}}{\beta_{\mathrm{lb}}(\mu)}$. Since we take the norms in the space $H^1_D(\Omega)$, we can define a lower bound to the inf-sup constant as $\beta_{\mathrm{lb}}(\mu) := \min_{i=1,\dots,B} \mu_i$. The basis is constructed on a finite training set consisting of $|\mathscr{P}_{\mathrm{train}}| = 1\,000$ random elements chosen uniformly from $\mathscr{P}$. We fix the tolerance $\rho = 10^{-3}$ which yields a basis for the approximation of size $N = 62$.

For the construction of the approximation space X^E_M for the approximation of the ALP $\mathrm{DG}|_{u_N(\mu)}(E(\mu)) = G(u_N(\mu);\mu)$, where $u_N(\mu) \in X_N$ is the RB-approximation of the solution, we perform another greedy procedure, i.e. we apply Algorithm 2. In particular, we again invoke the standard residual-based error estimator applied to the ALP and define

$$\delta_E(\mu) := \frac{\|\,\mathrm{DG}|_{u_N(\mu)}(\widehat{E}(\mu)) - G(u_N(\mu);\mu)\|_{X'}}{\beta_{\mathrm{lb}}(\mu)},$$

where $u_N(\mu)$ is the RB approximation obtained from the 62-dimensional subspace and $\widehat{E}(\mu)$ is the current approximation of the ALP for the parameter μ. We run the greedy algorithm for the ALP with the very low tolerance $\rho_E = 10^{-8}$. Furthermore, we reuse the same 1 000 parameters that were chosen for the greedy procedure for the main problem. The basis construction results in a subspace X_M^E of dimension $M = 118$. Note that due to linearity of the problem the ALP corresponds to solving another thermal block problem with a distributed source term that is given by the residual $G(u_N(\mu); \mu)$.

Figure 1b shows the decay of the error indicators for the main problem and the approximation of the ALP. We infer that the initial error for the ALP that is measured by the error indicator δ_E is very low and in the magnitude of the true error, which is why we had to set the tolerance to $\rho_E = 10^{-8}$.

In the following, we compare the improved error estimation techniques that are presented in this paper to the standard error bounds that are very widely used in the RB context. As a first test, we use the $H_D^1(\Omega)$-norm for the evaluation of the error bound and pick 20 random test parameters for the evaluation. For this test, we calculate the exact value of the stability constant $\gamma(\mu)$ by solving a d-dimensional eigenvalue problem. Clearly this is not online efficient but the purpose of the first test is to solely demonstrate the improved quality of the error estimates. The results are presented in Fig. 2: We show the absolute value of the true error $\|u_N(\mu) - u^*(\mu)\|_{H_D^1(\Omega)}$ as well as the standard RB-error bound $\Delta_{\text{RB}}(\mu) = \gamma(\mu)\|G(u_N(\mu); \mu)\|_{H_D^1(\Omega)}$. Recall that the latter choice corresponds to the split bound. The results in Fig. 2 show a very large overestimation in the range of $10-100$ for all test parameters. To demonstrate the improved error estimation and the arbitrarily high effectivity in the linear case, the error bound $\Delta_{\text{ub}}(\mu)$ with the improved approximation of $\epsilon(\mu)$ is shown in the same figure. To this end, we choose decreasing tolerances ρ_E and pick the subspace for $\widehat{E}(\mu)$ according to these tolerances. The plot clearly shows that an increasing dimension M improves the quality of the error

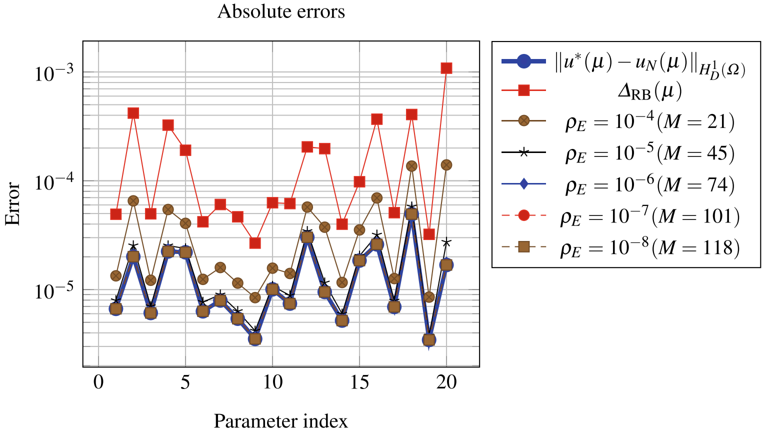


Fig. 2 Test 1: Absolute error measured in the $H_D^1(\Omega)$-norm for 20 random test parameters

estimation. In particular for $\rho_E \in \{10^{-6}, 10^{-7}, 10^{-8}\}$ ($M \in \{74, 101, 118\}$), we get almost exact error prediction.

In Fig. 3, we plot the true effectivities $\mathrm{eff}(\mu)$ along with the effectivity predictions from Lemma 4. Starting from $\rho_E = 10^{-5}$, the bound on the effectivity is applicable whereas for $\rho_E = 10^{-7}$, the predicted effectivity is close to the actual effectivity. The inapplicability of the effectivity bound can also be concluded from the results depicted in Figs. 4 and 5, where the approximation of the nonsplit residual $\|\widehat{E}\|_X$ and weighted error residual $\gamma\|R_E\|_Y$ is depicted for basis sizes $N \in \{40, 62\}$ and and error space sizes $M \in \{20, 40, 60, 80, 100, 118\}$. To this end, we recall that the effectivity bound holds only if the quotient $\frac{\gamma\|R_E\|_Y}{\|\widehat{E}\|_X}$ is bounded by $\frac{1}{2}$. We also note a rapid decline in the absolute value of the quotient which in turn translates to a effectivity (bound) close to 1, as can be seen in Fig. 3.

For the next test, we pick a larger test set $\mathscr{P}_{\text{test}} \subset \mathscr{P}$ consisting of 100 randomly chosen parameters. We compare the error estimation for the $H^1_D(\Omega)$, $L^2(\Omega)$ and the so-called energy norm $\|\cdot\|_{\bar{\mu}}$, which is defined as $\|u\|_{\bar{\mu}} := \sqrt{a(u, u; \bar{\mu})}$ for $u \in X$ and a fixed parameter $\bar{\mu} \in \mathscr{P}$. We pick the parameter $\bar{\mu} = (1, 2, 1, 2, \ldots, 1)^T$. It as well-known fact that the error estimation in the energy norm delivers very accurate results, which is also visible in Table 1 where the maximum and mean effectivity for all three norms are provided. The first row corresponds to the standard RB bound. The column entitled λ shows the factor by which we overestimate $\gamma(\mu)$, i.e. we pick the upper bound $\gamma_{\text{ub}}(\mu) := \lambda\gamma(\mu)$ for the calculations. In all cases, we observe a decay for decreasing tolerances ρ_E, i.e. richer subspaces X^E_M for the ALP. In particular, for the largest basis ($\rho_E = 10^{-8}$) and for $\lambda = 1$, we get exact error prediction over the whole parameter test set in the $H^1_D(\Omega)$ and energy norm. As expected, the influence of large overestimations of $\gamma(\mu)$ has much less impact on the improved norm in all three cases. Recall that for the classical RB bound, the scaling directly enters in the bound, i.e. $\lambda = 100$ means an additional degradation in the effectivity of factor 100 whereas we observe only factor $1.2 - 38$, depending on the chosen norm. Please note

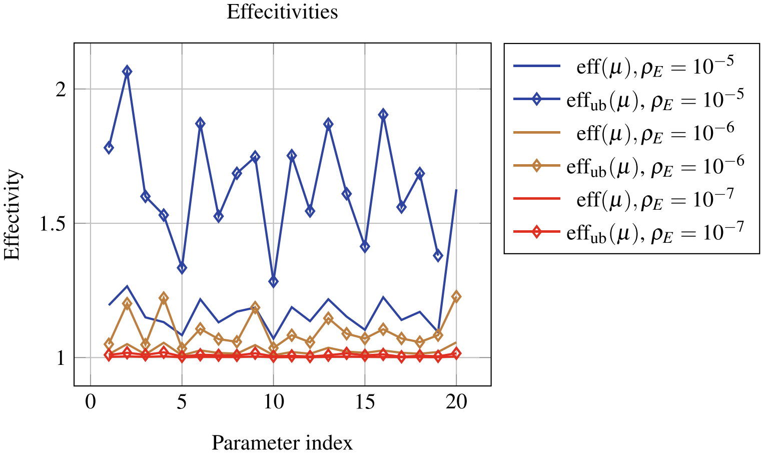


Fig. 3 Test 1: Effectivity of the $H^1_D(\Omega)$-norm error bound for 20 random test parameters

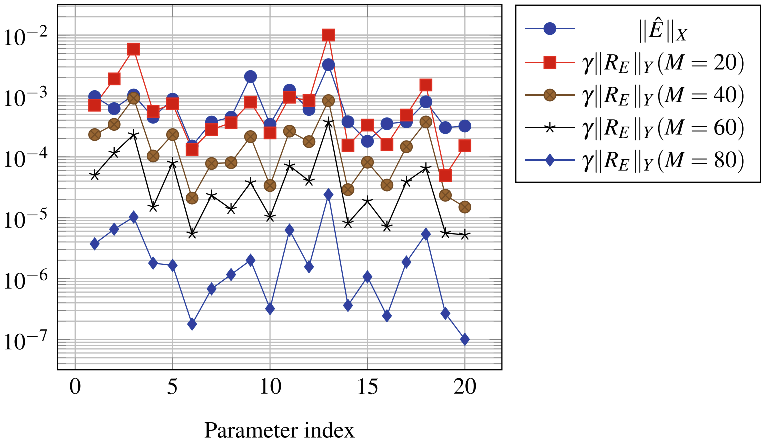


Fig. 4 Comparison of the nonsplit residual approximation and the weighted error residual for approximation space size $N = 40$ and varying error space sizes M

that in general, the residuals are only elements of $(H_D^1)'$, depending on the source term in the heat equation, and therefore the use of the $L^2(\Omega)$ norm is not possible. However, in our case, (no source term) and in the discrete setting, it is still applicable and was chosen to illustrate that the proposed method can be used for a variety of different norms.

Finally, we want to investigate the relation between the size of X_N and X_M which is needed to achieve a prescribed effectivity. For this purpose, we run the sequential double greedy algorithm (Algorithm 2) for $N \in \{10, 20, 30, 40, 50, 60\}$. We then select the basis size M of X_M in such a way that the effectivity of the error bound Δ_{ub} is smaller than 8,4,2, and 1.1 on a test set of 50 randomly chosen parameters. The results are depicted in Fig. 6. For the cases eff $\leq$ 4, 2, 1.1, we

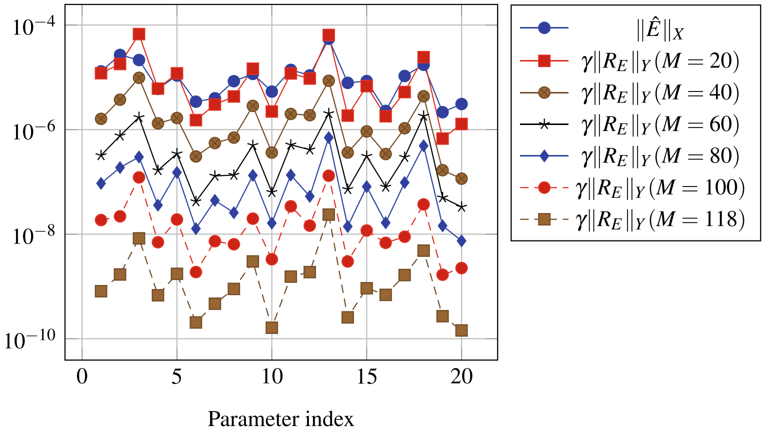


Fig. 5 Comparison of the nonsplit residual approximation and the weighted error residual for approximation space size $N = 62$ and varying error space sizes M

Table 1 Test 1: Maximum and mean effectivity of the error estimate for the thermal block example in three different norms

			Maximum			Mean		
ρ_E	M	λ	$\lVert\cdot\rVert_{L^2(\Omega)}$	$\lVert\cdot\rVert_{H^1_D(\Omega)}$	$\lVert\cdot\rVert_{\bar{\mu}}$	$\lVert\cdot\rVert_{L^2(\Omega)}$	$\lVert\cdot\rVert_{H^1_D(\Omega)}$	$\lVert\cdot\rVert_{\bar{\mu}}$
		1	10,018.07	64.27	42.83	3516.73	9.22	9.24
$1\cdot10^{-4}$	21	1	2, 025.01	9.94	7.02	513.37	2.29	2.28
$1\cdot10^{-5}$	45	1	384	2.86	2.25	73.46	1.19	1.18
$1\cdot10^{-6}$	74	1	55.5	1.31	1.21	8.58	1.02	1.02
$1\cdot10^{-7}$	101	1	7.63	1.03	1.02	1.88	1	1
$1\cdot10^{-8}$	118	1	1.58	1	1	1.11	1	1
$1\cdot10^{-8}$	118	10	6.48	1.02	1.01	2.08	1	1
$1\cdot10^{-8}$	118	100	59.36	1.22	1.15	11.79	1.02	1.02

The first row shows the results for the standard RB bound Δ_{RB}, the remaining rows for the proposed improved error estimate

notice an initial linear correlation between N, M. However, for larger values of N, a decay in the value of M can be observed. This can be attributed to the qualitatively better approximation $\widehat{E}$ of E for larger values of N. For the same pairs of N, M, we average the computation time for the calculation of the approximation u_N and the improved error bound Δ_{ub} for 200 randomly chosen parameter. The relative computational overhead which is required to compute the error bound is plotted in Fig. 7. We observe an overhead between 3 and 7 which decreases for increasing basis size M. This stems from the fact that for the basis sizes M used in this context, the computational cost of Δ_{ub} is dominated by the assembly of the reduced ALP. The relative impact of the assembly then decreases as the basis sizes N and M grow in value which in turn leads to a decrease in the overall relative computational overhead.

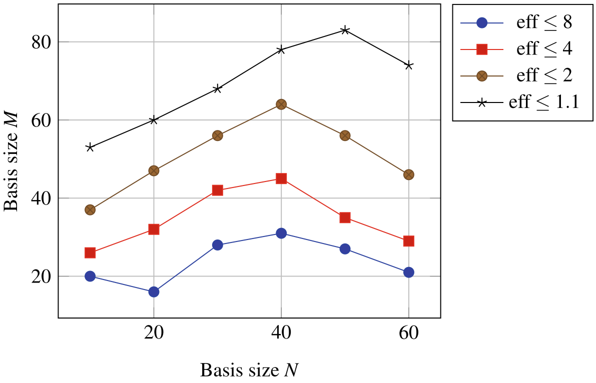


Fig. 6 Basis size $M(N)$ required to achieve prescribed effectivities

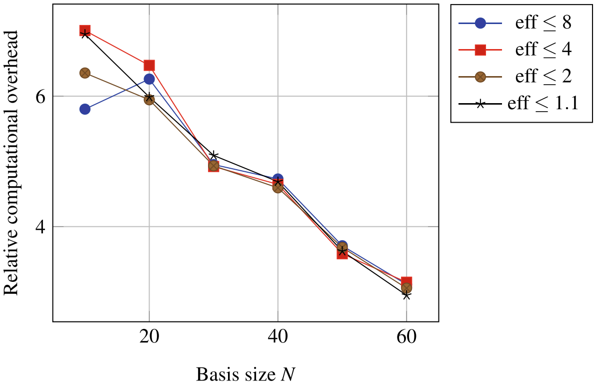


Fig. 7 Relative overhead for the computation of the improved error bounds for varying effectivites

4.2 Nonlinear finite-dimensional parametric problems

The second example stems from the finite-difference discretization of a nonlinear reaction-diffusion-advection equation. The infinite-dimensional description for this problem is given by defining for $\xi \in \Omega := (0, 1)$ the PDE

$$-\mu_1 \partial_{\xi\xi} u(\xi; \mu) + \partial_\xi u(\xi; \mu) - \mu_2 u(\xi; \mu)^2 = f(\xi), \quad \xi \in \Omega.$$
$$u(0; \mu) = u(1; \mu) = 0.$$

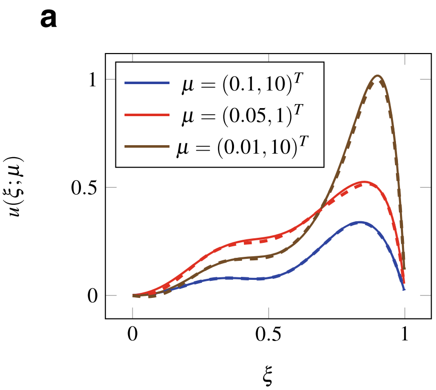
The parameter for this example stems from the set $\mu = (\mu_1, \mu_2)^T \in [0.1, 1]\times[1, 10]$, where μ_1 describes the diffusivity of the problem and μ_2 changes the influence of the nonlinearity. The right-hand side function (source term) is given as $f(\xi) := \sin(\xi\pi)^2$ for $\xi \in \Omega$. The PDE is discretized in space with a simple finite-difference scheme with upwind flux and results in a $d = 400$ dimensional nonlinear problem of the form $G(x; \mu) = 0$ with $G(x; \mu) := A(\mu_1)x + \mu_2 g(x) - f$ where $A(\mu) \in \mathbb{R}^{d\times d}$, $g : \mathbb{R}^d \to \mathbb{R}^d$ and $f \in \mathbb{R}^d$. In this case, we pick the finite-dimensional spaces $X = Y = \mathbb{R}^d$ and use the standard Euclidean norm for quantifying the error. We construct a subspace of dimension $N = 6$ by calculating snapshots for 100 random parameters chosen uniformly from $\mathscr{P}$ and by extracting the basis through the POD procedure. Note that this does not yield very accurate results for the RB-approximation but suffices to show the benefit of the improved error estimation theory presented in this paper. Figure 8a shows the solution to the full problem and to the reduced problem for three different parameters.

For evaluating the error bound, we have to calculate $\mathrm{DG}|_{\hat{x}(\mu)}$, which yields

$$\mathrm{DG}|_{\hat{x}(\mu)}(y) = A(\mu_1)y + 2\mu_2(\hat{x}(\mu) \circ y),$$

where $(a \circ b)_i := a_i b_i$ for $a, b \in \mathbb{R}^d$ denotes the componentwise product. From the explicit formula, we immediately get $L(\alpha; \mu) \le 2\mu_2\alpha =: L_{\text{ub}}(\alpha)$. In all of the following examples, the value of $\gamma(\mu) = \|DG(\cdot; \mu)|_{\hat{x}}^{-1}\|$ is calculated exactly by solving a high-dimensional eigenvalue problem.

For the application of the error bound, we have to construct a subspace X_M^E. To this end, we calculate solutions to the high-dimensional ALP for 50 random parameters



b

M	% valid	$\operatorname{mean}_{\mu\in\mathscr{P}_{\text{test}}} \text{eff}(\mu)$	$\max_{\mu\in\mathscr{P}_{\text{test}}} \text{eff}(\mu)$.
0	57	93.80	143.31
1	57	93.82	143.39
5	58	90.73	142.03
10	64	53.03	78.66
15	78	1.21	2.13
20	78	1.02	1.56
50	78	1.02	1.56

Fig. 8 **a** Example solutions for three different parameters. **b** Results for increasing basis size

from the training set and extract a basis by means of POD. For evaluating the bound, we calculate the error estimates for test parameters $\mathscr{P}_{\text{test}}$ consisting of 200 parameters chosen randomly from the parameter domain. We then vary the size of the RB space for the approximation of the ALP and evaluate the mean and maximum effectivity of the valid error estimations over the whole parameter test set $\mathscr{P}_{\text{test}}$. The results are presented in Fig. 8b. Note that a size of $M = 0$ corresponds to the split upper bound and represents the result that is classically used for error estimation in the RB context. Increasing the size again reveals a very accurate error prediction uniformly over the parameter space. Recall that in the nonlinear problem, the bound is only applicable if $\tau(\mu) \leq \tau_{\text{ub}}(\mu) \leq 1$, where $\tau_{\text{ub}}(\mu)$ is defined in (9). To show the benefit of using the improved error bound, the column entitled with "% valid" shows the fraction of valid error predictions. We observe that by increasing the dimension of X_M^E, the fraction increases from 57 to 78%. Note that the validity criterion is not always satisfied since the RB-approximation is too coarse. Hence, we cannot expect valid error estimations for all parameters unless we build richer subspaces X_N. Once again we want to highlight the fact that for the RB-approximation of the ALP, the solution to an M-dimensional linear problem is required.

4.3 Parametric algebraic Riccati equations

The ARE is a nonlinear matrix-valued equation with many applications in systems theory such as optimal (feedback) control or optimal state estimation, see [14]. For $X := \{A \in \mathbb{R}^{d\times d} \mid A = A^T\} =: \mathbb{R}^{d\times d}_{\text{sym}}$ and $\langle A, B\rangle_X := \text{trace}(A^T B)$, we define the mapping $G(\cdot;\mu) : X \to X$ via

$$G(P(\mu);\mu) := A(\mu)^T P(\mu) + P(\mu)A(\mu) - P(\mu)F(\mu)P(\mu) + Q(\mu), \qquad (18)$$

where $A(\mu) \in \mathbb{R}^{d\times d}$ and $F(\mu), Q(\mu) \in \mathbb{R}^{d\times d}_{\text{sym}}$ with $F(\mu)$ and $Q(\mu)$ being positive-semidefinite matrices. It is a well-known fact that this equation has a unique (stabilizing) solution $P^*(\mu) \in \mathbb{R}^{d\times d}_{\text{sym}}$ (i.e. the eigenvalues of $(A(\mu) - F(\mu)P(\mu))$ have negative real part and $P(\mu)$ is positive-semidefinite) provided the matrix

$A(\mu)$, $F(\mu)$, and $Q(\mu)$ satisfy specific conditions [15]. Often, in high-dimensional applications, the solution matrices $P^*(\mu)$ are of low numerical rank, meaning they can be efficiently approximated by low-rank factorizations of the form $P^*(\mu) \approx Z^*(\mu)Z^*(\mu)^T$ for $Z^*(\mu) \in \mathbb{R}^{d\times K}$ with $K \ll d$ (cf. [2]). This special structure is exploited in [20] in a parametric setting, where the low-rank factor greedy algorithm (LRFG) is introduced to make use of the special structure and to construct a suitable subspace for the RB-approximation of the ARE. This is done by defining the N-dimensional subspace $X_N := \{VP_NV^T \mid P_N \in \mathbb{R}^{N\times N}_{\text{sym}}\} \subset X$ for a suitable basis matrix $V \in \mathbb{R}^{d\times N}$ which is then used for the approximation via the Galerkin-projection $V^TG(VP_N(\mu)V^T;\mu)V = 0$. It can be shown that $P_N(\mu)$ solves an N-dimensional ARE that can be solved very efficiently for low N. Based on the low-dimensional matrix $P_N(\mu)$, we then define the approximation $\hat{P}(\mu) := VP_N(\mu)V^T$ and the error $e(\mu) = \hat{P}(\mu) - P^*(\mu)$. For measuring the error, we pick the spectral norm in the space X. The application of the error bound requires the quantities $L(\alpha;\mu)$ and $\|\mathrm{D}G(\cdot;\mu)^{-1}\|_{\mathscr{L}(X,X)}$. Due to the quadratic nature of the ARE, the derivative is readily calculated as

$$\mathrm{D}G(\cdot;\mu)|_{\hat{P}(\mu)}(P) = (A(\mu) - F(\mu)\hat{P}(\mu))^TP + P(A(\mu) - F(\mu)\hat{P}(\mu))$$

for some matrix $P \in \mathbb{R}^{d\times d}_{\text{sym}}$. Furthermore, an upper bound for $L(\alpha;\mu)$ can be derived

$$L(\alpha;\mu) \leq 2\|F(\mu)\|_X\alpha =: L_{\text{ub}}(\alpha).$$

The linearization $\mathrm{D}G(\cdot;\mu)|_{\hat{P}(\mu)}$ of the ARE results in a Lyapunov operator and the norm of its inverse is well studied (cf. [22]). For the following calculations, we use the exact value for $\gamma(\mu)$, which can be obtained by solving a high-dimensional Lyapunov equation and by taking the norm of the solution (cf. [13]).

We test the error estimation procedure by applying the RB-ARE method to a mathematical model of an optimal cooling process for steel profiles arising in rolling mills. The original model is a nonlinear heat equation for the temperature distribution of the cross-section of the steel profile with boundary control. In the technical application, the natural cooling process is supported by spraying cooling fluids onto the surface. The control objective is to optimally balance between a fast cooling process and an even temperature distribution. This is necessary to avoid deformations, brittleness, and other undesirable effects. A detailed explanation of the model and corresponding optimal control problem can be found in [5]. The resulting optimal control problem takes the form

$$\min_{u\in L^2(0,\infty;\mathbb{R}^7)} \int_0^\infty (y(t;\mu)^T\overline{Q}(\mu)y(t;\mu) + u(t)^T\overline{R}(\mu)u(t))\,\mathrm{d}t.$$
$$\text{subject to} \quad \overline{M}\dot{z}(t) = \overline{N}z(t) + \overline{H}u(t), \quad y(t;\mu) = \overline{C}z(t), \quad t \geq 0,$$

where we introduce parameter-dependent weights for the cost functional by setting $\overline{Q}(\mu) = I_6\mu_Q$ and $\overline{R}(\mu) = I_7\mu_R$ with $\mu_Q \in [10^{-4}, 0.1]$ and $\mu_R \in [10^{-4}, 1]$ and I_n

denoting the n-dimensional identity matrix. The unusual overbars are used to prevent notation doubles. The system matrices are given as $\overline{M}, \overline{N} \in \mathbb{R}^{d\times d}$, $\overline{H} \in \mathbb{R}^{d\times 7}$, and $\overline{C} \in \mathbb{R}^{6\times d}$ with $d = 20\ 209$. It is a well-known fact that the solution to this optimal control problem is given by finding the stabilizing solution $P^*(\mu)$ to the ARE (18) with $A(\mu) := \overline{M}^{-1}N$, $B(\mu) := \overline{M}^{-1}\overline{H}$, $F(\mu) := B(\mu)\overline{R}(\mu)^{-1}B(\mu)$ and $Q(\mu) := \overline{C}^T\overline{Q}(\mu)\overline{C}$ and by defining $u(t) = -\overline{R}(\mu)^{-1}B(\mu)P^*(\mu)z(t)$. The matrices can be downloaded from the MORWiki.[2]

The high dimension of the parametric ARE raises the need for efficient techniques for its solution. Hence, we apply the RB technique to the ARE. For testing the error bound, we construct an $N = 137$ dimensional subspace $X_N \subset \mathbb{R}^{d\times d}_{\mathrm{sym}}$ by running the LRFG algorithm on a test set consisting of 900 training parameters that were chosen from a grid consisting of 30×30 logarithmically distributed points in the parameter domain. Recall that the equation under consideration is matrix-valued. Hence, the reduction from d to N provides a huge computational benefit and speed-up. The basis generation for the ALP is performed by calculating the solutions $E(\mu)$ to the ALP $\mathrm{D}G(\cdot;\mu)|_{\hat{P}(\mu)}(E(\mu)) = G(\hat{P}(\mu);\mu)$ for a prescribed set of 50 parameters chosen randomly from the parameter domain. The basis is then extracted via a POD of the columns of the matrix $S := [E(\mu_1), \ldots, E(\mu_{50})]$ with a prescribed tolerance ρ_E for the “energy” that is contained in the singular values σ_k, i.e. we extract l basis elements with

$$l := \underset{j\in\{0,\ldots,d\}}{\arg\min}\ \frac{\sum_{k=1}^{j}\sigma_k^2}{\sum_{k=1}^{d}\sigma_k^2} \geq 1 - \rho_E.$$

We pick $\rho_E = 10^{-3}$, which results in an $M = 189$ dimensional subspace for the ALP. For details about the basis generation, we refer to [19].

For testing the error bound, we pick a test set $\mathscr{P}_{\mathrm{test}} \subset \mathscr{P}$ consisting of 100 parameters chosen randomly from the parameter domain. Figure 9a shows the evaluation of the maximum effectivity for the full error bound $\Delta(\mu)$, the upper bound $\Delta_{\mathrm{ub}}(\mu)$ with using $\epsilon_{\mathrm{ub}}(\mu)$, and the split bound $\Delta_{\mathrm{split}}(\mu)$. First of all, we observe a very good agreement of the true error and the full error bound, where no additional upper bounds for the quantities are employed. The mean overestimation of the nonsplit upper bound is relatively low which indicates good error prediction. However, when going from the nonsplit bound to the split bound, we see a large gap between the results. These results show that the improved error estimation presented in this paper is vital to get accurate predictions of the true error. As a last test, we explore the influence of the dimension of the subspace onto the quality of the error estimate. To this end, we vary the tolerance ρ_E for the construction of X_M^E, determine the corresponding subspaces, calculate the error bound for those subspaces, and plot the worst-case (maximum) effectivity along with the dimension M in Fig. 9b. Of course, the dimension of the subspace grows with the extraction of more information up to $M = 254$. On the other hand, the maximum effectivity decreases to an almost perfect error prediction for $\rho_E = 10^{-6}$ with a maximum effectivity of only 2.6.

[2] http://modelreduction.org

a

	mean	max
$\|e\|_X$	$1.001 \cdot 10^0$	$1.030 \cdot 10^0$
Δ_{ub}	$9.363 \cdot 10^1$	$1.182 \cdot 10^2$
Δ_{split}	$3.502 \cdot 10^5$	$8.059 \cdot 10^5$

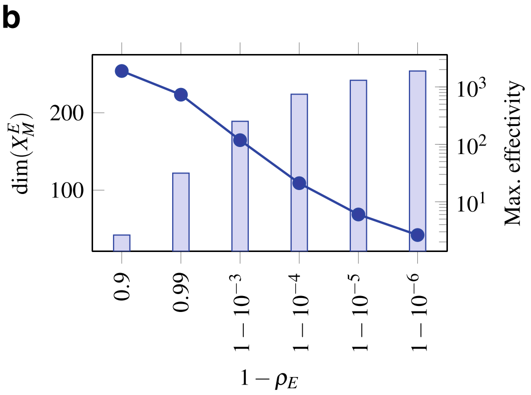


Fig. 9 **a** Mean and maximum effectivity for error bounds for the ARE. **b** Dimension of the subspace X_M^E (bars) and maximum effectivity (line) for varying ρ_E

5 Conclusion and outlook

In this article, we presented a novel improvement of error bounding techniques for problems which can be described as a zero value problem for differentiable operators over two Banach spaces. This was achieved by introducing and solving an auxiliary problem which counteracts the often severe overestimation that occurs when applying standard error bounding techniques and resulted in the here presented ALP-based error bounds. The resulting a-posteriori error bound shows significant improvement in its effectivity. Furthermore, the quality of the error prediction can be tuned by choosing richer subspaces for the approximation of $\widehat{E}$. The technique was then applied in the context of RB methods, where comparisons with standard error estimates were studied. Numerical examples for both, linear and nonlinear problems, highlight the benefits of the presented technique and show that we can reach effectivities that are very close to one in all examples.

Future work will study the application of the here presented method to time dependent-problems both continuous and discrete in time, as well as the applicability as an effective estimator for adaptive approximation schemes. A comparison of the proposed method to [11], especially in the case of linear problems, might be insightful.

Acknowledgments We also thank K. Smetana and the anonymous referees for valuable comments.

Funding information Open Access funding provided by Projekt DEAL. The authors would like to acknowledge the German Research Foundation (DFG) for financial support of the project within the Cluster of Excellence in Simulation Technology (EXC 310/2) at the University of Stuttgart.

Appendix A. Proof of Theorem 1

In the following, we make frequent use of the identity

$$G(x) - G(x') = \int_0^1 \mathrm{D}G|_{x'+t(x-x')}(x - x')\,\mathrm{d}t, \quad x, x' \in X \tag{19}$$

which is a direct application of the fundamental theorem of calculus.

Let $H : X \to X$ be defined via $H(x) := x - \mathrm{DG}|_{\hat{x}}^{-1}(G(x))$. The proof works by showing the existence of a fixed point $x^* \in X$ of the mapping H in the vicinity of the approximate solution $\hat{x}$. It is an easy observation that $G(x) = 0 \Leftrightarrow H(x) = x$, which motivates the application of Banach's fixed-point theorem. To this end, we define the set $M = \overline{B_{2\varepsilon}}(\hat{x}) := \{x \in X \mid \|x - \hat{x}\|_X \leq 2\varepsilon\}$, i.e. the closed ball around the approximate solution $\hat{x}$ with radius 2ε. In order to be able to apply Banach's fixed-point theorem to H in M, we have to prove that H is a self-mapping and a contraction in M.

Let $x \in M$. Consider

$$\begin{aligned} \|H(x) - \hat{x}\|_X &= \|x - \mathrm{DG}|_{\hat{x}}^{-1}(G(x)) - \hat{x}\|_X \\ &= \|\,\mathrm{DG}|_{\hat{x}}^{-1}\left[\mathrm{DG}|_{\hat{x}}(x - \hat{x}) - (G(x) - G(\hat{x}))\right] - \mathrm{DG}|_{\hat{x}}^{-1}(G(\hat{x}))\|_X \\ &= \|\,\mathrm{DG}|_{\hat{x}}^{-1}\left[\int_0^1 (\mathrm{DG}|_{\hat{x}} - \mathrm{D}G|_{\hat{x}+t(x-\hat{x})})(x - \hat{x})\mathrm{d}t\right] - \mathrm{DG}|_{\hat{x}}^{-1}(G(\hat{x}))\|_X. \end{aligned}$$

Since $\hat{x} + t(x - \hat{x}) \in M$ for $t \in [0, 1]$ we get the estimate

$$\begin{aligned} \|H(x) - \hat{x}\|_X &\leq \gamma \sup\nolimits_{z \in M} \|\,\mathrm{DG}|_{\hat{x}} - \mathrm{D}G|_z\|_{\mathscr{L}(X,Y)}\|z - \hat{x}\|_X + \epsilon \\ &\leq 2\gamma L(2\epsilon)\epsilon + \epsilon \leq 2\epsilon, \end{aligned}$$

which shows that $H(x) \in M$ for $x \in M$. Thus, H is a self-mapping in M. By making use of (19), we obtain the bound

$$\begin{aligned} \|H(x_1) - H(x_2)\|_X &= \|\,\mathrm{DG}|_{\hat{x}}^{-1}(\mathrm{DG}|_{\hat{x}}(x_1 - x_2) - (G(x_1) - G(x_2)))\|_X \\ &= \|\,\mathrm{DG}|_{\hat{x}}^{-1}\left(\int_0^1 (\mathrm{D}G|_{x_1} - \mathrm{D}G|_{x_1+t(x_2-x_1)})(x_1 - x_2)\mathrm{dt}\right)\|_X \\ &\leq \gamma L(2\varepsilon)\|x_1 - x_2\| \leq \frac{1}{2}\|x_1 - x_2\|, \end{aligned}$$

which proves the contraction property.

Hence, we can apply Banach's fixed-point theorem and prove the existence of $x^* \in M$ with $G(x^*) = 0$. We furthermore directly get the bound $\|x^* - \hat{x}\|_X \leq 2\epsilon$.

However, the bound can be slightly refined by considering for $x \in M$

$$\begin{aligned} \|x^* - x\|_X &= \|H(x^*) - x\|_X \\ &= \|\,\mathrm{DG}|_{\hat{x}}^{-1}\left(-G(x) + \int_0^1 (\mathrm{DG}|_{\hat{x}} - \mathrm{D}G|_{x^*+t(x-x^*)})(x - x^*)\,\mathrm{d}t\right)\|_X. \\ &\leq \epsilon + \gamma L(2\epsilon)\|x^* - x\|_X, \end{aligned}$$

from which we get for $x = \hat{x} \in M$ the final estimate

$$\|x^* - \hat{x}\|_X \leq \frac{\epsilon}{1 - \gamma L(2\epsilon)} \leq 2\epsilon.$$

Appendix B. Proof of Lemma 1

We only have to slightly modify the proof of Theorem 1. We consider the set $M = \overline{B_\alpha(\hat{x})} := \{x \in X \mid \|x - \hat{x}\|_X \leq \alpha\}$ and determine the minimal radius α such that H is a contracting self-mapping on M. Analogous to the proof of Theorem 1, we get

$$\begin{aligned} \|H(x) - \hat{x}\|_X &\leq \gamma \sup_{z \in M} \|\,\mathrm{DG}|_{\hat{x}} - \mathrm{D}G|_z\|_{\mathscr{L}}\|z - x\|_X + \varepsilon \\ &\leq \gamma L(\alpha)\alpha + \varepsilon \leq \gamma C\alpha^2 + \varepsilon \overset{!}{\leq} \alpha. \end{aligned}$$

Solving the resulting quadratic inequality, we have that α is contained in the interval $[\alpha_-, \alpha_+]$, where

$$\alpha_\pm = \frac{1 \pm \sqrt{1 - 4\gamma C\varepsilon}}{2\gamma C} = \frac{2}{1 \mp \sqrt{1 - 4\gamma C\varepsilon}}\varepsilon.$$

Hence, the smallest α for which H is a self-mapping is given by $\alpha_- = \frac{2}{1+\sqrt{1-4\gamma C\varepsilon}}\varepsilon$. For the proof of the contracting property, no further modification of the proof of Theorem 1 is necessary. Finally, it follows that

$$\|x^* - \hat{x}\| \leq \frac{2}{1 + \sqrt{1 - 4\gamma C\varepsilon}}\varepsilon.$$

References

1. Ainsworth, M., Oden, J.T.: A Posteriori Error Estimation in Finite Element Analysis. Wiley (2000)
2. Benner, P., Bujanović, Z.: On the solution of large-scale algebraic Riccati equations by using low-dimensional invariant subspaces. Linear Algebra Appl. **488**, 430–459 (2016). https://doi.org/10.1016/j.laa.2015.09.027
3. Benner, P., Gugercin, S., Willcox, K.: A survey of projection-based model reduction methods for parametric dynamical systems. SIAM Rev. **57**(4), 483–531 (2015). https://doi.org/10.1137/130932715
4. Benner, P., Ohlberger, M., Cohen, A., Willcox, K.: Model Reduction and Approximation. Society for Industrial and Applied Mathematics, Philadelphia (2017). http://epubs.siam.org/doi/abs/10.1137/1.9781611974829
5. Benner, P., Saak, J.: A semi-discretized heat transfer model for optimal cooling of steel profiles. In: Dimension Reduction of Large-Scale Systems, pp. 353–356. Springer (2005)

6. Braess, D.: Finite Elements. Cambridge University Press. https://books.google.de/books?id=PizECgOWoGgC (2007)
7. Caloz, G., Rappaz, J.: Handbook of Numerical Analysis, vol. 5, chap Numerical Analysis for Nonlinear and Bifurcation Problems, pp. 487–637 (1997)
8. Chaturantabut, S., Sorensen, D.: Nonlinear model reduction via discrete empirical interpolation. SIAM J. Sci. Comput. **32**(5), 2737–2764 (2010). https://doi.org/10.1137/090766498
9. Dolejší, V., Feistauer, M.: Discontinuous Galerkin Method Springer International Publishing. https://doi.org/10.1007/978-3-319-19267-3 (2015)
10. Haasdonk, B.: Reduced basis methods for parametrized PDEs – a tutorial introduction for stationary and instationary problems. In: Benner, P., Cohen, A., Ohlberger, M., Willcox, K. (eds.) Model Reduction and Approximation: Theory and Algorithms, pp. 65–136. SIAM, Philadelphia (2017). http://www.simtech.uni-stuttgart.de/publikationen/prints.php?ID=938
11. Hain, S., Ohlberger, M., Radic, M., Urban, K.: A hierarchical a posteriori error estimator for the reduced basis method. Advances in Computational Mathematics. https://doi.org/10.1007/s10444-019-09675-z (2019)
12. Huynh, D., Knezevic, D., Chen, Y., Hesthaven, J., Patera, A.: A natural-norm successive constraint method for inf-sup lower bounds. Comput. Methods Appl. Mech. Eng. **199**, 1963–1975 (2010). https://doi.org/10.1016/j.cma.2010.02.011. http://www.sciencedirect.com/science/article/pii/S0045782510000691
13. Kenney, C., Hewer, G.: The sensitivity of the algebraic and differential Riccati equations. SIAM J. Control. Optim. **28**(1), 50–69 (1990). https://doi.org/10.1137/0328003
14. Kwakernaak, H., Sivan, R.: Linear Optimal Control Systems, vol. 1. Wiley-interscience, New York (1972)
15. Lancaster, P., Rodman, L.: Algebraic Riccati Equations. Oxford University Press (1995)
16. Maday, Y., Nguyen, N.C., Patera, A.T., Pau, S.H.: A general multipurpose interpolation procedure: The magic points. Commun. Pure Appl. Anal. **8**(1534-0392-2009-1-383), 383 (2009). https://doi.org/10.3934/cpaa.2009.8.383. http://aimsciences.org//article/id/30a3894d-b0c8-4e29-8b8d-2611be32876f
17. Ohlberger, M.: A Posteriori Error Estimates and Adaptive Methods for Convection Dominated Transport Processes. Albert-Ludwigs-Universität Freiburg, Ph.D. thesis (2001)
18. Patera, A., Rozza, G.: Reduced Basis Approximation and a Posteriori Error Estimation for Parametrized Partial Differential Equations. To appear in (tentative) MIT Pappalardo Graduate Monographs in Mechanical Engineering. MIT (2007) http://augustine.mit.edu/methodology/methodology_book.htm
19. Schmidt, A.: Feedback Control for Parametric PDEs Using Reduced Basis Surrogate Models. Ph.D. thesis University of Stuttgart (2018)
20. Schmidt, A., Haasdonk, B.: Reduced basis approximation of large scale parametric algebraic Riccati equations. ESAIM: Control Optim. Calculus Variations **24**(1), 129–151 (2018). https://doi.org/10.1051/cocv/2017011
21. Steck, S., Urban, K.: A reduced basis method for the Hamilton-Jacobi-Bellmann equation with application to the european union emission trading scheme. Preprint, University of Ulm (2015)
22. Stykel, T.: Numerical solution and perturbation theory for generalized Lyapunov equations. Linear Algebra Appl. **349**(1-3), 155–185 (2002). https://doi.org/10.1016/S0024-3795(02)00255-0
23. Veroy, K., Patera, A.: Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: Rigorous reduced-basis a posteriori error bounds. Int. J. Numer. Methods Fluids **47**, 773–788 (2005)
24. Veroy, K., Prud'homme, C., Rovas, D., Patera, A.: A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. In: 16th AIAA Computational Fluid Dynamics Conference. American Institute of Aeronautics and Astronautics (2003). https://doi.org/10.2514/6.2003-3847. Paper 2003-3847
25. Volkwein, S.: Proper Orthogonal Decomposition: Theory and Reduced-Order Modelling. Lecture notes, Universität Konstanz (2013) http://www.math.uni-konstanz.de/numerik/personen/volkwein/teaching/POD-Book.pdf