Mathematical Modelling and Numerical Analysis Will be set by the publisher
Modélisation Mathématique et Analyse Numérique

A PRIORI CONVERGENCE OF THE GREEDY ALGORITHM FOR THE PARAMETRIZED REDUCED BASIS METHOD

A. Buffa[1], Y. Maday[2], A.T. Patera[3], C. Prud'homme[4] and G. Turinici[5]

Abstract. The convergence and efficiency of the reduced basis method used for the approximation of the solutions to a class of problems written as a parametrized PDE depends heavily on the choice of the elements that constitute the "reduced basis". The purpose of this paper is to analyze the *a priori* convergence for one of the approaches used for the selection of these elements, the greedy algorithm. Under natural hypothesis on the set of all solutions to the problem obtained when the parameter varies, we prove that three greedy algorithms converge; the last algorithm, based on the use of an *a posteriori* estimator, is the approach actually employed in the calculations.

Résumé. La convergence et l'efficacité de la méthode des bases réduites pour l'approximation de la solution d'une classe de problèmes écrits sous la forme d'une EDP paramétrique dépendent fortement du choix des éléments qui constituent la "base réduite". Ce travail est une contribution à l'analyse de la convergence *a priori* de l'algorithme glouton qui est l'une des approches utilisées pour la sélection de ces éléments. On démontre, sous des hypothèses naturelles sur l'ensemble des solutions du problème lorsque le paramètre varie, que trois algorithmes gloutons convergent, le dernier, basé sur l'utilisation d'un estimateur *a posteriori*, étant celui effectivement mis en oeuvre dans les calculs.

1991 Mathematics Subject Classification. 41A45, 41A65, 65N15.

September 19, 2010.

Introduction

The reduced basis method is a discretization approach for the approximation of the solutions of parameter dependent partial differential equations. Some solutions are assumed to be known (or at least very well approximated by a classical discretization method) for certain, well chosen, parameters from a preliminary (offline)

Keywords and phrases: greedy algorithm, reduced basis approximations, a priori analysis, best fit analysis

[1] Istituto di Matematica Applicata e Tecnologie Informatiche - CNR, Via Ferrata 1, 27100 Pavia, Italy
[2] UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France and Division of Applied Mathematics, Brown University, Providence, RI, USA
[3] Massachusetts Institute of Technology, Department of Mechanical Engineering, Room 3-266, 77 Mass. Ave., Cambridge, MA 02139-4307 USA
[4] Université de Grenoble 1-Joseph Fourier, Laboratoire Jean Kuntzmann, 51 rue des Mathmatiques, BP53, 38041, Grenoble Cedex 9, France
[5] Université Paris Dauphine, CEREMADE, Place du Marechal de Lattre de Tassigny, 75016 Paris, France

step; these solutions constitute the basis of the reduced basis method. The solution for (a large number of) new parameters is then approximated as a linear combination of the elements of this basis. Most often, this approximation is based on the variational equivalent formulation of the problem, the reduced basis approximation then being defined through a Galerkin process. In previous works [3, 4] exponential convergence with respect to the number N of basis functions is proved for a one-dimensional parameter case, and numerical experiments [7–9] illustrate the same behavior (or even faster) even in situations where the dimension of the parameter space P is larger. This is the case only when the elements of the basis — i.e. the parameters in the offline process — are sufficiently well chosen. The offline selection of these parameters is critical and various methods have been proposed for this purpose. These methods differ in their essence, in their efficiency both in the offline stage and in the online stage, and in whether they rely on random arguments or deterministic frameworks such as principal component analysis or greedy algorithms.

The aim of this paper is to provide an analysis of the greedy algorithm that is very commonly used in practice. Note that the concept of reduced basis approximation implies some structure on the set of all solutions of the parameter dependent partial differential equation under consideration. There is no reason why a reduced basis approach should be a viable alternative to classical discretizations such as finite element, finite volume or spectral methods in the most general case where the solutions do not depend smoothly with respect to the parameter. We thus start by making precise the feature that the set of all solutions must satisfy.

Let us first introduce the notations: $u(x, \boldsymbol{\mu})$ is the solution of a parameter dependent partial differential equation (PDE) set on a bounded spatial domain $\Omega \subset \mathbb{R}^d$ and on a closed parametric domain $D \subset \mathbb{R}^P$. For each $\boldsymbol{\mu}$ the solution $u(\cdot, \boldsymbol{\mu})$ belongs to $\mathcal{X} \subset L^2(\Omega)$, a functional space adapted to the PDE, e.g. $\mathcal{X} = H_0^1(\Omega)$ or $\mathcal{X} = L^2(\Omega)$. We will assume D to be compact, but we make no further hypothesis on Ω other than those required by the PDE itself.

The weak form of our partial differential equation reads: given $\boldsymbol{\mu} \in D$, find $u(\boldsymbol{\mu}) \in \mathcal{X}$ which satisfies

$$\mathcal{A}(u(\boldsymbol{\mu}), v; \boldsymbol{\mu}) = g(v), \quad \forall v \in \mathcal{X} \ , \tag{1}$$

where the form $\mathcal{A}(\cdot, \cdot; \boldsymbol{\mu}) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ encodes the description of the PDE and g is an element of $\mathcal{X}'$. We assume that the bilinear form $\mathcal{A}(\cdot, \cdot; \boldsymbol{\mu})$ is continuous and coercive on $\mathcal{X}$, uniformly with respect to the parameters $\boldsymbol{\mu}$:

there exists two positive constants M and $\alpha_{\rm coer}$ (independent of the parameters $\boldsymbol{\mu}$) such that

$$\begin{aligned}
&\forall \boldsymbol{\mu} \in D, \quad \mathcal{A}(u,v;\boldsymbol{\mu}) \leq M\|u\|_{\mathcal{X}}\|v\|_{\mathcal{X}} \qquad \forall\, u\,,\ v \in \mathcal{X};\\
&\forall \boldsymbol{\mu} \in D, \quad \mathcal{A}(u,u;\boldsymbol{\mu}) \geq \alpha_{\rm coer}\|u\|^2_{\mathcal{X}} \qquad \forall\, u \in \mathcal{X}.
\end{aligned} \tag{2}$$

For simplicity, we shall also assume that $\mathcal{A}$ is symmetric, $\mathcal{A}(u,v;\boldsymbol{\mu}) = \mathcal{A}(v,u;\boldsymbol{\mu})$, $\forall\, u,v \in \mathcal{X}$, although this hypothesis is not central to the results of the paper.

The reduced basis method consists in approximating the solution $u(\boldsymbol{\mu})$ of the parameter dependent problem (1) by a linear combination of appropriate, pre-computed, solutions $u(\boldsymbol{\mu}_i)$ for well chosen parameters $\boldsymbol{\mu}_i$, $i = 1,\ldots,N$. The approximation method of choice is a Galerkin procedure that reads: given $\boldsymbol{\mu} \in D$, find $u_N(\boldsymbol{\mu}) \in X_N = \text{Span}\{u(\boldsymbol{\mu}_i), i = 1,..,N\}$ such that

$$\mathcal{A}(u_N(\boldsymbol{\mu}), v_N;\boldsymbol{\mu}) = g(v_N), \quad \forall v_N \in X_N\ . \tag{3}$$

Cea's lemma provides the following bound

$$\|u(\boldsymbol{\mu}) - u_N(\boldsymbol{\mu})\|_{\mathcal{X}} \leq c \inf_{v_N \in X_N} \|u(\boldsymbol{\mu}) - v_N\|_{\mathcal{X}}\ , \tag{4}$$

where in fact $c = \sqrt{M/\alpha_{\rm coer}}$.

The rationale for this approach relies on the fact that the right-hand side of the bound (4) is very small, at least in many cases of importance. This, in turns, follows from the fact that the set $\mathcal{S}(D) = \{u(\boldsymbol{\mu})$ of all solutions to (1) when $\boldsymbol{\mu} \in D\}$ behaves well. In order to comprehend in which sense the good behavior of $\mathcal{S}(D)$ should be understood, it is helpful to introduce the notion of n-width following Kolmogorov [2] (see also [6])

Definition 1. Let F be a subset of $\mathcal{X}$ and Y_n be a generic n-dimensional subspace of $\mathcal{X}$. The angle between F and Y_n is

$$E(F;Y_n) := \sup_{x\in F} \inf_{y \in Y_n} \|x-y\|_{\mathcal{X}}.$$

The *Kolmogorov n-width* of F in $\mathcal{X}$ is given by

$$\begin{aligned}
d_n(F,\mathcal{X}) &:= \inf\{E(F;Y_n) : Y_n \text{ a } n\text{-dimensional subspace of } \mathcal{X}\}\\
&= \inf_{Y_n} \sup_{x\in F} \inf_{y\in Y_n} \|x-y\|_{\mathcal{X}}\ .
\end{aligned} \tag{5}$$

The n-width of F thus measures the extent to which F may be approximated by an n-dimensional subspace of $\mathcal{X}$. These concepts have been used to analyze the effectiveness of hp-finite elements in [5]. There are many reasons why this n-width may go to zero rapidly as n goes to infinity. In our case, where $F = \mathcal{S}(D)$, we can refer to regularity of the solutions $u(\boldsymbol{\mu})$ with respect to the parameter $\boldsymbol{\mu}$, or even to analyticity. Indeed, an upper bound for the asymptotic rate at which the n-width tends to zero is provided by the example from Kolmogorov stating that $d_n(\widetilde{B}_2^{(r)}; L^2) = \mathcal{O}(n^{-r})$ where $\widetilde{B}_2^{(r)}$ is the unit ball in the Sobolev space of all 2π-periodic, real valued, $(r-1)$-times differentiable functions whose $(r-1)$st derivative is absolutely continuous and whose rth derivative belongs to $L^2(\mathbb{R})$. In fact, exponential convergence is achieved when analyticity exists in the parameter dependency.

The knowledge of the n-width of F is not sufficient: of theoretical interest is the determination of an optimal finite dimensional space Y_n that realizes the infimum in d_n (provided it exists) or that is "close enough" to d_n. For practical reasons, we shall restrict ourselves to finite dimensional spaces that are spanned by elements of $\mathcal{S}(D)$. The greedy algorithms, a first definition of which is presented below, permit to construct such a space with good approximation properties.

Let us assume that the subset F in $\mathcal{X}$ is compact (consistent with the fact that D is assumed to be compact). In the general setting, the greedy algorithm is defined as follows:

- $f_1 := \operatorname{argmax} \|f\|_{\mathcal{X}}$
- Assume $f_1, \ldots, f_{i-1}$ are defined, consider $F_{i-1} := \operatorname{Span}\{f_1, \ldots, f_{i-1}\}$
- $f_i := \operatorname{argmax} \|f - P_{F_{i-1}}(f)\|_{\mathcal{X}}$

where $P_{F_{i-1}}$ denotes the orthogonal projection on F_{i-1} for the scalar product in $\mathcal{X}$.

1. Analysis of the approximation properties of F_k.

Assume that the construction of f_i does not end (which is equivalent to the fact that $\operatorname{Span}(F)$ is an infinite dimensional space). We start by orthogonalizing the elements provided by the algorithm, hence define

- $\xi_1 = f_1$
- $\xi_i = f_i - P_{F_{i-1}}(f_i)$.

It is an easy matter to check that the expression of $P_{F_i}(f)$, for any $f \in X$, is facilitated in this basis; indeed

$$\forall f \in X, \quad P_{F_i}(f) = \sum_{\ell=1}^{i} \alpha_\ell(f)\xi_\ell \tag{6}$$

with

$$\alpha_\ell(f) = \frac{\langle f, \xi_\ell \rangle_{\mathcal{X}}}{\|\xi_\ell\|^2_{\mathcal{X}}} \ . \tag{7}$$

Due to the orthogonality of ξ_ℓ with $F_{\ell-1}$, we deduce that

$$\begin{aligned} |\alpha_\ell(f)| &= \frac{|\langle f - P_{F_{\ell-1}} f, \xi_\ell \rangle_{\mathcal{X}}|}{\|\xi_\ell\|^2_{\mathcal{X}}} \\ &\leq \frac{\|f - P_{F_{\ell-1}} f\|_{\mathcal{X}}}{\|\xi_\ell\|_{\mathcal{X}}} = \frac{\|f - P_{F_{\ell-1}} f\|_{\mathcal{X}}}{\|f_\ell - P_{F_{\ell-1}} f_\ell\|_{\mathcal{X}}}, \end{aligned}$$

and hence from the maximization definition of f_ℓ we conclude that

$$\forall f \in F, \quad |\alpha_\ell(f)| \leq 1. \tag{8}$$

In what follows we denote by $\alpha_\ell^j := \alpha_\ell(f_j)$.

With this notation, we can write

$$\begin{aligned} \xi_2 &= f_2 - \alpha_1^2 f_1 \\ \xi_3 &= f_3 - \alpha_1^3 f_1 - \alpha_2^3 (f_2 - \alpha_1^2 f_1) \\ \xi_4 &= f_4 - \alpha_1^4 f_1 - \alpha_2^4 (f_2 - \alpha_1^2 f_1) - \alpha_3^4 (f_3 - \alpha_1^3 f_1 - \alpha_2^3 (f_2 - \alpha_1^2 f_1)) \\ \xi_5 &= \dots \end{aligned}$$

and thus

$$\xi_j = \sum_{\ell=1}^{j} \beta_\ell^j f_\ell \tag{9}$$

with

$$\begin{aligned} \beta_j^j &= 1 \\ \beta_\ell^j &= -\sum_{i=\ell}^{j-1} \alpha_i^j \beta_\ell^i. \end{aligned}$$

This, combined with (8), allows us to derive by induction that, for $j \geq \ell$,

$$\beta_\ell^j \leq 2^{j-\ell} \ . \tag{10}$$

Let now k be given. From the definition of the Kolmogorov n-width we know that, for any given $\lambda > 1$, there exists a finite dimensional space Y_k such that $E(F; Y_k) \leq \lambda d_k(F; \mathcal{X})$. This means that for any $\ell \leq k$, there exists a $v_\ell \in Y_k$ such that

$$\|f_\ell - v_\ell\|_{\mathcal{X}} \leq \lambda d_k(F; \mathcal{X}) \ . \tag{11}$$

Let us now set

$$\zeta_j = \sum_{\ell=1}^{j} \beta_\ell^j v_\ell, \tag{12}$$

which are elements in Y_k; these elements satisfy

$$\|\xi_\ell - \zeta_\ell\|_{\mathcal{X}} \leq 2^\ell \lambda d_k(F; \mathcal{X}) \ . \tag{13}$$

Let us now consider the family ζ_i for $i = 1, \ldots, k+1$. Since these $k+1$ vector belong to Y_k, which is k-dimensional, we deduce that there exist γ_i, $\|\boldsymbol{\gamma}\|_{\ell^2} = 1$, such that $\sum_{i=1}^{k+1} \gamma_i \zeta_i = 0$. We then know that

$$\left\| \sum_{i=1}^{k+1} \gamma_i \xi_i \right\|_{\mathcal{X}} = \left\| \sum_{i=1}^{k+1} \gamma_i (\xi_i - \zeta_i) \right\|_{\mathcal{X}} \leq 2^{k+1} \sqrt{k+1} \lambda d_k(F; \mathcal{X}) \ . \tag{14}$$

We know that there exists a j such that $\gamma_j > 1/\sqrt{k+1}$. Thus,

$$\left\| \xi_j + \gamma_j^{-1} \sum_{i<j} \gamma_i \xi_i + \gamma_j^{-1} \sum_{i>j} \gamma_i \xi_i \right\|_{\mathcal{X}} \leq 2^{k+1} (k+1) \lambda d_k(F; \mathcal{X}) \ .$$

Now, since the functions ξ_i are orthogonal, we obtain

$$\|\xi_j\|_{\mathcal{X}} \leq 2^{k+1} (k+1) \lambda d_k(F; \mathcal{X}) \ .$$

Recalling the very definition of ξ_j, we have that, for all $f \in F$,

$$\|f - P_{F_{j-1}} f\|_{\mathcal{X}} \leq \|f_j - P_{F_{j-1}} f_j\|_{\mathcal{X}} = \|\xi_j\|_{\mathcal{X}} \leq 2^{k+1} (k+1) \lambda d_k(F; \mathcal{X}) \ .$$

Hence, for any given $\lambda > 1$

$$\|f - P_{F_k} f\|_{\mathcal{X}} \leq \|f - P_{F_{j-1}} f\|_{\mathcal{X}} \leq 2^{k+1} (k+1) \lambda d_k(F; \mathcal{X}) \ .$$

We have thus proven

Theorem 1. *Assume that the set F has an exponentially small Kolmogorov n-width $d_k(F;\mathcal{X}) \leq ce^{-\alpha k}$ with $\alpha > \log 2$, then there exists $\beta > 0$ such that the set F_k yielded by the greedy algorithm has exponential approximation properties in the sense that*

$$\|f - P_{F_k} f\|_{\mathcal{X}} \leq Ce^{-\beta k} \ .$$

Remark 1. *It is instructive to exhibit examples that prove that the loss of the factor 2^n between the best choice indicated by the Kolmogorov n-width and the choice resulting from the greedy algorithm can be realized. Indeed, we have the following statement : For any $n > 0$, there exists a set $E_{n+1} = \{v_1, v_2, ..., v_{n+1}\}$ of vectors in $\mathbb{R}^{n+1}$ such that*

- *Regarding the choice of the greedy algorithm : for any k, $1 \leq k \leq n$*

$$F_k = \mathrm{Span}\{v_1, v_2, ..., v_k\},$$

- *Regarding the approximation properties*

$$\|v_{n+1} - P_{F_n} v_{n+1}\|_{\mathbb{R}^{n+1}} \simeq 2^n d_n(E_{n+1}, \mathbb{R}^{n+1}).$$

An example of such a set is as follows : let $e_1, e_2, .., e_{n+1}$ be the canonical basis of $\mathbb{R}^{n+1}$, $\varepsilon > 0$ be small enough, and $0 \leq \delta_1 \leq \delta_2 \ll \varepsilon^n$ $(\delta_2 > 0)$ then the above statement holds for the choice

$$v_1 = (1 + n\varepsilon^2)e_1 + \delta_1 e_{n+1}$$

$$v_2 = (1 + (n-1)\varepsilon^2)(e_1 + \varepsilon e_2) - \delta_1 e_{n+1}$$

$$v_3 = (1 + (n-2)\varepsilon^2)(e_1 - \varepsilon e_2 + \varepsilon^2 e_3) - \delta_1 e_{n+1}$$

$$v_4 = (1 + (n-3)\varepsilon^2)(e_1 - \varepsilon e_2 - \varepsilon^2 e_3 + \varepsilon^3 e_4) - \delta_1 e_{n+1}$$

$$..........$$

$$v_n = (1 + \varepsilon^2)(e_1 - \varepsilon e_2 - \varepsilon^2 e_3 - .... - \varepsilon^{n-2} e_{n-1} + \varepsilon^{n-1} e_n) - \delta_1 e_{n+1}$$

$$v_{n+1} = (e_1 - \varepsilon e_2 - \varepsilon^2 e_3 - .... - \varepsilon^{n-1} e_n) - \delta_2 e_{n+1}$$

First, it is obvious that $d_n(E_{n+1}, \mathbb{R}^{n+1}) \leq \mathcal{O}\delta_2$, since δ_2 is the angle between E_{n+1} and $\mathrm{Span}\{e_1, e_2, ..., e_n\}$.

Second, the prefactor $(1 + k\varepsilon^2)$ *is responsible for the order in which the greedy algorithm selects the elements and explains the first item above. This fact is obvious in the case when* $\delta_1 = 0$ *and remains true by continuity for* $\delta_1 > 0$*, small enough; indeed, if* $\delta_1 = 0$*, the norm of* v_1 *is equal to* $1 + n\varepsilon^2$ *and the norm of* v_2 *is, provided* ε *is small enough, of the order of* $1 + (n - 1/2)\varepsilon^2$*. This proves in particular that*

$$F_n = \mathrm{Span}\{v_1, v_2, ..., v_n\}. \tag{15}$$

Lastly, in order to understand the second item, it suffices again to analyze first the case where $\delta_1 = 0$*. Indeed, in this situation, due to (15), we can demonstrate that the best approximation of* v_{n+1} *in* F_n *is realized by* $P_{F_n} v_{n+1} = 2^{n-1}\frac{v_1}{1+n\varepsilon^2} - 2^{n-2}\frac{v_2}{1+(n-1)\varepsilon^2} \dots - 2\frac{v_{n-1}}{1+2\varepsilon^2} - \frac{v_n}{1+\varepsilon^2}$*. This remains the case, with slight modifications of the coefficients, so that even if* $\delta_1 \leq \delta_2 \ll \varepsilon^n$

$$v_{n+1} - P_{F_n} v_{n+1} \simeq 2^n \delta_2 e_{n+1},$$

this concludes the statement.

2. The greedy algorithm for the reduced basis method

Let us focus here on the case where F is the set of all solutions $\mathcal{S}(D) = \{u(\boldsymbol{\mu}), \boldsymbol{\mu} \in D\}$ to (1). (In actual practice, remember that we consider $S_h(D) = \{u_h(\boldsymbol{\mu}), \boldsymbol{\mu} \in D\}$, where $X_h \subset X$ is a suitably fine finite element approximation.) The greedy selection of the parameters varies slightly due to the natural variational framework of the problem:

Algorithm 1

i: The first parameter is defined as previously

$$\boldsymbol{\mu}_1 = \arg \sup_{\boldsymbol{\mu} \in D} \|u(\boldsymbol{\mu}; \cdot)\|_{\mathcal{X}} \ .$$

(Again, in actual practice, u is replaced by u_h.)

ii: Given $i-1$ samples in the parameters set, $\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_{i-1}$, we construct $U_{i-1} = \mathrm{Span}\{u(\boldsymbol{\mu}_1; \cdot), \dots, u(\boldsymbol{\mu}_{i-1}; \cdot)\}$, and we denote by $\Pi^{\boldsymbol{\mu}}_{i-1} : \mathcal{X} \to U_{i-1}$ the elliptic (Galerkin) projection onto the space U_{i-1} :

$$\mathcal{A}(\Pi^{\boldsymbol{\mu}}_{i-1} u, v; \boldsymbol{\mu}) = \mathcal{A}(u, v; \boldsymbol{\mu}), \quad \forall\, v \in U_{i-1} \ .$$

The next parameters are defined as follows

$$\boldsymbol{\mu}_i = \arg\sup_{\boldsymbol{\mu}\in D} \|u(\boldsymbol{\mu};\cdot) - \Pi^{\boldsymbol{\mu}}_{i-1} u(\boldsymbol{\mu};\cdot)\|_{\mathcal{X}} \ ,$$

iii: Iterate until $\arg\sup_{\boldsymbol{\mu}\in D} \|u(\boldsymbol{\mu};\,\cdot\,) - \Pi^{\boldsymbol{\mu}}_n u(\boldsymbol{\mu};\,\cdot\,)\|_{\mathcal{X}} < \text{tol}$.

Note that, from (1) and (3) we have $\forall \boldsymbol{\mu} \in D$ and $\forall m$, $\Pi^{\boldsymbol{\mu}}_m u(\boldsymbol{\mu}) = u_m(\boldsymbol{\mu})$.

The basis so generated is now orthogonalized with respect to the $\mathcal{X}$ scalar product, and we denote by $\{\xi_1, \dots, \xi_n\}$ the resulting basis.

Note that in the orthogonalization process, we cannot use $\Pi^{\boldsymbol{\mu}}_n$ since this operator depends on $\boldsymbol{\mu}$: this is why, in what follows, the orthogonalization is performed through the $\mathcal{X}$-topology. We denote by $P_{U_i} : \mathcal{X} \to U_i$ the orthogonal projection with respect to the $\mathcal{X}$ topology (which thus differs from $\Pi^{\boldsymbol{\mu}}_i$). The orthogonalization process gives:

$$\xi_1 = u(\boldsymbol{\mu}_1;\,\cdot\,), \quad \xi_i = u(\boldsymbol{\mu}_i;\,\cdot\,) - P_{U_{i-1}} u(\boldsymbol{\mu}_i;\,\cdot\,), \quad i = 2, \dots, n \ . \tag{16}$$

In particular $P_{U_i} u(\boldsymbol{\mu};\,\cdot\,) = \sum_{\ell=1}^{i} \alpha_\ell(\boldsymbol{\mu})\xi_\ell$, with

$$\alpha_\ell(\boldsymbol{\mu}) = \frac{\langle u(\boldsymbol{\mu};\,\cdot\,), \xi_\ell\rangle_{\mathcal{X}}}{\|\xi_\ell\|^2_{\mathcal{X}}} \ ,$$

where we recall that $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathcal{X}}$ denotes the scalar product in $\mathcal{X}$. We then have

$$|\alpha_\ell(\boldsymbol{\mu})| = \frac{|\langle(u(\boldsymbol{\mu};\,\cdot\,) - \Pi^{\boldsymbol{\mu}}_{\ell-1} u(\boldsymbol{\mu};\,\cdot\,), \xi_\ell)\rangle_{\mathcal{X}}|}{\|\xi_\ell\|^2_{\mathcal{X}}}$$

because of the orthogonality of ξ_ℓ and $\xi_1, \dots, \xi_{\ell-1}$. We thus obtain

$$\begin{aligned} |\alpha_\ell(\boldsymbol{\mu})| &\le \frac{\|u(\boldsymbol{\mu};\,\cdot\,) - \Pi^{\boldsymbol{\mu}}_{\ell-1} u(\boldsymbol{\mu};\,\cdot\,)\|_{\mathcal{X}}}{\|u(\boldsymbol{\mu}_\ell;\,\cdot\,) - P_{U_{\ell-1}} u(\boldsymbol{\mu}_\ell;\,\cdot\,)\|_{\mathcal{X}}} \\ &\le \frac{\|u(\boldsymbol{\mu}_\ell;\,\cdot\,) - \Pi^{\boldsymbol{\mu}_\ell}_{\ell-1} u(\boldsymbol{\mu}_\ell;\,\cdot\,)\|_{\mathcal{X}}}{\|u(\boldsymbol{\mu}_\ell;\,\cdot\,) - P_{U_{\ell-1}} u(\boldsymbol{\mu}_\ell;\,\cdot\,)\|_{\mathcal{X}}} \end{aligned} \tag{17}$$

since $\boldsymbol{\mu}_\ell$ is the parameter value in D attaining the maximum. Finally, we conclude that

$$|\alpha_\ell(\boldsymbol{\mu})| \le \sqrt{\frac{M}{\alpha_{\text{coer}}}}$$

thanks to the Galerkin type estimate (4)

$$\|u(\boldsymbol{\mu}_\ell;\,\cdot\,) - \Pi^{\boldsymbol{\mu}}_{\ell-1} u(\boldsymbol{\mu}_\ell;\,\cdot\,)\|_{\mathcal{X}} \le \sqrt{\frac{M}{\alpha_{\text{coer}}}} \|u(\boldsymbol{\mu}_\ell;\,\cdot\,) - P_{U_{\ell-1}} u(\boldsymbol{\mu}_\ell;\,\cdot\,)\|_{\mathcal{X}} \ .$$

The convergence analysis from the above estimate compared to (8) leads to a deteriorated bound

$$\beta_\ell^j \le \left(1+\sqrt{\frac{M}{\alpha_{\rm coer}}}\right)^{j-\ell} . \qquad (18)$$

instead of (10) and the conclusion is then given in

Theorem 2. *Assume that the set of all solutions $\mathcal{S}(D) = \{u(\boldsymbol{\mu}), \boldsymbol{\mu} \in D\}$ to (1) has an exponentially small Kolmogorov n-width $d_k(\mathcal{S}(D), \mathcal{X}) \le ce^{-\alpha k}$ with $\alpha > \log\left(1+\sqrt{\frac{M}{\alpha_{\rm coer}}}\right)$; then the reduced basis method converges exponentially in the sense that there exists $\beta > 0$ such that*

$$\forall\, \boldsymbol{\mu} \in D, \quad \|u(\mu) - u_N(\mu)\|_{\mathcal{X}} \le Ce^{-\beta N} . \qquad (19)$$

3. A computable greedy algorithm via *a posteriori* error bounds

In practice, the optimization of Step **ii** of Algorithm 1 is very computationally intensive. In practice we first replace the sup over D with a sup over a very fine sample in D; this nevertheless still requires many expensive evaluations. In order to construct a computable algorithm, we need in addition to replace Step **ii** with a relatively inexpensive procedure that maintains the performance stated in the estimate (19). We thus replace Step **ii** with

ii′:

$$\boldsymbol{\mu}_i = \arg\sup_{\boldsymbol{\mu}\in D} \Delta_{i-1}(\boldsymbol{\mu})$$

where $\Delta_{i-1}(\boldsymbol{\mu})$ is an inexpensive *a posteriori* error estimator of the quantity $\arg\sup_{\boldsymbol{\mu}\in D} \|u(\boldsymbol{\mu};\cdot) - \Pi_{i-1}^{\boldsymbol{\mu}} u(\boldsymbol{\mu};\cdot)\|_{\mathcal{X}}$. We briefly introduce such an estimator and refer to [7] for further details. To begin, we define the residual

$$r_i(v;\boldsymbol{\mu}) = f(v) - \mathcal{A}(\Pi_i^{\boldsymbol{\mu}} u(\boldsymbol{\mu}), v; \boldsymbol{\mu}), \quad \forall\, v \in X ,$$

associated with equation (1). Then $\Delta_i(\boldsymbol{\mu})$ is defined by

$$\Delta_i(\boldsymbol{\mu}) = \frac{\|r_i(\,\cdot\,,\boldsymbol{\mu})\|_{X'}}{\alpha_{\rm coer}^{\rm LB}} ,$$

where $\alpha_{\rm coer}^{\rm LB}$ is a positive lower bound for the coercitivity constant $\alpha_{\rm coer}$ introduced in (2). We can then demonstrate that

$$\|u(\boldsymbol{\mu}) - \Pi_i^{\boldsymbol{\mu}}\, u(\boldsymbol{\mu})\|_{\mathcal{X}} \le \Delta_i(\boldsymbol{\mu}) \le \frac{M}{\alpha_{\rm coer}^{\rm LB}} \|u(\boldsymbol{\mu}) - \Pi_i^{\boldsymbol{\mu}}\, u(\boldsymbol{\mu})\|_{\mathcal{X}},$$

which proves that $\Delta_i(\boldsymbol{\mu})$ is a valid error estimator. Note that this result is valid for every i and $\boldsymbol{\mu} \in D$, and does not require any hypotheses about regularity or *a priori* convergence.

It readily follows that

$$
\begin{aligned}
\|u(\boldsymbol{\mu};\cdot) - \Pi^{\boldsymbol{\mu}}_{\ell-1}\, u(\boldsymbol{\mu};\cdot)\|_{\mathcal{X}} \leq \quad & \Delta_{\ell-1}(\boldsymbol{\mu}) \\
\leq \quad & \Delta_{\ell-1}(\boldsymbol{\mu}_\ell) \\
\leq \quad & \frac{M}{\alpha^{\mathrm{LB}}_{\mathrm{coer}}} \|u(\boldsymbol{\mu}_\ell;\cdot) - \Pi^{\boldsymbol{\mu}_\ell}_{\ell-1}\, u(\boldsymbol{\mu}_\ell;\cdot)\|_{\mathcal{X}} \\
\leq \quad & \frac{M}{\alpha^{\mathrm{LB}}_{\mathrm{coer}}} \left(\frac{M}{\alpha_{\mathrm{coer}}}\right)^{1/2} \|u(\boldsymbol{\mu}_\ell;\cdot) - P_{U_{\ell-1}}\, u(\boldsymbol{\mu}_\ell;\cdot)\|_{\mathcal{X}} \ ; \qquad (20)
\end{aligned}
$$

hence, for Step $\mathbf{ii}'$, we obtain a slight modification to (17),

$$
|\alpha^i_\ell(\boldsymbol{\mu})| \leq \frac{M}{\alpha^{\mathrm{LB}}_{\mathrm{coer}}} \left(\frac{M}{\alpha_{\mathrm{coer}}}\right)^{1/2} . \qquad (21)
$$

(Note that, typically, $\alpha^{\mathrm{LB}}_{\mathrm{coer}}$ is quite close to α_{coer}.)

We can thus obtain

Corollary 3. *Theorem 2 applies to the Greedy algorithm with error bounds (Step* $\mathbf{ii}'$ *in place of* $\mathbf{ii}$*) if we strengthen our requirement on the exponent* α *to* $\alpha > \log(1 + (M/\alpha^{\mathrm{LB}}_{\mathrm{coer}})\sqrt{M/\alpha_{\mathrm{coer}}})$.

4. Conclusion

The results proven in this paper provide a first a priori analysis of the greedy algorithm for the selection of the elements used in the reduced basis for the approximation of parameter dependent PDE's. It is proven that the approximation properties of such a basis lead to an error that is distant from the best possible choice (given by the definition of the *Kolmogorov n-width*) by an exponential factor (see e.g. theorem 1). In the case in which the n-width is going to zero exponentially fast, the greedy maintains a exponential convergence in the reduced basis approximation.

Three comments are in order :

(1) We first report that in many cases that we have encountered, the convergence of the reduced basis method holds with a rate faster than exponential indicating that the assumed exponential decay of the

Kolmogorov n-width is conservative. In these cases, the loss of an exponential factor does not affect much the optimal rate.

(2) We have exhibited an example where the maximal loss predicted by our analysis is actually obtained when the convergence rate with a basis of dimension n is compared to the *Kolmogorov n-width*

(3) A very recent contribution [1] reports another comparison between the convergence rate obtained with a basis of dimension m and the *Kolmogorov n-width* with $n < m$. The loss is then different and much weaker if the *Kolmogorov n-width* has a polynomial decay. Nevertheless for faster decays — in particular those that we observe in our computations — this new analysis [1] provides a weaker convergence rate for the a priori analysis.

References

[1] Peter Binev, Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *Preprint : http://www.ann.jussieu.fr/ cohen/BCDDPW_10.pdf.*

[2] A. Kolmogoroff. Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse. *Ann. of Math. (2)*, 37(1):107–110, 1936.

[3] Y. Maday, A.T. Patera, and G. Turinici. A priori convergence theory for reduced-basis approximations of single-parametric elliptic partial differential equations. *Journal of Scientific Computing*, 17(1-4):437–446, 2002.

[4] Yvon Maday, T. Patera, Anthony, and Gabriel Turinici. Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations. *C. R. Acad. Sci., Paris, Sér. I, Math.*, 335:289–294, 2002.

[5] J. M. Melenk. On n-widths for elliptic problems. *J. Math. Anal. Appl.*, 247(1):272–289, 2000.

[6] Allan Pinkus. *n-widths in approximation theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1985.

[7] 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 — application to transport and continuum mechanics. *Archives of Computational Methods in Engineering*, 15(3):229–275, 2008.

[8] S. Sen. Reduced-basis approximation and a posteriori error estimation for many-parameter heat conduction problems. *Numerical Heat Transfer, Part B*, 54(5):369–389, 2008.

[9] K. Veroy, C. Prud'homme, D. V. Rovas, and A. T. Patera. *A Posteriori* error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. In *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*, 2003. Paper 2003-3847.