

# Generalized reduced basis methods and $n$ -width estimates for the approximation of the solution manifold of parametric PDEs

Toni Lassila, Andrea Manzoni, Alfio Quarteroni, and Gianluigi Rozza

**Abstract** The set of solutions of a parameter-dependent linear partial differential equation with smooth coefficients typically forms a compact manifold in a Hilbert space. In this paper we review the generalized reduced basis method as a fast computational tool for the uniform approximation of the solution manifold. We focus on operators showing an *affine parametric dependence*, expressed as a linear combination of parameter-independent operators through some smooth, parameter-dependent scalar functions. In the case that the parameter-dependent operator has a dominant term in its affine expansion, one can prove the existence of exponentially convergent uniform approximation spaces for the entire solution manifold. These spaces can be constructed without any assumptions on the parametric regularity of the manifold – only spatial regularity of the solutions is required. The exponential convergence rate is then inherited by the generalized reduced basis method. We provide a numerical example related to parametrized elliptic equations confirming the predicted convergence rates.

## 1 Introduction

Reduced order models (ROMs) are a crucial ingredient of many applications of increasing complexity in scientific computing related e.g. to parameter estimation, sensitivity analysis, optimal control, and design or shape optimization. In this paper we consider the reduced basis method for the numerical approximation of parameter-dependent partial differential equations ( $\mu$ -PDEs). The set of solutions of such an equation depends on a finite-dimensional vector of parameters related e.g. to physical coefficients, geometrical configuration, source terms, and boundary conditions. Solving the  $\mu$ -PDE for many different values of the parameters entails the exploration of the manifold of solutions, and is not affordable if each  $\mu$ -PDE requires an expensive numerical approximation, such as the one built over the finite element method. Suitable structural assumptions about the parametrization enable one to decouple the computational effort into two stages. A (very expensive) pre-processing step that is performed once (“offline”) – consisting in the construction of a reduced basis for the representation of the manifold of solutions, followed by very inexpensive calculations performed “online” for each new input-output evaluation required.

In the reduced basis method, numerical solutions for certain parameters values are computed *offline* by a classical discretization technique. These solutions give a basis for approximating *online* the PDE solution (for a large number of new parameter values) as a linear combination of the basis elements. The rationale of this approach stands on the very fast (often exponential) convergence – with respect

---

Toni Lassila · Alfio Quarteroni · Gianluigi Rozza

MATHICSE-CMCS Modelling and Scientific Computing, École Polytechnique Fédérale de Lausanne, Switzerland, e-mail: [toni.lassila@epfl.ch](mailto:toni.lassila@epfl.ch), [alfio.quarteroni@epfl.ch](mailto:alfio.quarteroni@epfl.ch), [gianluigi.rozza@epfl.ch](mailto:gianluigi.rozza@epfl.ch)

Andrea Manzoni

SISSA Mathlab, SISSA - International School for Advanced Studies, Trieste, Italy, e-mail: [amanzoni@sissa.it](mailto:amanzoni@sissa.it)

Alfio Quarteroni

MOX Modeling and Scientific Computing, Dipartimento di Matematica, Politecnico di Milano, Italy

to the number of basis – is exhibited by approximation spaces. We point out that in the *real-time* or *many-query* contexts, where the goal is to achieve a very low *marginal cost* per input-output evaluation, we can accept an increased “offline” cost – not tolerable for a single or few evaluations – in exchange for greatly decreased “online” cost for each new/additional input-output evaluation.

In previous works (see e.g. [18, 19, 20]) *a priori* exponential convergence with respect to the number of basis functions is proved in the case of elliptic PDEs depending on one-dimensional parameters; several computational tests shown e.g. in [28] provide a numerical assessment of this behavior, also for larger parameter space dimensions. Several new results, such as the ones presented in [4], address an *a priori* convergence analysis in the more general case where a *greedy* algorithm is employed to build the reduced space in an automatic, adaptive way. A further improvement has been proposed in [3], where an error estimate for the greedy algorithm has been developed in terms of the Kolmogorov  $n$ -width. After recalling the basic features of a generalized version of the reduced basis method, and the main convergence results in this field, the goal of this paper is to provide both a convergence analysis for the greedy algorithm and a numerical proof of this behavior, in order to extend the *a priori* convergence results presented in [19, 20]. To do this, we rely on the introduction of a *fundamental basis*, a suitable error representation formula and an upper bound estimate for the  $n$ -width of the solution set of an elliptic parametric PDE under suitable assumptions on its parametric form.

We proceed to describe the functional setting of our problems. Let  $\Omega \subset \mathbb{R}^d$ ,  $d \in \{1, 2, 3\}$ , be a bounded domain and  $X = X(\Omega)$  a Hilbert space of functions defined on  $\Omega$  with inner product  $(\cdot, \cdot)_X$  and induced norm  $\|\cdot\|_X = \sqrt{(\cdot, \cdot)_X}$ . We consider the following problem: given a vector of parameters  $\mu \in \mathcal{D}$  from a compact parameter set  $\mathcal{D} \subset \mathbb{R}^P$ , find  $u(\mu) \in X$  s.t.

$$a(u(\mu), v; \mu) = f(v; \mu) \quad \text{for all } v \in X, \quad (1)$$

where the parameters can enter in the bilinear form  $a(\cdot, \cdot; \mu)$  in several possible ways: as variable coefficients, as coefficient entering in the parametrization of the domain  $\Omega \subset \mathbb{R}^d$  of the problem, in the definition of the right-hand side  $f$  that account for either forcing terms and/or boundary conditions. We denote the  $P$ -manifold of solutions

$$\mathcal{K} := \{u(\mu) \in X : \mu \in \mathcal{D} \subset \mathbb{R}^P\}$$

in the space  $X$ . In many applications  $\mathcal{K}$  is a differentiable manifold. We also allow the case of a manifold that is not locally smooth at some isolated points, e.g. the parametric Helmholtz equation  $\nabla \cdot (a(\cdot; \mu) \nabla u) + u = 0$ , which has a smooth solution manifold except at the eigenvalues of the parametric Laplacian,  $-\nabla \cdot (a(\cdot; \mu) \nabla u) = \lambda(\mu)u$ . We considered this problem in [17].

A typical objective in applications is to provide a numerical approximation  $\tilde{u}(\mu)$  for  $u(\mu) \in \mathcal{K}$  that is *uniform* over the entire manifold  $\mathcal{K}$ . The other option is to consider *local* or *sequential* approximations of the manifold, such as tracking a path on the manifold starting from a certain point and proceeding via a continuation method. In such cases we are usually not interested in the global behavior of the manifold. To fulfill the request of uniform approximation, a necessary condition is that, for any  $\varepsilon > 0$ , we find a linear subspace  $X_N \subset X$  of dimension  $N$  s.t.

$$\inf_{\tilde{u} \in X_N} \|u(\mu) - \tilde{u}\|_X < \varepsilon \quad \text{for all } \mu \in \mathcal{D},$$

where the dimension  $N$  is as small as possible. The question is: how small can we expect  $N$  to be?

To address this question, we introduce e.g. the finite element (FE) approximation of problem (1): given a vector of parameters  $\mu \in \mathcal{D} \subset \mathbb{R}^P$ , find  $u_{h,p}(\mu) \in X_{h,p} = X_{h,p}(\Omega)$  s.t.

$$a(u_{h,p}, v_{h,p}; \mu) = f(v_{h,p}; \mu) \quad \text{for all } v_{h,p} \in X_{h,p}, \quad (2)$$

where  $X_{h,p} \subset X$  is a conforming FE subspace spanned by piecewise polynomial shape functions of degree  $p$  defined on a quasi-uniform mesh of maximum element size  $h$ . Due to classical a priori error estimates such an approximation will eventually approximate well all the solutions on the manifold as the

dimension  $N := \dim(X_{h,p})$  increases, but only for quite large  $N$  we can expect a uniformly small error in the approximation. When  $X = H^1(\Omega)$  is the standard Sobolev space, the classical a priori estimates for piecewise polynomial approximations are as follows [2]:

$$\|u(\mu) - u_{h,p}(\mu)\|_1 \leq \begin{cases} C(\mu)N^{-\min\{s-1,p\}/d} & \text{with } h\text{-refinement} \\ C(\mu)N^{-(s-1-\delta)/d} & \text{with } p\text{-refinement} \end{cases}, \quad (3)$$

where  $s > 1$  denotes the number of weak derivatives of  $u(\mu)$ , and  $\delta > 0$  is arbitrarily small. If the solution  $u$  is analytic ( $s = \infty$ ), one obtains exponential convergence as a result of  $p$ -refinement, i.e.

$$\|u(\mu) - u_{h,\text{spectral}}(\mu)\|_1 \leq C(\mu) \exp(-\gamma N),$$

and this leads to the study of spectral methods. It should be cautioned that even if a spectral approximation can obtain in theory exponential convergence across the entire parameter range, the constants in front depend on both the dimension  $d$  and the number of parameters  $P$  of the problem. The assumption of analyticity of solutions is also often violated.

An efficient method for the approximation of the parametric manifold  $\mathcal{K}$  should (i) provide exponential convergence in the dimension  $N$  of the approximation space; (ii) have the same convergence rate irrespective of the number of parameters  $P$ ; and (iii) entail a computational cost that scales only moderately in  $N$ . Exploiting the *structure of the manifold*  $\mathcal{K}$  is key to finding uniform approximations that satisfy (i)–(iii). Our technique for proving exponentially convergent approximation estimates for the manifold of solutions relies on a series expansion of the solution  $u(\mu)$ . Series expansion solutions, either by separation of variables or by power series expansion for PDEs with analytical coefficients, are classical tools for existence proofs. Analytical power series expansions, such as the decomposition method of Adomian [1], are not competitive against good numerical approximation schemes in actually providing approximate solutions to PDEs, but they do provide an interesting approach to constructing convergence estimates. The novel contribution of this work is to consider the power series expansion method for parameter-dependent PDEs by searching for solutions in a parametrically separable form

$$u(\mu) = \sum_{k=0}^{\infty} \Theta_k(\mu) \Psi_k, \quad (4)$$

where the  $\Psi_k$  do not depend on  $\mu$  and the scalar functions  $\Theta_k(\mu)$ . The expansion (4) together with standard estimates for convergent power series then provides a construction of approximation spaces that are uniformly exponentially convergent over the entire parameter range.

In order to achieve separation w.r.t to the parameters, we must make suitable structural assumptions on the PDE. A typical assumption is that of affine dependence on the parameter, i.e. problem (1) is assumed to be of the form

$$\sum_{q=1}^{Q_a} \Theta_q^a(\mu) a_q(u, v) = \sum_{q=1}^{Q_f} \Theta_q^f(\mu) f_q(v) \quad \text{for all } v \in X, \quad (5)$$

where every  $a_q : X \times X \rightarrow \mathbb{R}$  and  $f_q : X \rightarrow \mathbb{R}$  are parameter-independent bilinear and linear forms respectively, whereas  $\Theta_q^a : \mathcal{D} \rightarrow \mathbb{R}$  and  $\Theta_q^f : \mathcal{D} \rightarrow \mathbb{R}$  are scalar coefficient functions depending only on the parameter (but not necessary in a smooth way). We shall next describe the generalized reduced basis method (GRBM) that, given assumption (5), satisfies (i)–(iii). We then discuss some recent theoretical approximation results linking the best possible approximation space for  $\mathcal{K}$  with the convergence rate obtained by the GRBM, and exhibit a model problem where, at least in the case we have  $Q_a = 2$  and some additional special structure on bilinear forms  $a_q$ , we indeed observe in practice the exponential convergence predicted by theory.

## 2 Generalized reduced basis method for uniform approximation of $\mu$ -PDEs

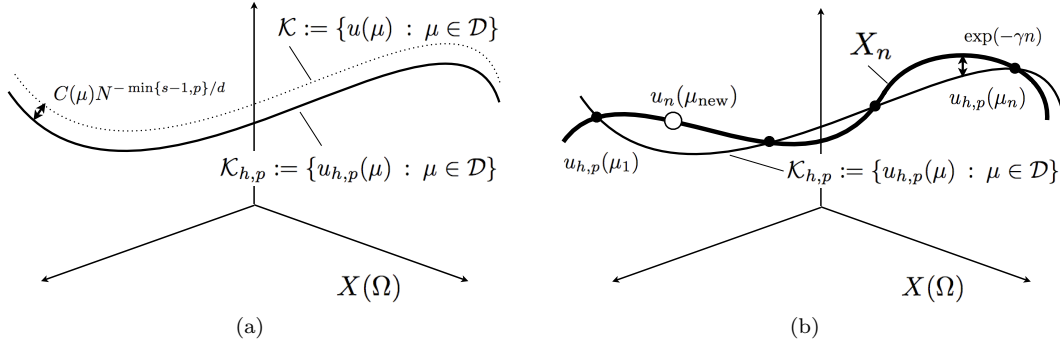
It is clear that the FE approximation  $u_{h,p}(\mu)$  of (1) can be made arbitrarily accurate for all possible parameters  $\mu \in \mathcal{D}$ , but this usually require a considerable computational cost. In order to overcome this (sometimes unaffordable) difficulty, a possible idea is to instead consider the manifold of discrete solutions  $u_{h,p}(\mu)$  given by

$$\mathcal{K}_{h,p} := \{u_{h,p}(\mu) \in X_{h,p} : \mu \in \mathcal{D} \subset \mathbb{R}^P\},$$

as a surrogate for  $\mathcal{K}$ , and then to look for approximations of  $\mathcal{K}_{h,p}$  that converge exponentially fast (see Fig. 1 for a graphical sketch). Specifically, we consider subspaces  $X_n \subset X_{h,p}$  (for  $n \ll N$ ) that are constructed by using information coming from the snapshot solutions  $u_{h,p}(\mu_i)$  computed at well-chosen points  $\mu_i$ ,  $i = 1, \dots, n$ . More precisely,  $X_n$  is the *span* of the snapshot solutions  $u_{h,p}(\mu_i)$ ,  $i = 1, \dots, n$ . This leads to the Reduced Basis (RB) method [27, 28], which is, in brief, a Galerkin projection on an  $n$ -dimensional approximation space relying on the parametrically induced manifold  $\mathcal{K}_{h,p}$ . Assuming that the solutions  $u_{h,p}$  satisfy the a priori convergence estimate (3) and that the approximation  $u_n(\mu)$  converges exponentially to  $u_{h,p}(\mu)$ , we can write the total error of the reduced solution as

$$\|u(\mu) - u_n(\mu)\|_X \leq \|u(\mu) - u_{h,p}(\mu)\|_X + \|u_{h,p}(\mu) - u_n(\mu)\|_X \leq C(\mu) \left[ N^{-\min\{s-1,p\}/d} + \exp(-\gamma n) \right],$$

and for  $N$  sufficiently large the exponential term in  $n$  dominates the error. The additional strength of this method is that in the best case we only need to solve  $n$  times the FE problem for  $u_{h,p}$ , and that the solution for  $u_n(\mu)$  can be done with complexity depending on  $n$  but not  $N$  after some initial preprocessing steps, so that indeed  $N$  can be chosen fairly large in order to obtain highly accurate reduced order solutions.



**Fig. 1** (a) Low-dimensional manifold  $\mathcal{K}_{h,p}$  on which the field variable resides and (b) approximation of a new solution at  $\mu_{\text{new}}$  with the “snapshots”  $u_{h,p}(\mu_m)$ ,  $1 \leq m \leq n$ .

For notational simplicity we present here the case  $P = 1$  and formulate the RB method in a slightly more general form than usually given. It can be expressed in three distinct steps:

- i) **Choice of the reduced subspace.** The basic idea of every reduced basis method is to choose a finite training sample  $\Xi_{\text{train}} \subset \mathcal{D}$ ,  $|\Xi_{\text{train}}| = M$ , in the parameter space and to use the information contained in the corresponding solutions  $u_{h,p}(\mu)$  for each  $\mu \in \Xi_{\text{train}}$  (called snapshots) to find a representative subspace for the approximation of the manifold  $\mathcal{K}_{h,p}$ . The reduced subspace  $X_n$  of dimension  $n$  is found by solving [28]

$$X_n := \arg \inf_{X^* \subset X_{\text{train}}, \dim(X^*)=n} \delta(X^*, \mathcal{K}_{h,p}; X) \quad (6)$$

where  $X_{\text{train}} := \text{span}\{u_{h,p}(\mu) : \mu \in \Xi_{\text{train}}\}$  is the space containing all the snapshots, and the function  $X^* \mapsto \delta(X^*, \mathcal{K}_{h,p}; X) \in \mathbb{R}$  measures the distance between any subspace  $X^* \subset X$  and the manifold  $\mathcal{K}_{h,p}$  and is defined by

$$\delta(X_n, \mathcal{K}_{h,p}; X) := \sup_{u \in \mathcal{K}_{h,p}} \inf_{\tilde{u} \in X_n} \|u - \tilde{u}\|_X.$$

Since the exact distance of the subspace to the manifold is usually unknown, we must resort to computable surrogates to solve (6). Nonetheless, we mention that the exact distance is used in the so-called *strong greedy* algorithm introduced in [3] for the theoretical analysis of convergence rates of reduced basis methods. Thus we replace (6) with

$$X_n := \arg \inf_{X^* \subset X_{\text{train}}, \dim(X^*)=n} M_{\mathcal{K}_{h,p}}(X^*), \quad (7)$$

where  $X^* \mapsto M_{\mathcal{K}}(X^*) \in \mathbb{R}$  is an *approximate distance* between any subspace  $X^* \subset X$  and the manifold  $\mathcal{K}_{h,p}$ . The choice of the function  $M_{\mathcal{K}}$  to be used for the approximation of  $\delta(\cdot, \cdot; \cdot)$  defines which algorithm we use to choose the subspace. This is by far the most common way of constructing reduced subspaces, and we call these approaches *Lagrange ROMs*. In the GRBM we consider also the parametric sensitivities up to a suitable order, say  $K - 1$  (with  $K \geq 1$ ) as part of the snapshot set

$$X_{\text{train}} := \text{span} \left\{ \frac{\partial^k u_{h,p}}{\partial \mu^k}(\mu) : \mu \in \Xi_{\text{train}}, \quad k = 0, \dots, K - 1 \right\},$$

giving a total number of  $MK$  snapshots. They can be computed from the *discrete sensitivity equation(s)*: find  $w_k(\mu) = \frac{\partial^k u_{h,p}}{\partial \mu^k}(\mu) \in X_{h,p}$  s.t.

$$a(w_k(\mu), v_{h,p}; \mu) = \frac{\partial^k f}{\partial \mu^k}(v_{h,p}; \mu) - \sum_{\ell=1}^k \binom{k}{k-\ell} \frac{\partial^\ell a}{\partial \mu^\ell}(w_{k-\ell}(\mu), v_{h,p}) \quad \text{for all } v_{h,p} \in X_{h,p}, \quad (8)$$

for all  $k = 1, \dots, K$ . Only the right-hand side of the system changes with  $k$  and thus any preconditioners or matrix decompositions used for the primal problem can be reused. The information contained in these snapshots is then used to build the reduced space  $X_n$  with dimension  $\dim X_n = n \ll MK$  in what can also be understood as a *data compression problem*. If  $K = 2$  and  $M > 1$  we have a *Hermite ROM*, and if  $K > 1$  and  $M = 1$  we have a *Taylor ROM*.

Two standard choices for  $M_{\mathcal{K}}$  are:

1. The *proper orthogonal decomposition* (POD), where

$$M_{\mathcal{K}}(X_n) = \frac{1}{M} \sum_{i=1}^M \|u_{h,p}(\mu_i) - \Pi_{X_n}(u_{h,p}(\mu_i))\|_X^2, \quad (9)$$

and  $\Pi_{X_n} : X \rightarrow X_n$  is the orthogonal projection w.r.t. the inner product of  $X$ . In this case, we choose the basis by minimizing the  $\ell^2(\Xi_{\text{train}})$  error in parameter space. It turns out that the optimal bases are hierarchical and are spanned by the leading  $n$  eigenvectors of the correlation matrix

$$\mathbb{C}_{ij} = \frac{1}{M} (u_{h,p}(\mu_j) - \bar{u}, u_{h,p}(\mu_i) - \bar{u})_X, \quad 1 \leq i, j \leq M,$$

where we have subtracted the mean of the snapshots

$$\bar{u} = \frac{1}{M} \sum_{i=1}^M u_{h,p}(\bar{\mu}).$$

The eigenpairs  $(\lambda_j, \psi_j)_{j=1}^M$  of  $\mathbb{C}$  (with the eigenvalues ordered in the decreasing order) are the solutions of

$$\mathbb{C}\psi_j = \lambda_j \psi_j, \quad j = 1, \dots, |\Xi_{\text{train}}|.$$

Then, the optimal basis for the  $n$ -th dimensional space  $X_n$  generated by minimizing (9) is given by

$$\chi_0 = \bar{u}, \quad \chi_j = \sum_{i=1}^M [\psi_j]_i (u_{h,p}(\mu_i) - \bar{u}), \quad 1 \leq j \leq n,$$

being  $[\psi_j]_i$  the  $i$ -th component of the  $j$ -th eigenvector. Extensions of the standard POD basis to incorporate parametric sensitivities (8) were presented in [6, 12, 13] and are not discussed in detail here.

2. The *greedy algorithm*, where

$$M_{\mathcal{K}}(X_n) = \sup_{\mu \in \Xi_{\text{train}}} \|u_{h,p}(\mu_i) - \Pi_{X_n}(u_{h,p}(\mu_i))\|_X \quad (10)$$

i.e. minimization of the  $\ell^\infty(\Xi_{\text{train}})$  error in parameter space. In practice no efficient algorithm exists to solve (6) for large-scale problems, so we approximate it by its relaxation

$$M_{\mathcal{K}}(X_n) = \sup_{\mu \in \Xi_{\text{train}}} \Delta_n(\tilde{u}(\mu)) \quad (11)$$

where  $\Delta_n(\tilde{u}(\mu))$  is a computationally inexpensive *a posteriori* error estimator for the quantity  $\|u_{h,p}(\mu) - \tilde{u}(\mu)\|_X$  that should satisfy

$$C_1 \Delta_n(\tilde{u}(\mu)) \leq \|u(\mu) - \tilde{u}(\mu)\|_X \leq C_2 \Delta_n(\tilde{u}(\mu)), \quad \forall \mu \in \mathcal{D} \quad (12)$$

for some constants  $C_1 > 0$ ,  $C_2 \geq 1$ . This corresponds to the approximate minimization of the  $\ell^\infty(\Xi_{\text{train}})$  error in parameter space. In this case we have a *weak greedy algorithm* as defined in [3].

Note that while conceptually the POD and the greedy algorithms can be cast in a similar framework, their practical implementations are quite different. The training set  $\Xi_{\text{train}} \subset \mathcal{D}$  needs to be reasonably dense in the parameter space for  $\mathcal{M}_{\mathcal{K}_{h,p}}(X^*)$  to be a good approximation of the true distance  $\delta(X^*, \mathcal{K}_{h,p}; X)$  for all subspaces  $X^* \subset X$ . In the POD one needs to compute the FE approximations  $u_{h,p}(\mu)$  for all the points in  $\Xi_{\text{train}}$ , which amounts to a considerable computational undertaking. In contrast, the weak greedy algorithm only needs to compute the exact solutions (and their parametric derivatives) at the  $n$  snapshots comprising the RB and only the computationally inexpensive *a posteriori* estimator  $\Delta_n(u_n(\mu))$  needs to be evaluated over the entire training set. The difference in the norms used ( $\ell^\infty$  for the greedy vs.  $\ell^2$  for the POD) also results in slightly different approximation behavior of the resulting bases. Typically the POD basis needed to reach a given tolerance is smaller in size but tends to be not as robust far away from the snapshots (see e.g. [28, 29]).

ii) **(Petrov-)Galerkin projection of the equations.** In the second step we perform projection of the original problem onto the reduced trial subspace  $X_n$  using the reduced test subspace  $Y_n$  to obtain the reduced basis approximation: find  $u_n(\mu) \in X_n$  s.t.

$$a(u_n(\mu), v_n; \mu) = f(v_n; \mu) \quad \text{for all } v_n \in Y_n, \quad (13)$$

where  $X_n = \text{span}\{\varphi_j\}_{j=1}^n$  and  $Y_n = \text{span}\{\psi_j\}_{j=1}^n$ . If  $Y_n = X_n$  this is a pure Galerkin method, otherwise it is a Petrov-Galerkin method. The Petrov-Galerkin approach is adopted if the underlying system is either nonsymmetric or noncoercive and can be interpreted as a form of stabilization of the ROM. Applying the assumption (5) to (13) leads to the discrete system

$$\sum_{q=1}^{Q_a} \Theta_q^a(\mu) A_q^n U^n = \sum_{q=1}^{Q_f} \Theta_q^f(\mu) F_q^n \quad (14)$$

where the matrices and vectors

$$\begin{bmatrix} A_q^n \end{bmatrix}_{i,j} := a_q(\varphi_j, \psi_i), \quad \begin{bmatrix} F_q^n \end{bmatrix}_i := F_q(\psi_i), \quad U^n \text{ the reduced solution,} \quad (15)$$

are dense but only of dimension  $n$  and more importantly can be assembled once and then stored. The system (14) is then assembled by evaluating the coefficient functions  $\Theta_q^a, \Theta_q^f$  and summing together the weighted contributions from all the parts of the decomposition, and solving one small dense linear

system. Assuming all the FE degrees of freedom are nodal, we can write the discrete projectors

$$\begin{bmatrix} \mathbb{X}_n \\ N \times n \end{bmatrix}_{i,j} := \varphi_j(x_i), \quad \begin{bmatrix} \mathbb{Y}_n \\ N \times n \end{bmatrix}_{i,j} := \psi_j(x_i) \quad (16)$$

where  $x_i$  are the nodal points in the full space  $X_N$ . The discrete matrices and vectors (15) can then be obtained as

$$A_q^n = \mathbb{Y}_n A_q \mathbb{X}_n^T, \quad F_q^n = \mathbb{Y}_n F_q,$$

and the approximation of the solution  $u_N(\mu)$  is obtained as  $u_n(x_i) = [\mathbb{X}_n^T U_n]_i$  for  $i = 1, \dots, N$ . From here on we use mainly the discrete forms of the equations.

iii) **Certification of the ROM with error bounds.** *A posteriori* error bounds are used to both (i) certify the GRBM solution during the online stage, and (ii) construct the reduced space by means of the *weak greedy* algorithm. For the sake of simplicity we treat the case of linear, elliptic and coercive  $\mu$ -PDEs – extensions to noncoercive and nonlinear problems can be found in [5, 8, 11, 21, 30]. Our error bounds rely on two basic ingredients: the dual norm of the residual and a lower bound of the stability factor (in this case, of the parameter-dependent coercivity constant). The residual  $r(v; \mu) \in X'_{h,p}$  is defined as

$$r(v; \mu) \equiv f(v; \mu) - a(u_n(\mu), v; \mu), \quad \forall v \in X_{h,p} \quad (17)$$

so that exploiting (2) and the bilinearity of  $a(\cdot, \cdot; \mu)$  we have the error representation for  $e(\mu) = u_{h,p}(\mu) - u_n(\mu) \in X_{h,p}$  given by

$$a(e(\mu), v; \mu) = r(v; \mu), \quad \forall v \in X_{h,p}. \quad (18)$$

As a second ingredient, we need a positive lower bound  $\alpha_h^{LB}(\mu)$  for the (discrete) coercivity constant  $\alpha_h(\mu)$ :

$$0 \leq \alpha_h^{LB}(\mu) \leq \alpha_h(\mu) := \inf_{w \in X_{h,p}} \frac{a(w, w; \mu)}{\|w\|_{X_{h,p}}^2} \quad \forall \mu \in \mathcal{D}, \quad (19)$$

whose efficient evaluation as a function of  $\mu$  is made possible thanks to the so-called successive constraint method (see e.g. [14, 15, 17] for a general description of this procedure). By combining (18) with 19 and using the Cauchy-Schwarz inequality, the following *a posteriori* error estimate in the energy norm holds (see [28] for a proof):

$$\left( \frac{\gamma_a(\mu)}{\alpha_h^{LB}(\mu)} \right)^{-1/2} \frac{\|r(u_n(\mu); \mu)\|_{X'_{h,p}}}{\sqrt{\alpha_h^{LB}(\mu)}} \leq \|u_n(\mu) - u_{h,p}(\mu)\|_X \leq \frac{\|r(u_n(\mu); \mu)\|_{X'_{h,p}}}{\sqrt{\alpha_h^{LB}(\mu)}}, \quad (20)$$

so that expression (12) is now made explicit, being

$$\Delta_n(\mu) := \frac{\|r(u_n(\mu); \mu)\|_{X'_{h,p}}}{\sqrt{\alpha_h^{LB}(\mu)}}, \quad C_1 := \inf_{\mu \in \mathcal{D}} \left\{ \left( \frac{\gamma_a(\mu)}{\alpha_h^{LB}(\mu)} \right)^{-1/2} \right\}, \quad C_2 := 1.$$

We conclude this section by mentioning a few works where generalized RB/POD methods were applied. In one of the the pioneering works on RBM, Noor [24] used instead a Taylor ROM to build a local reduced space that was used to trace the post-buckling behavior of a nonlinear structure. The continuation idea was used also by Peterson [25] to compute Navier-Stokes solutions with increasing Reynolds number flow over a forward facing step. Again a Taylor ROM was constructed and used to extrapolate an initial guess for the Newton method at a slightly higher Reynolds number. Ito and Ravindran [16] were perhaps the first ones to suggest using a Hermite ROM in a uniform approximation context, rather than in a pure continuation method. The Lagrange and Hermite ROMs were compared on a driven cavity problem, where the Hermite approach was somewhat superior. No stability problems were reported and the Hermite basis with only two basis functions was able to extrapolate solutions to much larger Reynolds numbers. In the works of Hay et al. [12, 13] sensitivity information was introduced into the proper orthogonal

decomposition framework. The parametric sensitivities of the POD modes were derived and computed. The test problems were related with channel flow around a cylindrical obstacle, either by using a simple parametrization as the Reynolds number, or a more involved geometric parametrization of the obstacle. The use of a Hermite ROM considerably improved the validity of the reduced solutions away from the parametric snapshots. However, in the more involved geometrical parametrization case the Hermite ROM failed completely, as it did not converge to the exact solution even when the number of POD modes was increased. Carlberg and Farhat [6] proposed an approach they call “compact POD”, based on goal-oriented Petrov-Galerkin projection to minimize the approximation error subject to a chosen output criteria, and including sensitivity information with proper weighting coming from the Taylor-expansion and including “mollification” of basis functions far away from the snapshot parameter. The application was the optimization of an aeroelastic wing configuration by building local ROMs along the path to the optimal wing configuration.

### 3 Approximation theoretical basis for the generalized reduced basis method

We now turn to the convergence analysis of approximations in the reduced subspaces that are obtained by (7) and the choice (11) for  $M_{\mathcal{K}_{h,p}}$ . We recall some recent theoretical results and provide an extension through an exponential convergence result. To do this, we rely on the introduction of a *fundamental* basis and on an intuitive error representation formula, which will be exploited in the numerical example discussed in the following section. We define the best approximation error of  $\mathcal{K}_{h,p}$  obtained by the greedy algorithm (6) as

$$\sigma_n(\mathcal{K}_{h,p}; X) := \sup_{u_{h,p} \in \mathcal{K}_{h,p}} \inf_{\tilde{u} \in X_n} \|u_{h,p} - \Pi_{X_n}(u_{h,p})\|_X.$$

A priori convergence estimates for reduced basis approximations have been demonstrated in simple cases, such as in [19], where it was found that for a specific problem exponential convergence was achieved

$$\sigma_n(\mathcal{K}_{h,p}; X) \leq C \exp(-n^\alpha), \quad \text{for some } \alpha > 0.$$

Recently much interest has been devoted to understanding why the weak greedy method (11) is able to give an approximation space  $X_n$  that exhibits exponential convergence in  $n$ . To express how well we are able to uniformly approximate a given manifold of solutions  $\mathcal{K}_{h,p}$  with a finite-dimensional subspace, we recall the notion of  $n$ -width [26, 22] that is used to measure the degree in which we can uniformly approximate a subset of the space  $X$  using finite-dimensional subspaces  $X_n$ . The Kolmogorov  $n$ -width is defined as

$$d_n(\mathcal{K}_{h,p}; X) := \inf_{X_n \subset X} \sup_{u_{h,p} \in \mathcal{K}_{h,p}} \inf_{\tilde{u} \in X_n} \|u_{h,p} - \tilde{u}\|_X \quad (21)$$

where the first infimum is taken over all linear subspaces  $X_n \subset X$  of dimension  $n$ . We also define the discrepancy between the subspace  $X_n$  and the manifold  $\mathcal{K}_{h,p}$  as

$$\delta(X_n, \mathcal{K}_{h,p}; X) = \sup_{u_{h,p} \in \mathcal{K}_{h,p}} \inf_{\tilde{u} \in X_n} \|u_{h,p} - \tilde{u}\|_X.$$

The subspace  $X_n$  is said to be optimal if  $\delta(X_n, \mathcal{K}_{h,p}; X) = d_n(\mathcal{K}_{h,p}; X)$ . In general, the optimal subspace w.r.t the Kolmogorov  $n$ -width (21) is not spanned by elements of the set  $\mathcal{K}_{h,p}$  being approximated, so that possibly  $d_n(\mathcal{K}_{h,p}; X) \ll \sigma_n(\mathcal{K}_{h,p}; X)$ . In the recent work [3] it was shown that

$$\sigma_n(\mathcal{K}_{h,p}; X) \leq \frac{2^{n+1}}{\sqrt{3}} d_n(\mathcal{K}_{h,p}; X),$$

and that this estimate cannot in general be improved. However, it was also shown that if the  $n$ -width converges at an exponential rate, say  $d_n(\mathcal{K}_{h,p}; X) \leq C \exp(-cn^\beta)$  for all  $n > 0$  and some  $\tilde{C}, c > 0$ , then

$$\sigma_n(\mathcal{K}_{h,p}; X) \leq \tilde{C} \exp\left(-\alpha n^{\beta/(\beta+1)}\right). \quad (22)$$



A tighter estimate was proved for the case of algebraic convergence: if  $d_n(\mathcal{K}_{h,p}; X) \leq CMn^{-\alpha}$  for all  $n > 0$  and some  $M, \alpha > 0$ , then also

$$\sigma_n(\mathcal{K}_{h,p}; X) \leq CMn^{-\alpha}. \quad (23)$$

The fast (exponential) convergence of numerical approximations is often linked to spectral approximations. In this way, the reduced basis method can be understood as a spectral method, where instead of using generic global polynomial basis functions we use problem-dependent global smooth approximation basis. The analyticity of the solutions of elliptic PDEs was exploited e.g. in a recent work [7] in the special case where  $\mathcal{K}_{h,p}$  is an analytic manifold. Using complex analysis techniques and a Taylor expansion approximation of the solution  $u_{h,p}$  and its parametric derivatives  $w_{m,k}$  for  $k = 1, 2, \dots$ , the authors obtained a convergence rate for a reduced basis approximation as

$$\|u_{h,p}(\mu) - \tilde{u}(\mu)\|_X \leq C(\mu) n^{-(1/p-1)},$$

where  $0 < p < 1$  is the  $\ell^p$ -summability exponent of a sequence related only to the diffusion coefficients of the problem. In particular, the convergence rate was independent of the spatial dimension  $d$  and the number of parameters  $P$ . In general, the reduced basis approximation of solutions of elliptic equations with regular coefficients has indeed been very successful. We would however like to convince the reader that analytic regularity of the solution manifold  $\mathcal{K}_{h,p}$  is not necessary in order to successfully apply the reduced basis method.

Unfortunately, very little seems to be known about the  $n$ -width of manifolds of solutions of  $\mu$ -PDEs. Very specific results concern special subspaces [10, 23]. For instance, if  $Y \subset X$  is a dense, compactly embedded, and bounded subspace with inner product  $(\cdot, \cdot)_Y$  then the  $n$ -width of a ball  $B_Y \subset Y \subset X$  of finite radius is

$$d_n(B_Y; X) = \sqrt{\lambda_{n+1}},$$

where  $\lambda_n$  are the eigenvalues of the problem

$$\begin{cases} \lambda \in \mathbb{R}, u \in V, u \neq 0 \\ (u, v)_Y = \lambda(u, v)_X \end{cases}$$

sorted in descending order (see [10], Theorem 4.5). We then obtain an algebraic decay rate for the  $n$ -width of  $B_Y$

$$d_n(B_Y; X) = Cn^{-s/d},$$

where  $s \geq 1$  similarly to (3). Provided that  $\mathcal{K}_{h,p} \subset B_Y$  we obtain also an upper bound for the  $n$ -width of  $\mathcal{K}_{h,p}$ , since then  $d_n(B_Y; X) \geq d_n(\mathcal{K}_{h,p}; X)$ , consequently only algebraic convergence of the reduced basis method is predicted by (23). Such results can be misleadingly pessimistic compared to practical experiences with reduced basis methods because they do not take into account the structure of the manifold  $\mathcal{K}_{h,p}$  nor the fact that approximation (4) inherits in some sense the structure of the manifold.

## 4 An extended result of exponential convergence

Let us now give an example of a  $\mu$ -PDE where the explicit dependence of the solution manifold on the parameters can be exhibited. Let us consider the parameter-dependent problem after discretization:

$$(\Theta_1^a(\mu)A_1 + \Theta_2^a(\mu)A_2)u = \sum_{q=1}^{Q_f} \Theta_q^f(\mu)F_q. \quad (24)$$

We assume that (i) the operator  $A_1$  is invertible; (iii) the problem satisfies a global condition for the spectral radius  $\rho$  being

$$\rho \left( \frac{\Theta_2^a(\mu)}{\Theta_1^a(\mu)} A_1^{-1} A_2 \right) < 1 \quad \text{for all } \mu \in \mathcal{D}, \quad (25)$$

which we interpret as meaning that the term  $\Theta_1^a(\mu)A_1$  dominates the original differential operator. Under this assumption, for a parametrized PDE under the form (24), we can prove an exponential convergence result, which extends the one reported in [19].

Let us remark that a problem such as (24) can arise for example from a discretized advection-diffusion or reaction-diffusion problem, where  $A_1$  contains the (dominant) diffusion operator and  $A_2$  contains all the other terms. We proceed to write explicitly the solution of this problem as

$$u = \left( I + \frac{\Theta_2^a(\mu)}{\Theta_1^a(\mu)} A_1^{-1} A_2 \right)^{-1} (\Theta_1(\mu)A_1)^{-1} \left( \sum_{q=1}^{Q_f} \Theta_q^f(\mu) F_q \right),$$

which by exploiting the global spectral condition (25) leads to the series expansion for the solution

$$u = \sum_{k=0}^{\infty} \sum_{q=1}^{Q_f} \frac{(-1)^k \Theta_2^k(\mu) \Theta_q^f(\mu)}{\Theta_1^{k+1}(\mu)} [A_1^{-1} A_2]^k A_1^{-1} F_q.$$

By defining the fundamental basis vectors  $\Psi_{k,q} := [A_1^{-1} A_2]^k A_1^{-1} F_q$ , for  $k = 0, 1, \dots$  and  $q = 1, \dots, Q_f$ , we can write the solution as a series

$$u = \sum_{k=0}^{\infty} \sum_{q=1}^{Q_f} \frac{(-1)^k [\Theta_2^a(\mu)]^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} \Psi_{k,q}. \quad (26)$$

Several remarks can be made about formula (26):

1. In the special case  $A_2 = 0$  the parametric dependence enters only through the r.h.s and as a consequence the series (26) truncates to a finite one

$$u = \sum_{q=1}^{Q_f} \frac{\Theta_q^f(\mu)}{\Theta_1^a(\mu)} \Psi_{0,q}, \quad (27)$$

and so the greedy algorithm will always terminate after  $Q_f$  steps.

2. If the decay of the series coefficients in (26) is rapid, the solutions  $u$  can be well approximated by only the first few fundamental basis functions  $\Psi_{k,q}$ ,  $k = 0, 1, \dots, K$  and  $q = 1, \dots, Q_f$ . They can be computed according to an iterative procedure

$$\Psi_{0,q} = A_1^{-1} F_q, \quad \Psi_{k+1,q} = A_1^{-1} A_2 \Psi_{k,q} \quad \text{for all } q = 1, \dots, Q_f.$$

requiring at each step one matrix multiplication and one backward substitution after obtaining once and for all the LU-decomposition of  $A_1$ .

3. In general the  $\Psi_{k,q}$  are not linear combinations of solutions of (24) so that they do not constitute a reduced basis approximation. They are, however, useful for estimating the  $n$ -width of the solution set. Provided that there exist positive sequences  $\{\gamma_{k,q}\}_{k=1}^{\infty}$  s.t.  $\|\Psi_{k,q}\|_X \leq \gamma_{k,q}$  for each  $q = 1, \dots, Q_f$ , we obtain an upper bound estimate for the  $n$ -width of the solution set  $\mathcal{U}$  of (24)

$$d_m(\mathcal{U}; X) \leq \sup_{\mu \in \mathcal{D}} \sum_{k=n}^{\infty} \sum_{q=1}^{Q_f} \left| \frac{[\Theta_2^a(\mu)]^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} \right| \gamma_{k,q} \quad (28)$$

by using the definition of the  $n$ -width, estimating upwards, and using formula (26) as

$$\begin{aligned}
d_m(\mathcal{U}; X) &= \inf_{X_m \subset X} \sup_{\mu \in \mathcal{D}} \inf_{\tilde{u} \in X_m} \|u_{h,p}(\mu) - \tilde{u}\|_X \leq \sup_{\mu \in \mathcal{D}} \inf_{\tilde{u} \in X_m^\Psi} \|u_{h,p}(\mu) - \tilde{u}\|_X \\
&\leq \sup_{\mu \in \mathcal{D}} \left\| u_{h,p}(\mu) - \sum_{k=0}^{n-1} \sum_{q=1}^{Q_f} \frac{(-1)^k [\Theta_2^a(\mu)]^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} \Psi_{k,q} \right\|_X \\
&= \sup_{\mu \in \mathcal{D}} \left\| \sum_{k=n}^{\infty} \sum_{q=1}^{Q_f} \frac{(-1)^k [\Theta_2^a(\mu)]^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} \Psi_{k,q} \right\|_X \leq \sup_{\mu \in \mathcal{D}} \sum_{k=n}^{\infty} \sum_{q=1}^{Q_f} \left| \frac{[\Theta_2^a(\mu)]^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} \right| \|\Psi_{k,q}\|_X,
\end{aligned}$$

where  $m := Q_f \cdot n$  and  $X_m^\Psi := \text{span}\{\Psi_{k,q} : k = 0, \dots, n-1, q = 1, \dots, Q_f\}$ , i.e. the first  $m$  fundamental basis vectors. We have in fact decomposed the description of the manifold of solutions  $\mathcal{U}$  into two parts: the *parametric regularity* is carried by the coefficients  $\Theta_1, \Theta_2, \Theta_q^f$ , which can be taken just in  $L^\infty(\mathcal{D})$  without affecting the  $n$ -width, and the *spatial regularity*, which is contained in the norm estimates  $\gamma_{k,q}$  for the fundamental basis functions.

4. If the solution of (24) is approximated by the projection-based ROM in (14), i.e.

$$\mathbb{Y}_n [\Theta_1^a(\mu) A_1 + \Theta_2^a(\mu) A_2] \mathbb{X}_n^T U_n = \sum_{q=1}^{Q_f} \Theta_q^f(\mu) \mathbb{Y}_n F_q,$$

where the projectors were defined in (16), we obtain a similar formula for the reduced solution

$$u_n(\mu) = \sum_{k=0}^{\infty} \sum_{q=1}^{Q_f} \frac{(-1)^k [\Theta_2^a(\mu)]^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} \Psi_{k,q}^n,$$

but now with the reduced fundamental basis functions  $\Psi_{k,q}^n$  defined as

$$\Psi_{0,q}^n = (\mathbb{Y}_n A_1 \mathbb{X}_n^T)^{-1} \mathbb{Y}_n F_q, \quad \Psi_{k+1,q}^n = (\mathbb{Y}_n A_1 \mathbb{X}_n^T)^{-1} \mathbb{Y}_n A_2 \mathbb{X}_n^T \Psi_{k,q}^n.$$

As a result, we obtain immediately the error representation formula

$$\varepsilon_n(\mu) := [u_{h,p} - \mathbb{X}_n^T u_n](\mu) = \sum_{k=0}^{\infty} \sum_{q=1}^{Q_f} \frac{(-1)^k [\Theta_2^a(\mu)]^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} (\Psi_{k,q} - \mathbb{X}_n^T \Psi_{k,q}^n).$$

Thus the quality of the ROM can directly be measured by observing how well it approximates the fundamental basis vectors, i.e. by looking at  $\|\Psi_{k,q} - \mathbb{X}_n^T \Psi_{k,q}^n\|_X$  for all  $k$  and  $q$ .

5. Even if the global spectral condition (25) does not hold, we can try to expand the solution locally around different  $\mu^*$  and obtain local approximation bases. This leads one to consider the  $hp$ -reduced basis method [9], where different reduced bases (analogous to  $p$ -refinement in the FEM) are constructed at different parts of the parameter domain (analogous to  $h$ -refinement in the FEM). Let  $\mathcal{D}_1, \dots, \mathcal{D}_M$  be a nonoverlapping subdivision of the original parameter domain  $\mathcal{D}$  into  $M$  subdomains. The local spectral condition requires that in each subdomain  $\mathcal{D}_m$

$$\exists i(m) : \quad \rho \left( \frac{\Theta_{j(m)}^a(\mu)}{\Theta_{i(m)}^a(\mu)} A_{i(m)}^{-1} A_{j(m)} \right) < 1 \quad \text{for all } \mu \in \mathcal{D}_m, \quad \text{for } j(m) \neq i(m),$$

that is to say in each parameter subdomain  $\mathcal{D}_m$  one of the terms  $A_q$  dominates, but the dominant part of the operator can change from subdomain to subdomain. If such a local spectral condition holds, our results extend straightforwardly to show the existence of local exponentially convergent approximation spaces.

With the  $n$ -width estimate (28) we can give an exponential convergence result for a problem under the form (24), extending that of [19]:

**Proposition 1.** *Assume that the series (26) converges, so that*

$$\exists \varepsilon > 0 \quad \text{s.t.} \quad \left| \frac{\Theta_2(\mu)}{\Theta_1(\mu)} \right| \leq \frac{1 - \varepsilon}{\|A_1^{-1}A_2\|_X} \quad \text{for all } \mu \in \mathcal{D}. \quad (29)$$

Then the  $n$ -width of the solution set  $\mathcal{U}$  of (24) converges exponentially, i.e.

$$d_n(\mathcal{U}; X) \leq Ce^{-\alpha n} \quad \text{for some } C, \alpha > 0. \quad (30)$$

*Proof.* The  $n$ -width upper bound (28) gives for  $m = n \cdot Q_f$

$$\begin{aligned} d_m(\mathcal{U}; X) &\leq \sup_{\mu \in \mathcal{D}} \sum_{k=n}^{\infty} \sum_{q=1}^{Q_f} \left| \frac{[\Theta_2^a(\mu)]^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} \right| \|(A_1^{-1}A_2)^k A_1^{-1}F_q\|_X \\ &\leq Q_f \cdot \sup_{\mu, q} \left\{ \left| \frac{\Theta_q^f(\mu)}{\Theta_1(\mu)} \right| \|A_1^{-1}F_q\|_X \right\} \cdot \sum_{k=n}^{\infty} \left| \frac{[\Theta_2^a(\mu)]^k}{[\Theta_1^a(\mu)]^k} \right| \|A_1^{-1}A_2\|_X^k \\ &= Q_f \cdot \sup_{\mu, q} \left\{ \left| \frac{\Theta_q^f(\mu)}{\Theta_1(\mu)} \right| \|A_1^{-1}F_q\|_X \right\} \cdot (1 - \varepsilon)^n \sum_{k=0}^{\infty} (1 - \varepsilon)^k \\ &= \frac{Q_f}{\varepsilon} \cdot \sup_{\mu, q} \left\{ \left| \frac{\Theta_q^f(\mu)}{\Theta_1(\mu)} \right| \|A_1^{-1}F_q\|_X \right\} \cdot \exp\left(\frac{\log(1 - \varepsilon)}{Q_f} m\right), \end{aligned}$$

so that the result holds with  $\alpha = -\log(1 - \varepsilon)/Q_f$  and  $C = \frac{Q_f}{\varepsilon} \cdot \sup_{\mu, q} \left\{ \left| \frac{\Theta_q^f(\mu)}{\Theta_1(\mu)} \right| \|A_1^{-1}F_q\|_X \right\}$ .  $\square$

Exponential convergence of the greedy reduced basis algorithm is then predicted by [3] as in (22). It should be understood that tight  $n$ -width estimates for the rate of exponential convergence cannot be obtained by such series expansions – indeed the coefficient  $\alpha$  will tend to 0 if we let  $\varepsilon \rightarrow 0$ . The factor  $1/Q_f$  in the exponential is also excessively pessimistic provided that not for all  $q$  the terms converge at the same rate. In the next section we will demonstrate a problem where much faster exponential convergence of the greedy algorithm is observed, even in the parametric region when the fundamental series no longer converges rapidly.

To close this section let us briefly consider the more general case  $Q_a > 2$ :

$$\left( \Theta_1^a(\mu)A_1 + \sum_{r=2}^{Q_a} \Theta_r^a(\mu)A_2 \right) u = \sum_{q=1}^{Q_f} \Theta_q^f(\mu)F_q. \quad (31)$$

If the global spectral condition

$$\rho \left( \sum_{r=2}^{Q_a} \frac{\Theta_r^a(\mu)}{\Theta_1^a(\mu)} A_1^{-1}A_r \right) < 1, \quad (32)$$

is satisfied, we can write the solution as

$$u = \left( I + \sum_{r=2}^{Q_a} \frac{\Theta_r^a(\mu)}{\Theta_1^a(\mu)} A_1^{-1}A_r \right)^{-1} (\Theta_1^a(\mu)A_1)^{-1} \left( \sum_{q=1}^{Q_f} \Theta_q^f(\mu)F_q \right),$$

and applying (32) leads to

$$u = \left( \sum_{k=0}^{\infty} \frac{(-1)^k}{[\Theta_1^a(\mu)]^{k+1}} \left[ \sum_{r=2}^{Q_a} \Theta_r^a(\mu) A_1^{-1} A_r \right]^k A_1^{-1} \right) \left( \sum_{q=1}^{Q_f} \Theta_q^f(\mu) F_q \right)$$

and finally

$$u = \sum_{k=0}^{\infty} \sum_{q=1}^{Q_f} \frac{(-1)^k \Theta_q^f(\mu)}{[\Theta_1^a(\mu)]^{k+1}} \Psi_{k,q}(\mu), \quad (33)$$

but now the fundamental basis vectors

$$\Psi_{0,q} = A_1^{-1} F_q, \quad \Psi_{k+1,q}(\mu) = \left[ \sum_{r=2}^{Q_a} \Theta_r^a(\mu) A_1^{-1} A_r \right] \Psi_{k,q}(\mu) \quad \text{for all } q = 1, \dots, Q_f. \quad (34)$$

depend explicitly on the parameter(s)  $\mu$ . Let  $\rho^{(k)} = (\rho_1, \rho_2, \dots, \rho_k)$  be a multi-index of dimension  $k$  and let  $\rho^{(0)} = \emptyset$ . We define a set of parameter-free basis functions  $\varphi_{k,q,\rho}$  according to the recursion

$$\varphi_{0,q,\rho^{(0)}} = A_1^{-1} F_q, \quad \varphi_{k+1,q,\rho^{(k+1)}} = A_1^{-1} A_{\rho_{k+1}} \varphi_{k,q,\rho^{(k)}}.$$

Using the parameter-free basis we can rewrite the recursion of the fundamental basis (34) as

$$\left\{ \begin{array}{l} \Psi_{0,q} = \varphi_{0,q,\rho^{(0)}} \\ \Psi_{1,q}(\mu) = \sum_{r=2}^{Q_a} \Theta_r^a(\mu) A_1^{-1} A_r \Psi_{0,q} = \sum_{r=2}^{Q_a} \Theta_r^a(\mu) \varphi_{1,q,(r)} \\ \Psi_{2,q}(\mu) = \sum_{r'=2}^{Q_a} \Theta_{r'}^a(\mu) A_1^{-1} A_{r'} \Psi_{1,q} = \sum_{r'=2}^{Q_a} \sum_{r=2}^{Q_a} \Theta_{r'}^a(\mu) \Theta_r^a(\mu) A_1^{-1} A_{r'} \varphi_{1,q,(r)} \\ \quad = \sum_{r'=2}^{Q_a} \sum_{r=2}^{Q_a} \Theta_{r'}^a(\mu) \Theta_r^a(\mu) \varphi_{2,q,(r',r)} \\ \Psi_{k,q}(\mu) = \sum_{r_1=2}^{Q_a} \dots \sum_{r_k=2}^{Q_a} \Theta_{r_1}^a(\mu) \dots \Theta_{r_k}^a(\mu) \varphi_{k,q,(r_k,r_{k-1},\dots,r_1)} \end{array} \right.$$

and so the  $k$ -th level expansion for  $\Psi_{k,q}$  will contain in general  $(Q_a - 1)^k$  terms, and the size of the expansion blows up exponentially. Without some strong structural assumptions the series expansion method is not suitable for deriving exponentially decaying  $n$ -width estimates in the case  $Q_a \gg 1$ .

## 5 Numerical example of a parameter-dependent diffusion problem

In this section we shall give numerical evidence of exponential convergence of  $n$ -width upper bounds (and consequently of the GRBM approximation) proved in the previous section. As a test problem we consider a diffusion problem in a disk with four circular subregions  $\Omega_1, \dots, \Omega_4$  as depicted in Fig. 2. The parametric problem can be formulated as follows: given  $\mu \in \mathcal{D} \subset \mathbb{R}^8$ , find  $u = u(\mu)$  s.t.

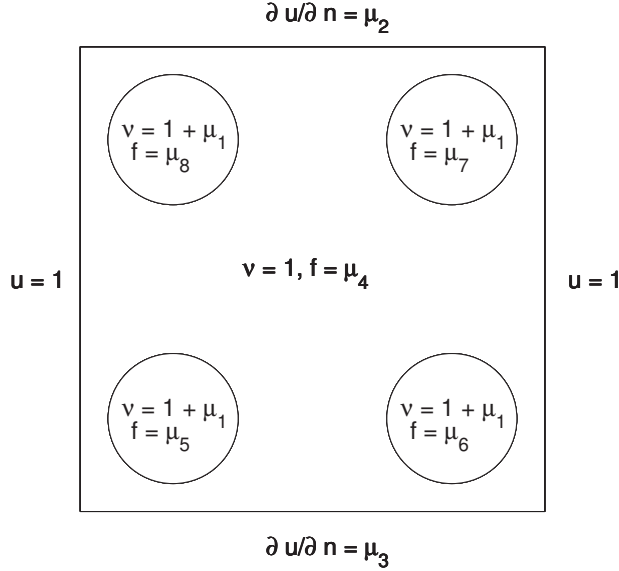
$$\begin{aligned} -(1 + \mu_1 \mathbb{I}_\omega) \Delta u &= \mu_4 \mathbb{I}_{\Omega \setminus \omega} + \sum_{q=1}^4 \mu_{q+4} \mathbb{I}_{\Omega_q}, & \text{in } \Omega \\ u &= 1 & \text{on } \Gamma_1 \cup \Gamma_4, \quad \frac{\partial u}{\partial n} = \mu_2 & \text{on } \Gamma_2, \quad \frac{\partial u}{\partial n} = \mu_3 & \text{on } \Gamma_3 \end{aligned}$$

where the  $\Gamma_k$  denote the four sides of the square, and  $\omega := \cup_{q=1}^4 \Omega_q$  is the union of the disks. The function  $\mathbb{I}_\Omega$  denotes the characteristic function of the subdomain  $\Omega$ . Thus the first parameter  $\mu_1$  controls the difference between the isotropic diffusion coefficient inside the disks versus the background conductivity,

while the rest of the parameters  $\mu_2, \dots, \mu_8$  enter into the boundary conditions and the source terms. This problem exhibits the same properties as the case discussed in the previous section, so that the solution can be written as the combination of the fundamental basis vectors thanks to the formula (26). In this case the affine expansion (5) of the problem is given by:

$$\left\{ \begin{array}{lll} \Theta_1^a(\mu) = 1, & a_1(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, d\Omega & \Theta_4^f(\mu) = \mu_5, \quad f_4(v) = \int_{\Omega_1} v \, d\Omega \\ \Theta_2^a(\mu) = \mu_1, & a_2(u, v) = \int_{\omega} \nabla u \cdot \nabla v \, d\Omega & \Theta_5^f(\mu) = \mu_6, \quad f_5(v) = \int_{\Omega_2} v \, d\Omega \\ \Theta_1^f(\mu) = \mu_2, & f_1(v) = \int_{\Gamma_2} v \, d\Omega & \Theta_6^f(\mu) = \mu_7, \quad f_6(v) = \int_{\Omega_3} v \, d\Omega \\ \Theta_2^f(\mu) = \mu_3, & f_2(v) = \int_{\Gamma_3} v \, d\Omega & \Theta_7^f(\mu) = \mu_8, \quad f_7(v) = \int_{\Omega_4} v \, d\Omega \\ \Theta_3^f(\mu) = \mu_4, & f_3(v) = \int_{\Omega \setminus \omega} v \, d\Omega & \end{array} \right.$$

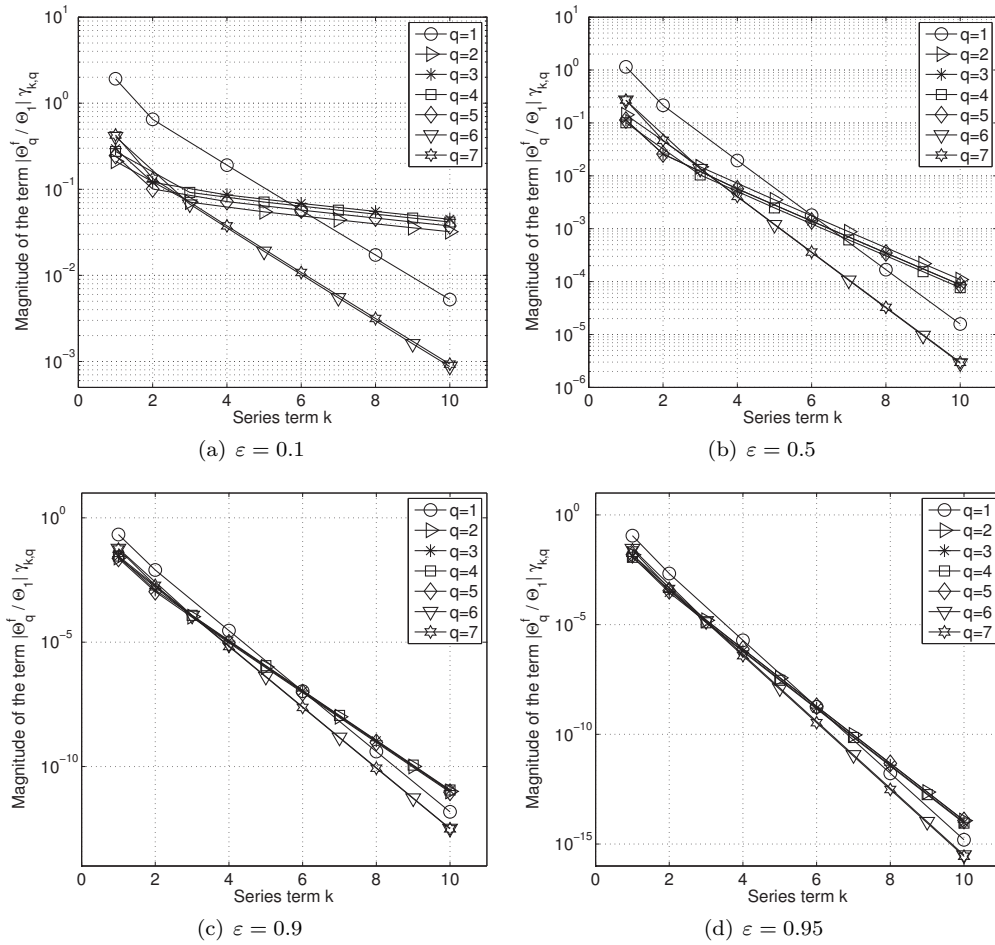
so that  $Q_a = 2$ ,  $Q_f = 7$ , and the problem satisfies the global spectral condition (25) provided that  $\mu_1 \in [-(1 - \varepsilon), 1 - \varepsilon]$  for some  $\varepsilon > 0$ .



**Fig. 2** Schematic description of the domain and boundary conditions of the model problem.

In order to compare the  $n$ -width bounds with the observed convergence rates of the weak greedy algorithm, we considered four different cases:  $\varepsilon = 0.1$ ,  $\varepsilon = 0.5$ ,  $\varepsilon = 0.9$ , and  $\varepsilon = 0.95$ . Note that if  $\varepsilon = 1$ , the manifold of parametric solutions dimension is limited to a  $Q_f$ -dimensional subspace of  $X$  as indicated by (27), and so the greedy algorithm terminates after exactly 7 iterations. In Fig. 3 we have plotted the convergence of the fundamental series terms  $\sup_{\mu} \| (\Theta_2^a(\mu)\Theta_q^f(\mu)/\Theta_1^a(\mu)) \Psi_{k,q} \|_X$  that dictate the convergence rate of the  $n$ -width upper bound (30). For the value  $\varepsilon = 0.1$  very weak convergence of the fundamental series is observed for some of the terms, namely  $q = 2, 3, 4, 5$ .

To obtain the GRBM approximation the weak greedy algorithm was driven by the residual-based *a posteriori* error estimator (20). In both cases the greedy was run until an absolute  $H^1$ -error below  $10^{-3}$  was reached. This required  $n = 25$  basis functions for the case  $\varepsilon = 0.1$ ,  $n = 21$  basis functions for the case  $\varepsilon = 0.5$ ,  $n = 15$  basis functions for the case  $\varepsilon = 0.9$ , and  $n = 14$  basis functions for the case  $\varepsilon = 0.95$ .

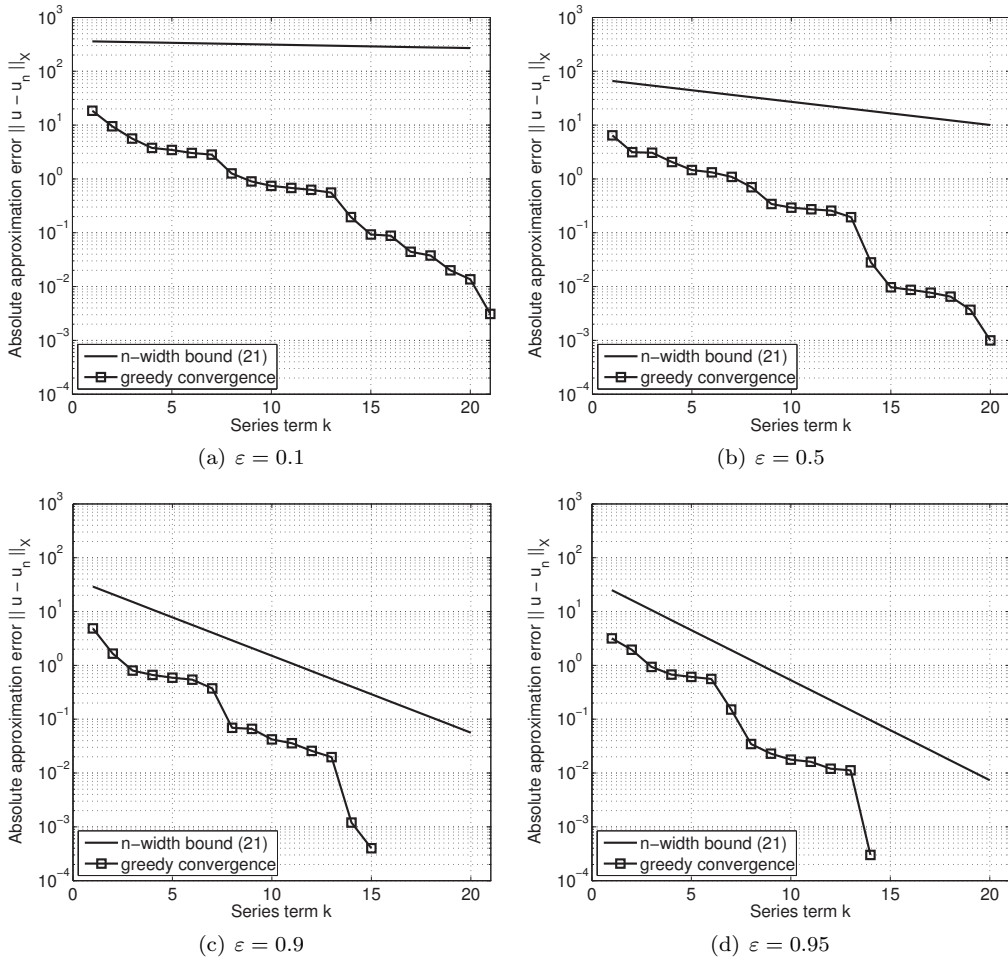


**Fig. 3** Convergence of the fundamental series (26) coefficients for different values of  $\varepsilon$  in (29).

In Fig. 4 we have plotted the corresponding convergence rates of the greedy algorithm compared to the  $n$ -width upper bound predictions given by (30). In each case exponential convergence of the GRBM approximation is observed. The actual exponential decay rate depends on  $\varepsilon$ , where for  $\varepsilon = 0.1$  the  $n$ -width estimate is much too pessimistic when compared to the true rate of convergence. This is likely due to the weak convergence of some of the fundamental series terms (see Fig. 3(a)), and the result could be improved by considering more carefully the cutoff point for the different series terms for different  $q$ . However, as  $\varepsilon \rightarrow 1$  the  $n$ -width estimate (30) becomes more and more indicative of the convergence rate observed during the greedy algorithm. According to Fig. 3(c–d)) at the limit all the fundamental series coefficients converge at roughly the same rate, so that the bound (30) is expected to sharpen considerably.

## 6 Conclusions

We have reviewed the generalized reduced basis method for the uniform approximation of manifolds of solutions of parametric partial differential equations. These methods are typically driven by a greedy algorithm for selecting near-optimal reduced approximation subspaces. It has recently been shown that the convergence rate of the generalized reduced basis approximations is linked to the Kolmogorov  $n$ -width of the manifold of solutions.



**Fig. 4** Comparison of the  $n$ -width upper bound estimate (30) and the greedy algorithm convergence rate.

We have exhibited a model problem where the exact parameter-dependent solution can be expanded as a Neumann series, leading to a constructive proof that the  $n$ -width of the solution set in this case converges exponentially. Numerical experiments confirm that the reduced basis approximation also converges exponentially, and with a rate that is comparable to the one predicted by our  $n$ -width upper bound estimate. The predicted convergence rate is independent of the parametric regularity of the solution manifold and the number of parameters, but it does depend on the size of the affine expansion of the parametric problem.

Future work involves finding more cases of parameter-dependent problems, where explicit solution formulas could be used to prove more general  $n$ -width estimates.

## References

1. G. Adomian. *Solving Frontier Problems of Physics: The Decomposition Method*. Kluwer Academic Publishers, 1994.
2. I. Babuška, B.A. Szabo, and I.N. Katz. The  $p$ -version of the finite element method. *SIAM J. Numer. Anal.*, 18(3):515–545, 1981.
3. P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM J. Math. Anal.*, 43(3):1457–1472, 2011.
4. A. Buffa, Y. Maday, A. T. Patera, C. Prud’homme, and G. Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM Math. Model. Numer. Anal.*, 46(3):595–603, 2012.



5. C. Canuto, T. Tonn, and K. Urban. A-posteriori error analysis of the reduced basis method for non-affine parameterized nonlinear PDEs. *SIAM J. Numer. Anal.*, 47(3):2001–2022, 2009.
6. K. Carlberg and C. Farhat. A low-cost, goal-oriented ‘compact proper orthogonal decomposition’ basis for model reduction of static systems. *Int. J. Numer. Methods Engrg.*, 86(3):381–402, 2011.
7. A. Cohen, R. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs. *Preprint*, 2010.
8. S. Deparis. Reduced basis error bound computation of parameter-dependent Navier-Stokes equations by the natural norm approach. *SIAM J. Num. Anal.*, 46(4):2039–2067, 2008.
9. J.L. Eftang, D.J. Knezevic, and A.T. Patera. An  $hp$  certified reduced basis method for parametrized parabolic partial differential equations. *Math. Comput. Modelling Dynam. Systems*, 17(4):395–422, 2011.
10. J.A. Evans, Y. Bazilevs, I. Babuška, and T.J.R. Hughes.  $N$ -widths, sup-infs, and optimality ratios for the  $k$ -version of the isogeometric finite element method. *Comput. Methods Appl. Mech. Engrg.*, 198(21-26):1726–1741, 2009.
11. M.A. Grepl, Y. Maday, N.C. Nguyen, and A.T. Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM Math. Modelling Numer. Anal.*, 41(3):575–605, 2007.
12. A. Hay, J. Borggaard, I. Akhtar, and D. Pelletier. Reduced-order models for parameter dependent geometries based on shape sensitivity analysis. *J. Comp. Phys.*, 229(4):1327–1352, 2010.
13. A. Hay, J.T. Borggaard, and D. Pelletier. Local improvements to reduced-order models using sensitivity analysis of the proper orthogonal decomposition. *J. Fluid Mech.*, 629:41–72, 2009.
14. D.B.P. Huynh, D. Knezevic, Y. Chen, J. Hesthaven, and A.T. Patera. A natural-norm successive constraint method for inf-sup lower bounds. *Comput. Methods Appl. Mech. Engrg.*, 199(29–32), 2010.
15. D.B.P. Huynh, G. Rozza, S. Sen, and A.T. Patera. A successive constraint linear optimization method for lower bounds of parametric coercivity and inf-sup stability constants. *C. R. Acad. Sci. Paris. Sér. I Math.*, 345:473–478, 2007.
16. K. Ito and S.S. Ravindran. A reduced order method for simulation and control of fluid flows. *J. Comp. Phys.*, 143(2), 1998.
17. T. Lassila, A. Manzoni, and G. Rozza. On the approximation of stability factors for general parametrized partial differential equations with a two-level affine decomposition. *ESAIM Math. Model. Numer. Anal.*, 46:1555–1576, 2012.
18. Y. Maday. *Reduced basis method for the rapid and reliable solution of partial differential equations*, volume III, pages 1255–1270. Eur. Math. Soc., Zürich, 2006.
19. Y. Maday, A.T. Patera, and G. Turinici. Global a priori convergence theory for reduced-basis approximation of single-parameter symmetric coercive elliptic partial differential equations. *C. R. Acad. Sci. Paris. Sér. I Math.*, 335:289–294, 2002.
20. Y. Maday, A.T. Patera, and G. Turinici. A priori convergence theory for reduced-basis approximations of single-parametric elliptic partial differential equations. *J. Sci. Comput.*, 17(1–4):437–446, 2002.
21. A. Manzoni. *Reduced models for optimal control, shape optimization and inverse problems in haemodynamics*. PhD thesis, N. 5402, École Polytechnique Fédérale de Lausanne, 2012.
22. J.M. Melenk. On  $n$ -widths for elliptic problems. *J. Math. Anal. Appl.*, 247:272–289, 2000.
23. A.A. Melkman and C.A. Micchelli. Spline spaces are optimal for  $L^2$   $n$ -width. *Illinois J. Math.*, 22(4), 1978.
24. A.K. Noor. Recent advances in reduction methods for nonlinear problems. *Comput. Struct.*, 13(1-3):31–44, 1981.
25. J.S. Peterson. The reduced basis method for incompressible viscous flow calculations. *SIAM J. Sci. Stat. Comput.*, 10:777–786, 1989.
26. A. Pinkus.  *$n$ -Widths in Approximation Theory*. Springer-Verlag, Ergebnisse, 1985.
27. A. Quarteroni, G. Rozza, and A. Manzoni. Certified reduced basis approximation for parametrized PDEs and applications. *J. Math. Ind.*, 3(1), 2011.
28. G. Rozza, D.B.P. Huynh, and A.T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Arch. Comput. Methods Engrg.*, 15:229–275, 2008.
29. T. Tonn, K. Urban, and S. Volkwein. Comparison of the reduced-basis and POD a posteriori error estimators for an elliptic linear-quadratic optimal control problem. *Math. Comp. Model. Dyn.*, 17(4):355–369, 2011.
30. K. Veroy and A.T. Patera. Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: rigorous reduced-basis a posteriori error bounds. *Int. J. Numer. Meth. Fluids*, 47(8-9):773–788, 2005.