ESAIM: M2AN 54 (2020) 1509–1524
https://doi.org/10.1051/m2an/2020004

ESAIM: Mathematical Modelling and Numerical Analysis
www.esaim-m2an.org

REDUCED BASIS GREEDY SELECTION USING RANDOM TRAINING SETS

Albert Cohen[1] Wolfgang Dahmen[2] Ronald DeVore[3] and James Nichols[1]

Abstract. Reduced bases have been introduced for the approximation of parametrized PDEs in applications where many online queries are required. Their numerical efficiency for such problems has been theoretically con rmed in Binev *et al* (*SIAM J Math Anal* **4** (2011) 1457–1472) and DeVore *et al* (*Constructive Approximation* **7** (2013) 455–466), where it is shown that the reduced basis space V_n of dimension n, constructed by a certain greedy strategy, has approximation error similar to that of the optimal space associated to the Kolmogorov n-width of the solution manifold. The greedy construction of the reduced basis space is performed in an offline stage which requires at each step a maximization of the current error over the parameter space. For the purpose of numerical computation, this maximization is performed over a nite *training set* obtained through a discretization of the parameter domain. To guarantee a nal approximation error for the space generated by the greedy algorithm requires in principle that the snapshots associated to this training set constitute an approximation net for the solution manifold with accuracy of order . Hence, the size of the training set is the covering number for and this covering number typically behaves like $\exp(C^{1/s})$ for some $C > 0$ when the solution manifold has n-width decay $O(n^{-s})$. Thus, the shear size of the training set prohibits implementation of the algorithm when is small. The main result of this paper shows that, if one is willing to accept results which hold with high probability, rather than with certainty, then for a large class of relevant problems one may replace the ne discretization by a random training set of size polynomial in ^{1}. Our proof of this fact is established by using inverse inequalities for polynomials in high dimensions.

Mathematics Subject Classi cation. 62M45, 65D05, 68Q32, 65Y20, 68Q30, 35B30, 41A10, 41A25, 58D15, 65C99.

Received October 22, 2018. Accepted January 13, 2020.

1. Introduction

Complex systems are frequently described by parametrized PDEs that take the general form

$$\mathcal{P}(u, y) = 0. \tag{1.1}$$

Here $y = (y_j)_{j=1,\dots,d}$ is a vector of parameters ranging over some domain $Y \subset \mathbb{R}^d$ and $u = u(y)$ is the corresponding solution which is assumed to be uniquely defined in some Hilbert space V for every $y \in Y$.

Keywords and phrases. Reduced bases, performance bounds, random sampling, entropy numbers, Kolmogorov n-widths, sparse high-dimensional polynomial approximation.

[1] Laboratoire Jacques-Louis Lions, Sorbonne Université, Paris, France.

[2] Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA.

[3] Department of Mathematics, Texas A&M University, College Station, TX 77840, USA.

*Corresponding author: cohen@ann jussieu fr

Throughout this paper, we denote by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ the norm and inner product of V. In what follows, we assume that the parameters have been rescaled so that $Y = [\ \ 1,1]^d$. Here d is typically large and in some cases $d = \infty$. We seek results that are immune to the size of d, *i.e.*, are dimension independent.

Various *reduced modeling* approaches have been developed for the purpose of efficiently approximating the solution $u(y)$ in the context of applications where the solution map

$$y \mapsto u(y), \tag{1.2}$$

needs to be queried for a large number of parameter values $y \in Y$. This need occurs for example in optimal design or inverse problems where such parameters need to be optimized. The strategy consists in first constructing in some *offline stage* a linear space V_n of hopefully low dimension n, that provides a reduced map

$$y \mapsto u_n(y) \in V_n, \tag{1.3}$$

that approximates the solution map to the required target accuracy ε for all queries of $u(y)$. The reduced map is then implemented in the *online stage* with greatly reduced computational cost, typically polynomial in n.

As opposed to standard approximation spaces such as finite elements, the spaces V_n are specifically designed to approximate the image of u, *i.e.*, the elements in the parametrized family

$$\mathcal{M} := \{u(y)\ :\ y \in Y\}, \tag{1.4}$$

which is called the *solution manifold*. The optimal n-dimensional linear approximation space for this manifold is the one that achieves the minimum in the definition of the Kolmogorov n-width

$$d_n = d_n(\mathcal{M})_V := \inf_{\dim(E)=n} \sup_{u\in\mathcal{M}} \|u \quad P_E u\|. \tag{1.5}$$

This optimal space is computationally out of reach and the above quantity should be viewed as a benchmark for more practical methods. Here P_E denotes again the V-orthogonal projection onto E for any subspace E of V.

One approach for constructing a reduced space V_n, which comes with substantial theoretical footing and will be instrumental in our discussion, consists in proving that the solution map is analytic in the parameters y and has a Taylor expansion

$$u(y) = \sum_{\nu\in\mathcal{F}} t_\nu y^\nu, \tag{1.6}$$

where $y^\nu := \prod_{j=1}^d y_j^{\nu_j}$ and $\mathcal{F} := \{\nu \in \ \ ^d\}$. In the case of countably many parameters, $\mathcal{F}$ is the set of finitely supported sequences $\nu = (\nu_j)_{j\geq 1}$ with $\nu_j \in \ \ $. One then proves that the sequence $(t_\nu)_{\nu\in\mathcal{F}}$ of Taylor coefficients has some decay property. Two prototypical examples of results on decay are given in [1,7] for the elliptic equation

$$\operatorname{div}(a\nabla u) = f, \tag{1.7}$$

where the diffusion coefficient function has the parametrized form

$$a = a(y) = \overline{a} + \sum_{j\geq 1} y_j \psi_j, \tag{1.8}$$

for some given functions $\overline{a}$ and $(\psi_j)_{j\geq 1}$. These results (see *e.g.* Thms. 1.1 and 1.2 in [1]) show that, under mild decay or summability conditions on the functions ψ_j, one has that the sequence $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ is in ℓ^p for certain $p < 1$ with a bound on the ℓ^p quasi-norm. It then follows that for each n, there is a set $\ \ _n \subset \mathcal{F}$ with $\#(\ \ _n) = n$ such that

$$\sup_{y\in Y} \|u(y) \quad \sum_{\nu\in\ \ _n} t_\nu y^\nu\| \leq C n^{\ \ r}, \quad r := 1/p \quad 1. \tag{1.9}$$

Similar results have been obtained for more general models of linear and nonlinear PDEs, see in particular [5,6], using orthogonal expansions into tensorized Legendre polynomials in place of Taylor series.

Therefore the space $V_n := \operatorname{span}\{t_\nu \ : \ \nu \in \ _n\}$ has dimension n with an *a priori* bound Cn^{-r} on its approximation error for all members of the solution manifold $\mathcal{M}$. One choice of $_n$ giving (1.9) is the set of indices corresponding to the n largest $\|t_\nu\|_V$. Further analysis Cohen and Migliorat [8] shows that the same convergence estimate can be obtained imposing in addition that the sets $_n$ are *downward closed* (or *lower sets*), *i.e.* having the property

$$\nu \in \ _n \quad \text{and} \quad \mu \le \nu \implies \mu \in \ _n, \tag{1.10}$$

where $\mu \le \mu$ is to be understood componentwise. We stress that the rate of decay n^{-r} in the bound (1.9) may be suboptimal compared to the actual rate of decay of the n-width $d_n(\mathcal{M})$.

The present paper is concerned with another prominent reduced modeling strategy known as the *Reduced Basis Method* (RBM). In this approach [10,11,14,15], particular snapshots

$$u^k = u(y^k), \quad k = 1, 2, \dots, \tag{1.11}$$

are selected in the solution manifold and the space V_n is defined by

$$V_n := \operatorname{span}\{u^1, \dots, u^n\}. \tag{1.12}$$

A certain *greedy procedure* has been first proposed in [12] and analyzed in [3] for selecting these snapshots. It was shown in [2,9] that the approximation error

$$\sigma_n = \sigma_n(\mathcal{M})_V := \sup_{u \in \mathcal{M}} \|u \quad P_{V_n} u\|, \tag{1.13}$$

provided by the resulting spaces has the same rate of decay (polynomial or exponential) as that of d_n. In this sense, the method leads to reduced models with optimal performance, in contrast to sparse polynomial expansions.

In its simplest (and idealized) form, the greedy procedure can be described as follows: at the initial step, one sets $V_0 = \{0\}$, and given that V_n has been produced after n steps, one selects the new snapshot by

$$u^{n+1} := \operatorname*{argmax}_{u \in \mathcal{M}} \|u \quad P_{V_n} u\|. \tag{1.14}$$

Each greedy step thus amounts to maximizing

$$e_n(y) := \|u(y) \quad P_{V_n} u(y)\| \tag{1.15}$$

over the parameter domain Y. While in this precise form the scheme cannot be realized in practice an important modification of this greedy selection, known as *the weak greedy algorithm*, allows the selection to be done in a a practically feasible manner while retaining the same performance guarantees, see Section 2 below.

The optimization in the greedy algorithm is typically performed by replacing Y at each step by a discrete *training set* $\tilde{Y}$. In order to retain the performance guarantees of the greedy algorithm, this discretization should in principle be chosen fine enough so that the solution map $y \mapsto u(y)$ is resolved up to the target accuracy $\varepsilon > 0$, that is, the discrete set

$$\widetilde{\mathcal{M}} = \{u(y) \ : \ y \in \tilde{Y}\}, \tag{1.16}$$

is an ε-approximation net for $\mathcal{M}$.

Although performed in the offline stage, this discretization becomes computationally problematic when the parametric dimension d is either large or infinite, due to the prohibitive size of this net as $\varepsilon \to 0$. For example, in the typical case when the Kolmogorov width of $\mathcal{M}$ decays like $O(n^{-s})$ for some $s > 0$, we can invoke Carl's inequality [13] to obtain a sharp bound $e^{c\varepsilon^{-1/s}}$ for the cardinality of $\tilde{\mathcal{M}}$ and $\tilde{Y}$. This exponential growth

drastically limits the possibility of using ε-approximation nets in practical applications when the number of involved parameters become large.

There is a preference toward the use of the greedy constructions over the Taylor expansion constructions because they guarantee error decay comparable to the decay of Kolmogorov widths while the Taylor polynomial constructions do not provide any such guarantee. Therefore, it is of interest to understand whether the apparent impediment of requiring such a fine discretization of the solution manifold can somehow be avoided or significantly mitigated. The main result of this paper is to prove that this is indeed the case provided that one is willing to accept error guarantees that hold with *high probability* rather than with certainty. Our main result shows that a target accuracy ε can generally be met with high probability by searching over a randomly discretized set $\tilde{Y}$ whose size grows only polynomially in ε^{-1} rather than exponentially, in contrast to ε-approximation nets.

The paper is organized as follows. In Section 2, we elaborate on the weak form of the greedy algorithm, which is used in numerical computation, and recall some known facts on its performance and complexity. In Section 3, we use properties of downward closed polynomial approximation to show how a random sampling $\tilde{Y}$ provides an approximate solution of the optimization problem engaged at each step of the greedy algorithm. In Section 4, we formulate our modification of the greedy algorithm based on such random selection and then analyze its performance. Finally, we illustrate in Section 5 the validity of the randomized approach by some numerical tests performed in parametric dimensions up to $d = 64$ for which the size of ε-approximation nets become computationally prohibitive.

2. Performance and complexity of reduced basis greedy algorithms

The greedy selection process described in the introduction is not practically feasible, due to at least three obstructions:

(1) Given a parameter value y, the snapshot $u(y)$, in particular, the generators of the reduced spaces cannot be exactly computed.
(2) For a given $y \in Y$, the quantity $e_n(y)$ to be maximized cannot be exactly evaluated.
(3) The map $y \mapsto e_n(y)$ is non-convex/non-concave and therefore difficult to maximize, even if it could be exactly evaluated.

The first obstruction can be handled when a numerical solver is available for computing an approximation

$$y \mapsto u_h(y), \tag{2.1}$$

of $u(y)$ to any prescribed accuracy $\varepsilon_h > 0$, that is, such that

$$\sup_{y \in Y} \|u(y) \quad u_h(y)\| \leq \varepsilon_h. \tag{2.2}$$

Here $h > 0$ is a space discretization parameter: typically, u_h belongs to a finite element space V_h of meshsize h and (possibly very large) dimension n_h. The selected reduced basis functions are now given by $u_i = u_h(y^i) \in V_h$ and therefore the reduced basis space V_n is a subspace of V_h. Whenever the n-widths d_n decay much faster than the approximation order provided by V_h the reduced space V_n has typically much smaller dimension than V_h, that is

$$n \ll n_h. \tag{2.3}$$

This yields substantial computational savings when using the reduced basis discretization in the online stage.

Note that this numerical solver allows us in principle to also handle the second obstruction: we could now perform the greedy algorithm by maximizing at each step the quantity

$$e_{n,h}(y) := \|u_h(y) \quad P_{V_n} u_h(y)\|, \tag{2.4}$$

which, in contrast to $e_n(y)$, can be exactly computed and satisfies

$$|e_n(y) \quad e_{n,h}(y)| \leq \varepsilon_h, \quad y \in Y. \tag{2.5}$$

In other words, the greedy algorithm is applied on the approximate solution manifold

$$\mathcal{M}_h := \{u_h(y) \ : \ y \in Y\}. \tag{2.6}$$

While the quantity $e_{n,h}(y)$ can in principle be computed exactly, the complexity of this computation depends, at least in a linear manner, on the dimension $n_h = \dim(V_h)$, which is typically much higher than n. Substantial computational saving may still be obtained when maximizing instead an *a posteriori* estimator $\overline{e}_{n,h}(y)$ of this quantity that satisfies

$$\overline{e}_{n,h}(y) \leq e_{n,h}(y) \leq \beta \overline{e}_{n,h}(y), \quad y \in Y. \tag{2.7}$$

The computation of $\overline{e}_{n,h}(y)$ for a given $y \in Y$ is based in particular on replacing the orthogonal projection $P_{V_n} u_h(y)$ by a Galerkin projection. It, in turn, does *not* require the computation of $u_h(y)$ and entails a computational cost $c(n)$ depending on the small dimension n, typically in a polynomial way, rather than on the large dimension n_h. We refer to [6] for the derivation of a residual-based estimator $\overline{e}_{n,h}(y)$ having these properties in the case of elliptic PDEs with affine parameter dependence.

Maximizing the *a posteriori* estimator $\overline{e}_{n,h}(y)$ amounts to applying to $\mathcal{M}_h$ a so called *weak-greedy* algorithm, where u^{n+1} now satisfies

$$\|u^{n+1} \quad P_{V_n} u^{n+1}\| \geq \gamma \|u \quad P_{V_n} u\|, \quad u \in \mathcal{M}_h, \tag{2.8}$$

with parameter $\gamma := \quad /\beta \in]0,1[$.

For such an algorithm, it was proved in [2,9] that any polynomial or exponential rate of decay achieved by the Kolmogorov n-width d_n is retained by the error performance σ_n for this algorithm. More precisely, the following holds, for any compact set $\mathcal{K}$ in a Hilbert space V, see [6].

Theorem 2.1. *Let $d_n = d_n(\mathcal{K})_V$ be the n-widths of the solution manifold. Consider the weak greedy algorithm with threshold parameter γ. For any $C_0 > 0$ and $s > 0$, we have*

$$d_n \leq C_0(\max\{1,n\}) \ ^s, \quad n \geq 0 \quad \Longrightarrow \quad \sigma_n \leq C_s \gamma \ ^2(\max\{1,n\}) \ ^s, \quad n \geq 0, \tag{2.9}$$

where $C_s := 2^{4s+1} C_0$. For any $c_0, C_0 > 0$ and $s > 0$, we have

$$d_n \leq C_0 e \ ^{c_0 n^s}, \quad n \geq 0 \quad \Longrightarrow \quad \sigma_n \leq C_s \gamma \ ^1 e \ ^{c_1 n^s}, \quad n \geq 0, \tag{2.10}$$

where $c_1 = \frac{c_0}{2} 3 \ ^s$ and $C_1 := C_0 \max\{\sqrt{2}, e^{c_1}\}$.

Remark 2.2. If the same rates of d_n in the above theorem are only assumed within a limited range $0 \leq n \leq n^*$, then the same decay rates of σ_n are achieved for the same range $0 \leq n \leq n^*$, up to some minor changes in the expressions of the constants c_1 and C_1, independently of n^*.

Remark 2.3. The reduced basis algorithm aims to construct an n-dimensional linear space that is tailored for the approximation of *all* solutions that constitute the solution manifold $\mathcal{M}$. Therefore, its performance σ_n is always bounded from below by the n-width $d_n = d_n(\mathcal{M})$. Assuming some rate of decay on d_n, at least polynomial, is thus strictly necessary if we want the reduced basis method to converge at such a rate. As discussed in the introduction, in the high dimensional parametric context, such rate can be rigourously established for certain instances of linear elliptic PDE's with affine parametrization of the diffusion function. A more general approach applicable to nonlinear PDE's and non-affine parametrizations for proving such rates is given in [6]. On the other hand, it is known that $d_n(\mathcal{M})$ has poor decay for certain categories of parametrized problems. This includes, in particular, transport dominated problems with sharp transition locus that varies with the value of the parameter y. Such problems are thus intrinsically not well tailored to reduced basis methods.

The additional perturbations due to the numerical solver and the *a posteriori* error indicator can thus be incorporated in the analysis of the reduced basis algorithm. If $\varepsilon > 0$ *a posteriori* is our final target accuracy, we set the space discretization parameter h so that $\varepsilon_h = \frac{\varepsilon}{2}$ where ε_h is the space discretization error bound in (2.2). We then apply the greedy selection on $\mathcal{M}_h$ based on maximizing $\overline{e}_{n,h}(y)$ until we are ensured that $e_{n,h}(y) \leq \frac{\varepsilon}{2}$ for all $y \in Y$. The target accuracy $e_n(y) \leq \varepsilon$ is thus met for all $y \in Y$ in view of (2.5).

Note that a decay rate $d_n(\mathcal{M}) \leq \gamma(n)$ for some decreasing sequence $\gamma(n)$ implies a comparable rate $d_n(\mathcal{M}_h) \leq 2\gamma(n)$, for the range $n \leq n^*$ where n^* is the largest value of n such that $\gamma(n) \geq \varepsilon_h$. Therefore, using Remark 2.2 in conjunction with Theorem 2.1 applied to $\mathcal{M}_h$, we obtain an estimate on the number of greedy steps $n(\varepsilon)$ that are necessary to reach the target accuracy ε.

Corollary 2.4. *Let $d_n = d_n(\mathcal{M})_V$ be the n-width of the solution manifold. For any $s > 0$ and $C_0 > 0$,*

$$d_n \leq C_0(\max\{1,n\})^{-s}, \quad n \geq 0 \quad \Longrightarrow \; n(\varepsilon) \leq C_1 \varepsilon^{-\frac{1}{s}}, \quad \varepsilon > 0, \tag{2.11}$$

where C_1 depends on C_0, s and γ. For any $s > 0$ and $c_0, C_0 > 0$,

$$d_n \leq C_0 e^{-c_0 n^s}, \quad n \geq 0 \quad \Longrightarrow \quad n(\varepsilon) \leq C_1 \max\{\log(c_1\varepsilon)^{1/s}, 0\}, \quad \varepsilon > 0, \tag{2.12}$$

where C_1 and c_1 depend on C_0, c_0, s and γ.

The difficulty in item 3 is the most problematic one, in particular when the parametric variable y is high-dimensional, and is the main motivation for the present work. Since the quantities $e_n(y)$, $\overline{e}_n(y)$, $e_{n,h}(y)$ and $\overline{e}_{n,h}(y)$ may have many local maxima, continuous optimization techniques are not appropriate. A typically employed strategy is therefore to replace the continuous optimization over Y by its discrete optimization over a training set $\widetilde{Y} \subset Y$ of finite size. This amounts to applying the greedy or weak-greedy algorithm to the discretized manifold $\widetilde{\mathcal{M}}$ defined by (1.16) or, more practically, to its approximated version

$$\widetilde{\mathcal{M}}_h := \{u_h(y) \; : \; y \in \widetilde{Y}\}. \tag{2.13}$$

On a first intuition, the discretization should be sufficiently fine so that the manifold $\mathcal{M}_h$ is resolved with accuracy of the same order as the target accuracy ε. Recall that if K is a compact set in some normed space, a finite set S is called a δ-net of K if

$$K \subset \bigcup_{v \in S} B(v, \delta), \tag{2.14}$$

that is, any $u \in K$ is at distance at most δ from some $v \in S$.

The perturbation of the greedy algorithm due to this discretization can be accounted for jointly with the previously discussed perturbation, namely finite element approximation and *a posteriori* error estimation. Assuming for example that $\widetilde{\mathcal{M}}_h$ is a $\varepsilon/3$-net of $\mathcal{M}_h$, we set the space discretization parameter h so that $\varepsilon_h = \frac{\varepsilon}{3}$. We then apply the greedy selection on $\widetilde{\mathcal{M}}_h$ based on maximizing $\bar{e}_{n,h}(y)$ over $\widetilde{Y}$ until we are ensured that $e_{n,h}(y) \leq \frac{\varepsilon}{3}$ for all $y \in \widetilde{Y}$. By the covering property, we have $e_{n,h}(y) \leq 2\varepsilon/3$ for all $y \in Y$, and therefore the target accuracy $e_n(y) \leq \varepsilon$ is met for all $y \in Y$. In addition, since $d_n(\widetilde{\mathcal{M}}_h)_V \leq d_n(\mathcal{M}_h)_V$, the statement of Corollary 2.4 remains unchanged for this discretized algorithm.

The main problem with this approach is that the current estimates on the size of an ε-net of $\mathcal{M}$ or $\mathcal{M}_h$ become extremely large as $\varepsilon \to 0$, especially in high parametric dimension.

A first natural strategy to generate such an ε-net would be to apply the solution map $y \mapsto u(y)$ to an ε-net for Y in a suitable norm, relying on a stability estimate for this map. For example, in the simple case of the elliptic PDE (1.7) with parametrized coefficients (1.8), one can easily establish a stability estimate of the form

$$\|u(y) \quad u(\tilde{y})\| \leq C\|y \quad \tilde{y}\|_{\ell^\infty}, \quad y, \tilde{y} \in Y, \tag{2.15}$$

under the minimal uniform ellipticity assumption $\sum_{j\geq 1}|\psi_j| \leq \min \overline{a} \quad \delta$ for some $\delta > 0$, with C depending on $\min \overline{a}$ and δ. Thus, one possible ε-net of $\mathcal{M}$ or $\mathcal{M}_h$ in the V norm is induced by a $C \ {}^1\varepsilon$-net $\widetilde{Y}$ of Y in the ℓ^∞ norm. However, the size of such a net scales like

$$\#(\widetilde{Y}) \sim \varepsilon \ {}^{cd}. \tag{2.16}$$

with the parametric dimension d, therefore suffering from the curse of dimensionality. In the case $d = \infty$, one would have to truncate the parametric expansion (1.8) for a given target accuracy ε, resulting into an active parametric dimension $d(\varepsilon) < \infty$. Assuming a polynomially decaying error $\|\sum_{j>k}|\psi_j|\|_{L^\infty} \lesssim k \ {}^b$, the growth of $d(\varepsilon)$ as $\varepsilon \to 0$ is in $\mathcal{O}(\varepsilon \ {}^{1/b})$ resulting in a training set of size scaling like

$$\#(\widetilde{Y}) \sim \varepsilon \ {}^{c\varepsilon \ {}^{1/b}}, \tag{2.17}$$

which is extremely prohibitive.

One sharper way to obtain an estimate independent of the parametric dimension is to use a fundamental result that relates covering and widths. We define the entropy number $\varepsilon_n := \varepsilon_n(\mathcal{M})_V$ as the smallest value of $\varepsilon > 0$ such that there exists a covering of $\mathcal{M}$ by 2^n balls of radius ε. Then, Carl's inequality [13] states that for any $s > 0$,

$$(n+1)^s \varepsilon_n \leq C_s \sup_{k=0,\ldots,n} (k+1)^s d_k, \tag{2.18}$$

where $d_k = d_k(\mathcal{M})_V$ and C_s is a fixed constant. This inequality shows that, in the case of polynomial decay $n \ {}^s$ of the n-widths, there exists an ε-net $\widetilde{\mathcal{M}}$ associated with a training set of size

$$\#(\widetilde{Y}) \sim 2^{c\varepsilon \ {}^{1/s}}. \tag{2.19}$$

While this estimate is more favorable than (2.17), it is still prohibitive. Moreover, the construction of such an ε-net, as in the proof of Carl's inequality, necessitates the knowledge of the approximation spaces V_n that perform with the n-width accuracy $n \ {}^s$ which is precisely the objective of the greedy algorithm.

The computational cost at each step n of the offline stage is determined by the product between $\#(\widetilde{Y})$ and the cost $c(n)$ of evaluating the error bound $\overline{e}_{n,h}(y)$ for an individual $y \in \widetilde{Y}$. Therefore, the prohibitive number of error bound evaluations is the limiting factor in practice and poses the main obstruction to the feasibility of certified reduced basis methods in the regime of polynomially decaying n-widths, and hence in particular, in the context of high parametric dimension.

In what follows, we show that this obstruction can be circumvented by not searching for an ε-net of $\mathcal{M}$ but rather defining $\widetilde{Y}$ by random sampling of Y. This approach allows us to significantly reduce the size of training sets used in greedy algorithms while still obtaining reduced bases with the same guarantee of performance at least with high probability. In order to keep our arguments and notation as simple and clear as possible, we do not consider the issue of space discretization and error estimation, assuming that we have access to $e_n(y)$ for each individual $y \in Y$. As just described, a corresponding finer analysis can incorporate the perturbation of using instead $\overline{e}_n(y)$, $e_{n,h}(y)$ or $\overline{e}_{n,h}(y)$, with the same resulting overall performance.

3. Polynomial approximation

Let $V_n := \operatorname{span}\{u^1, \ldots, u^n\}$ be the reduced basis space at the n-th step of the weak greedy algorithm. The next step of the greedy algorithm is to search over Y to find a point $y \in Y$ where

$$e_n(y) := \|u(y) \quad P_{V_n} u(y)\|_V \tag{3.1}$$

(in practice $\bar{e}_{n,h}(y)$) is large, hopefully close to its maximum over Y. In this section we show that random sampling gives a discrete set $\widetilde{Y}$, of moderate size, on which the maximum of $e_n(y)$ can be compared with the

maximum of $e_n(y)$ over all of Y with high probability. To obtain a result of this type we use approximation by polynomials.

Recall that $\subset \mathcal{F}$ is said to be a *downward closed set* if whenever $\nu \in$ and $\mu \leq \nu$, then $\mu \in$, where $\mu \leq \mu$ is to be understood componentwise. To such a set , we associate the multivariate polynomial space

$$\mathbb{P} := \operatorname{span}\left\{ y \mapsto y^\nu := \prod_{j\geq 1} y_j^{\nu_j} \ : \ \nu \in \right\}. \tag{3.2}$$

We define

$$\mathcal{P} := V \otimes \mathbb{P} , \tag{3.3}$$

the space of V-valued polynomials spanned by the same monomials. Thus, any polynomial P in $\mathcal{P}$ takes the form $P(y) = \sum_{\nu\in} a_\nu y^\nu$ where the a_ν are in V. For any $m \geq 0$, we let

$$\Sigma_m := \bigcup_{\#( )=m} \mathcal{P} , \tag{3.4}$$

where the union is over all downward closed sets of size m. Note that the union of all downward closed sets of cardinality m is the so-called hyperbolic cross $H_{m,d}$ that consists of all ν such that $\prod_{j=1}^d (1+\nu_j) \leq m$.

Given a function v in $L^\infty(Y,V)$, we consider its approximation in $L^\infty(Y,V)$ by the elements of Σ_m and the error

$$\delta_m(v) := \inf_{P\in\Sigma_m} \sup_{y\in Y} \|v(y) \quad P(y)\|_V. \tag{3.5}$$

For $r > 0$, a function v in $L^\infty(Y,V)$ is said to be in the approximation class $\mathcal{A}^r = \mathcal{A}^r((\Sigma_m)_{m\geq 1})$ if

$$\delta_m(v) \leq C m^{\ r}, \quad m \geq 1, \tag{3.6}$$

and the smallest such C defines a quasi-seminorm $|v|_{\mathcal{A}^r}$ for this class which is a linear subspace of $L^\infty(Y,V)$. A quasi-norm for this space is defined by

$$\|v\|_{\mathcal{A}^r} := \max\{\|v\|_{L^\infty(Y,V)}, |v|_{\mathcal{A}^r}\}. \tag{3.7}$$

Several foundational results in parametric PDEs prove that the solution map $y \mapsto u(y)$ belongs to classes $\mathcal{A}^r$, as already mentioned in our introduction. An important observation to us is that whenever $u \in \mathcal{A}^r$ and V_n is a finite dimensional subspace of V then both $P_{V_n} u$ and $u \quad P_{V_n} u$ are also in $\mathcal{A}^r$. For example, for any downward closed set $\subset \mathcal{F}$ and an approximation $P(y) = \sum_{\nu\in} a_\nu y^\nu$ to $u(y)$, the polynomial $Q(y) := \sum_{\nu\in} P_{V_n} a_\nu y^\nu$ is in $\mathcal{P}$ and

$$\|P_{V_n} u(y) \quad Q(y)\|_V = \|P_{V_n}(u(y) \quad P(y))\|_V \leq \|u(y) \quad P(y)\|_V. \tag{3.8}$$

It follows that $\delta_m(P_{V_n} u) \leq \delta_m(u)$ for all $m \geq 1$. The same holds for $u \quad P_{V_n} u = P_{V_n^\perp} u$. From this, one derives that

$$\|P_{V_n} u\|_{\mathcal{A}^r} \leq \|u\|_{\mathcal{A}^r} \quad \text{and} \quad \|u \quad P_{V_n} u\|_{\mathcal{A}^r} \leq \|u\|_{\mathcal{A}^r}. \tag{3.9}$$

The next result shows that when a function belongs to the class $\mathcal{A}^r$, its maximum over a random set of point $\widetilde{Y}$ is above a fixed fraction of its maximum over Y with some controlled probability.

Lemma 3.1. *Let $r > 1$ and suppose that $v \in \mathcal{A}^r$ with $\|v\|_{\mathcal{A}^r} \leq M_0$ and $\|v\|_{L^\infty(Y,V)} = M$. Let m be an integer such that*

$$4 M_0 m^{\ r+1} < M. \tag{3.10}$$

If $\widetilde{Y}$ is any nite set of N points drawn at random from Y with respect to the uniform probability measure on Y, then

$$\sup_{y\in Y} \|v(y)\|_V \geq \frac{M}{8m} \tag{3.11}$$

with probability larger than $1 \quad 1 \quad \frac{3}{4m^2}\Big)^N$.

Proof. From the definition of m, there exists a downward closed set $\quad$ with $\#(\ \ ) = m$ and a V-valued polynomial $P \in \mathcal{P}$ $\quad$ such that

$$\|v \quad P\|_{L^\infty(Y,V)} \le \frac{M}{4m}. \tag{3.12}$$

We use the Legendre polynomials to represent P. We denote by $(L_j)_{j\ge 0}$ the sequence of univariate Legendre polynomials normalized in $L^2([\ \ 1,1], \frac{\mathrm{d}t}{2})$. Their multivariate counterparts

$$L_\nu(y) := \prod_{\nu_j \neq 0} L_{\nu_j}(y_j), \quad \nu \in \mathcal{F}, \tag{3.13}$$

are an orthonormal basis on $L^2(Y,\rho)$, where ρ is the uniform probability measure on Y. We write P in its Legendre expansion

$$P(y) = \sum_{\nu\in} c_\nu L_\nu(y), \quad c_\nu := \int_Y P(y)L_\nu(y)\,\mathrm{d}\rho, \tag{3.14}$$

where the coefficients c_ν are elements of V. We next invoke a result from [4] which says that for any downward closed set $\quad$, one has

$$\max_{y\in Y} \sum_{\nu\in} |L_\nu(y)|^2 \le \#(\ \ )^2. \tag{3.15}$$

Thus, it follows from the Cauchy-Schwartz inequality that for any $y \in Y$,

$$\|P(y)\|_V^2 \le \sum_{\nu\in} \|c_\nu\|_V^2 \sum_{\nu\in} |L_\nu(y)|^2 \le m^2 \int_Y \|P(y)\|_V^2\,\mathrm{d}\rho, \tag{3.16}$$

or equivalently

$$\|P\|_{L^\infty(Y,V)} \le m\|P\|_{L^2(Y,V)}. \tag{3.17}$$

Now, let $S := \{y \in Y \ : \ \|P(y)\|_V \ge \delta\}$ where $\delta := \frac{M_1}{2m}$ with $M_1 := \|P\|_{L^\infty(Y,V)}$. Then,

$$\|P\|^2_{L^2(Y,V)} \le \int_S \|P(y)\|_V^2\,\mathrm{d}\rho + \int_{S^c} \|P(y)\|_V^2\,\mathrm{d}\rho \le M_1^2\rho(S) + \delta^2. \tag{3.18}$$

Inserting this into (3.17) gives

$$M_1^2 m \ \ ^2 \le M_1^2\rho(S) + \delta^2 \le M_1^2\rho(S) + m \ \ ^2\frac{M_1^2}{4}. \tag{3.19}$$

In other words,

$$\rho(S) \ge \frac{3}{4m^2}. \tag{3.20}$$

Suppose now that $\widetilde{Y}$ is a set formed by N independent draws with respect to the uniform measure ρ on Y. The probability that none of these draws is in S is at most $(1 \quad \frac{3}{4m^2})^N$. So, with probability greater than $1 \quad (1 \quad \frac{3}{4m^2})^N$, we have

$$\max_{y\in Y} \|P(y)\|_V \ge \delta = \frac{M_1}{2m}. \tag{3.21}$$

Accordingly, with at least the same probability, we have from (3.12)

$$\max_{y\in Y} \|v(y)\|_V \ge \delta \quad \frac{M}{4m} \ge \frac{M_1}{2m} \quad \frac{M}{4m} \ge \frac{M \quad \frac{M}{4m}}{2m} \quad \frac{M}{4m} \ge \frac{M}{8m}, \tag{3.22}$$

which proves the lemma.

Remark 3.2. Intuitively, the above proof relies on the fact that v is close to a polynomial P, and that the L^∞ norm of P on a sufficiently fine discrete set $\tilde{Y}$ is comparable to its L^∞ norm on the continuous domain Y. A general line of research is to look for equivalences between discrete and continuous L^p norms for a given m-dimensional space X_m of functions, thus of the form

$$c_1\|v\|_{L^p(Y)} \leq \left(\frac{1}{\#(\tilde{Y})}\sum_{y\in\tilde{Y}}|v(y)|^p\right)^{1/p} \leq c_2\|v\|_{L^p(Y)}, \quad v \in X_m. \tag{3.23}$$

Such results are refered to as Marcinkiewicz-type discretization theorems, see [16] for a recent survey. In several settings where X_m consists of algebraic or trigonometric polynomials, it is known that random sampling for $\tilde{Y}$ yields such inequalities with high probability at a sampling budget $\#(\tilde{Y})$ that grows polynomially, and sometimes linearly, with m. Note that the inequality (3.15) is used in [4] to show that for L^2 norms and downward closed polynomials spaces in any dimension, the norm equivalence (3.23) holds with high probability for random samples of cardinality $\mathcal{O}(\#(\ \ )^2)$ up to logarithmic factors.

Remark 3.3. The result in the above lemma can be improved by sampling according to the tensor product Chebychev measure, that is, with ρ now defined as

$$\mathrm{d}\rho(y) := \bigotimes_{j\geq 1}\frac{\mathrm{d}y_j}{\pi\sqrt{1\quad y_j^2}}. \tag{3.24}$$

Indeed, we can apply a similar reasoning with the L_ν replaced by the tensorized Chebychev polynomials T_ν, for which it is proved in [4] that

$$\max_{y\in Y}\sum_{\nu\in}|L_\nu(y)|^2 \leq \#(\ \ )^2\ , \qquad := \frac{\ln 3}{2\ln 2}. \tag{3.25}$$

for any downward closed set . Therefore, the statement of the lemma is modified as follows: if m is such that $4M_0 m^{\ r+} < M$ and $\widetilde{Y}$ is any finite set of N points drawn at random from Y with respect ρ, then

$$\sup_{y\in Y}\|v(y)\|_V \geq \frac{M}{8m} \tag{3.26}$$

with probability larger than 1 1 $\frac{3}{4m^2}\Big)^N$.

4. The main result

We are now in position to formulate our main result. We suppose that we are given an error tolerance ε and we wish to use a greedy algorithm to construct a space V_n such that with high probability, say probability greater than 1 η, we have

$$\mathrm{dist}(\mathcal{M}, V_n) \leq \varepsilon, \tag{4.1}$$

with n hopefully small and the off-line complexity also acceptable. We assume that the solution map $y \mapsto u(y)$ belongs to $\mathcal{A}^r$ for some $r > 2$ and that we have an upper bound M_0 for $\|u\|_{\mathcal{A}^r}$. This assumption is known to hold in a great variety of settings of parametric PDEs, see [6].

Given r and the user prescribed η, we first define m as the smallest integer such that

$$32M_0 m^{\ r+2} \leq \varepsilon \quad \text{and} \quad 2^{4r+2}m^{\ r} \leq 1. \tag{4.2}$$

We then define N as the smallest integer such that

$$\left(1 \quad \frac{3}{4m^2}\right)^N \leq \frac{\eta}{m^2}. \tag{4.3}$$

We consider the following greedy algorithm for finding a reduced basis. In the first step, we make N independent draws of the parameter y according to the uniform measure ρ. This produces a set $\widetilde{Y}_0$ of cardinality N. We then use

$$u^0 := u(y^0), \quad y^0 := \operatorname*{argmax}_{y\in Y_0} \|u(y)\|_V. \tag{4.4}$$

At the general step, once $u^0, \ldots, u^n \ ^1$ and $V_n := \operatorname{span}\{u^0, \ldots, u^n \ ^1\}$ have been chosen, we make N independent draws of the parameter y, producing the set $\widetilde{Y}_n$, and then define

$$u^n := u(y^n), \quad y^n := \operatorname*{argmax}_{y\in Y_n} e_n(y), \tag{4.5}$$

where $e_n(y) := \|u(y) \quad P_{V_n}u(y)\|_V$. In practice, each $e_n(y)$ to be computed is only approximately computed through a surrogate $\overline{e}_{n,h}(y)$ and u^n is only approximated as described in Section 2. However, we do not incorporate these facts in the analysis that follows in order to simplify the presentation.

Let

$$\hat{\sigma}_n := \max_{y\in Y_n} e_n(y), \quad \sigma_n := \max_{y\in Y} e_n(y), \quad n \geq 0, \tag{4.6}$$

be the computed error and the true error for approximation of $\mathcal{M}$ by V_n. We terminate the algorithm at the smallest integer $n \leq m^2$ for which $\hat{\sigma}_n \leq \frac{\varepsilon}{8m}$. If this does not occur before m^2 steps we then terminate after step $n = m^2$ has been completed. The V_n is the output of the algorithm.

Theorem 4.1. *Assume that the solution map $y \mapsto u(y)$ belongs to $\mathcal{A}^r$ for some $r > 2$ with upper bound M_0 for $\|u\|_{\mathcal{A}^r}$. Then, with probability greater than $1 \quad \eta$ the following hold for the above numerical algorithm:*

(i) *The algorithm produces a reduced basis space V_n such that $\operatorname{dist}(\mathcal{M}, V_n) \leq \varepsilon$.*

(ii) *If for some $s > 0$ and $C_0 > 0$, we have $d_n(\mathcal{M}) \leq C_0 \max\{1, n\}^{\ s}$ for all $n \geq 0$, then the algorithm terminates in $n(\varepsilon)$ steps, where*

$$n(\varepsilon) \leq C\varepsilon^{\ (\frac{1}{s}+\frac{3}{s\ r\ 2)}}, \tag{4.7}$$

and requires $N(\varepsilon)$ error bound evaluations, where

$$N(\varepsilon) = C\varepsilon^{\ \frac{2s+r+1}{s\ r\ 2)}}(|\ln\eta| + |\ln\varepsilon|). \tag{4.8}$$

The constants C in the above bounds depend only on (r, s, C_0, M_0).

Remark 4.2. Note that the assumption $u \in \mathcal{A}^r$ implies in particular that the Kolmogorov n-width of $\mathcal{M}$ decays at least like $n^{\ r}$ and therefore we may assume that $s \geq r$ in the above theorem, although this is not used in the proof.

Proof. We first show that with probability greater than $1 \quad \eta$, the algorithm produces at each step $k \leq n$ a snapshot $u^k = u(y^k)$ which realizes a weak greedy algorithm, applied over *all* of Y, with parameter $\gamma := \frac{1}{8m}$. Indeed, for any k, let $v(y) = u(y) \quad P_{V_k}u(y)$ be the error function at the step after V_k is defined. As shown in the previous section, $v \in \mathcal{A}^r$ and $\|v\|_{\mathcal{A}^r} \leq M_0$. Since the algorithm has not terminated, we have

$$\|v\|_{L^\infty(Y,V)} = \sigma_k \geq \hat{\sigma}_k \geq \frac{\varepsilon}{8m} \geq 4M_0 m^{\ r+1}, \tag{4.9}$$

where the last inequality is the first condition in (4.2). Therefore, we can apply Lemma 3.1 to v and find that with probability greater than $1 \quad 1 \quad \frac{3}{4m^2}\Big)^N$, and thus from (4.3) with probability greater than $1 \quad \frac{\eta}{m^2}$, we have

$$\hat{\sigma}_k = \max_{y\in Y_k} \|v(y)\|_V \geq \gamma \sup_{y\in Y} \|v(y)\|_V = \gamma\sigma_k. \tag{4.10}$$

This means that with this probability the function u^{k+1} is a selection of the weak greedy algorithm with parameter γ. Since, the draws are independent and there are at most m^2 sets $\widetilde{Y}_k$, the union bound implies that with probability at least $1 \quad \eta$, the sequence $u^1, \ldots, u^n$ is a sequence that is the realization of the weak greedy algorithm with this parameter. For the remainder of the proof, we put ourselves in the case of favorable probability.

Now consider the termination of the algorithm. If $n < m^2$, then

$$\sigma_n \leq 8m\hat{\sigma}_n \leq \varepsilon, \tag{4.11}$$

and so $\operatorname{dist}(\mathcal{M}, V_n) \leq \varepsilon$. We now check the case $n = m^2$. Since by assumption, the solution map belongs to $\mathcal{A}^r$, we know that

$$d_k(\mathcal{M}) \leq \|u\|_{\mathcal{A}^r} \max\{1,k\}^{\ r} \leq M_0 \max\{1,k\}^{\ r}, \quad k \geq 0. \tag{4.12}$$

From the estimates on the performance of the weak greedy algorithm given in Theorem 2.1, we thus know that

$$\sigma_n \leq 2^{4r+1}64m^2M_0n^{\ r} = 2^{4r+2}32m^{2\ 2r}M_0 \leq \varepsilon, \tag{4.13}$$

where we have used the product of the two conditions in (4.2). Hence at step $n = m^2$ we have $\operatorname{dist}(\mathcal{M}, V_n) \leq \varepsilon$. Therefore, we have completed the proof of (i).

We next prove (ii). So assume that $d_n(\mathcal{M}) \leq C_0 \max\{1,n\}^{\ s}$, for some $s > 0$. Then, according to Theorem 2.1,

$$\operatorname{dist}(\mathcal{M}, V_n) = \sigma_n \leq 2^{4s+1}\gamma^{\ 2}C_0n^{\ s} \leq 2^{4s+1}64m^2C_0n^{\ s}, \quad n \geq 1. \tag{4.14}$$

It follows that the numerical algorithm will terminate at a $n(\varepsilon)$ with $n(\varepsilon) \leq n$ where n is the smallest integer that satisfies

$$2^{4s+1}64m^2C_0n^{\ s} \leq \frac{\varepsilon}{8m}. \tag{4.15}$$

Therefore,

$$n(\varepsilon) \leq 2^{4+\frac{10}{s}}C_0^{1/s}m^{\frac{3}{s}}\varepsilon^{\ \frac{1}{s}}. \tag{4.16}$$

Using the first condition in (4.2), this leads to the estimate (4.7) with multiplicative constant $C := 2^{4+\frac{10}{s}}C_0^{1/s}(32M_0)^{\frac{3}{s\ r\ 2)}}$.

Finally, to execute the algorithm, we will need to draw $n(\varepsilon)$ sets $(\widetilde{Y}_0, \ldots, \widetilde{Y}_{n\ 1})$, each of them of size N. The total number of error bound evaluation is thus

$$N(\varepsilon) = n(\varepsilon)N. \tag{4.17}$$

From the definition of N in (4.3) we derive that

$$N \leq 1 + \ \ln\ 1 \quad \frac{3}{4m^2}\Big)\Big)^{\ 1}(\ln|\eta| + 2\ln\ m) \leq Cm^2(|\ln\ \eta| + \ln\ m). \tag{4.18}$$

Using the first condition in the definition (4.2) of m, this leads to

$$N \leq C\varepsilon^{\ \frac{2}{r\ 2)}}(|\ln\eta| + |\ln\varepsilon|), \tag{4.19}$$

where C depends on r and M_0. Combining this with (4.7), we obtain (4.8), which concludes the proof of (ii).

Remark 4.3. The above theorem can be improved by sampling according the tensor product Chebychev measure (3.24), in view of Remark 3.3. Here, we require that the solution map $y \mapsto u(y)$ belongs to $\mathcal{A}^r$ for some $r > 2 \quad := \frac{\ln 3}{\ln 2}$. We then define m as the smallest integer such that

$$32M_0 m^{\ r+2} \leq \varepsilon \quad \text{and} \quad 2^{4r+2} m^{\ (2 \quad 1)r} \leq 1, \tag{4.20}$$

and N as the smallest integer such that

$$1 \quad \frac{3}{4m^2}\Big)^N \leq \frac{\eta}{m^2}\cdot \tag{4.21}$$

We terminate the algorithm at the smallest integer $n \leq m^2 \quad$ for which $\hat{\sigma}_n \leq \frac{\varepsilon}{8m}$. With the exact same proof, we reach the statement as Theorem 4.1, however with a number of step

$$n(\varepsilon) \leq C\varepsilon^{\ (\frac{1}{s}+\frac{3}{s\ r\ \ 2\ )}}, \tag{4.22}$$

and a number of error bound evaluations

$$N(\varepsilon) = C\varepsilon^{\ \frac{2s\ +r+}{s\ r\ \ 2\ )}}(|\ln\eta| + |\ln\varepsilon|). \tag{4.23}$$

Let us comment on the difference in performance between the above algorithm using random sampling and the greedy algorithm based on using an ε-net for the solution manifold as a training set. We aim at a target accuracy ε, and assume that the n-widths of the solution manifold decay like $d_n(\mathcal{M}) \leq Cn^{\ s}$.

Then, the approach based on an ε-net constructs a reduced basis space V_n of optimal dimension

$$n = n(\varepsilon) \sim \varepsilon^{\ 1/s}, \tag{4.24}$$

and the total number of error bound evaluations is at best of the order

$$N(\varepsilon) \sim n(\varepsilon)2^{c\varepsilon^{\ 1/s}}, \tag{4.25}$$

each of them having a cost $\mathrm{Poly}(n) = \mathrm{Poly}(\varepsilon^{\ 1})$, resulting in a prohibitive offline cost $\mathrm{Poly}(\varepsilon^{\ 1})2^{c\varepsilon^{\ 1/s}}$. Note that one should also add the cost of evaluating the $n(\varepsilon)$ reduced basis functions on the finite element space of large dimension $n_h \gg 1$.

In contrast, the approach based on random sampling constructs a reduced basis space V_n of sub-optimal dimension

$$n(\varepsilon) \leq C\varepsilon^{\ (\frac{1}{s}+\frac{3}{s\ r\ \ 2)}}, \tag{4.26}$$

but the total number of error bound evaluation is now of the order

$$N(\varepsilon) \sim \varepsilon^{\ \frac{2s+r+1}{s\ r\ \ 2)}}(|\ln\eta| + |\ln\varepsilon|), \tag{4.27}$$

where η is the probability of failure.

In summary, while our approach allows for a dramatic reduction in the offline cost, it comes with a loss of optimality in the performance of reduced basis spaces since $n(\varepsilon)$ scales with $\varepsilon^{\ 1}$ with an exponent larger than $\frac{1}{s}$. In particular, this affects the resulting online cost. Inspection of the proof of the main theorem reveals that this loss comes from the fact that the greedy selection from the random set can only be identified with a weak-greedy algorithm with a parameter $\gamma = \frac{1}{8m}$ which instead of being fixed becomes small as m grows, or equivalently as ε decreases. Let us still observe that the above perturbation of $\frac{1}{s}$ by $\frac{3}{s(r\ \ 2)}$ becomes negligible as r gets larger. We have also seen that this perturbation can be reduced to $\frac{3}{s(r\ \ 2\ )}$ by sampling randomly according to the Chebychev measure.

This leaves open the question of finding a sampling strategy for the training set which leads to reduced basis of optimal complexity $n(\varepsilon) \sim \varepsilon^{\ 1/s}$ and where the number of error bound evaluation in the offline stage remains polynomial in $\varepsilon^{\ 1}$.

5. Numerical illustration

The results that we have obtained in the previous section can be rephrased in the following way: a polynomial rate of decay of the error achieved by the greedy algorithm in terms of reduced basis space dimension n can be maintained when using a random training set of cardinality N that scales polynomially in n. More precisely, in view of (4.26) and (4.27), a sufficient scaling is

$$N \sim n^{\beta^*}, \quad \beta^* = \frac{2s+r+1}{s(r \quad 2)} \left(\frac{1}{s} + \frac{3}{s(r \quad 2)} \right)^{1} \tag{5.1}$$

independently of the parametric dimension d. Note that we obviously have

$$\beta^* \geq \frac{2s+r+1}{3} \geq \frac{3+2s}{3} > 1.$$

In this section, we illustrate these findings through the following numerical test: we consider the elliptic diffusion equation (1.7) set on the square domain $D =]0,1[^2$, and with affine parametrization (1.8) where $\bar{a} = 1 + \delta$ for some $\delta > 0$ and the ψ_j are given by

$$\psi_j = a_j \chi_{D_j} \tag{5.2}$$

where $\{D_1, \ldots, D_d\}$ is a uniform partition of D into $d = k \times k$ squares of equal size. We take $k = 8$ and therefore $d = 64$, and take

$$a_j = j^{\quad t}, \tag{5.3}$$

for some $t > 0$. The effect of taking t larger is to raise the anisotropy of the parametric dependence. The results from [1] show that this is directly reflected by a rate $n^{\quad r}$ of sparse polynomial approximation of the parameter to solution map, for any $r < t \quad \frac{1}{2}$, and therefore on the rate of decay of the n-width $d_n(\mathcal{M})$. The effect of taking δ closer to 0 is to make the problem more degenerate as y_1 gets close to $\quad 1$.

We test the performance σ_n of the reduced basis spaces generated by the greedy algorithm using random training sets $\widetilde{\mathcal{M}}$ of cardinality

$$N = N(n) = \lfloor n^{\beta} \rfloor, \tag{5.4}$$

for some $\beta \geq 1$. Note that the case $\beta = 1$ amounts in selecting the n reduced basis element completely at random since all the elements from $\widetilde{\mathcal{M}}$ are necessarily chosen. We expect the performance to improve as β becomes larger since we then perform a particular selection of the n reduced basis elements within $\mathcal{M}$ through the greedy algorithm. On the other hand, our theoretical results indicate that a fixed value $\beta = \beta^*$ should be sufficient to ensure that the algorithm performs almost as good as if the selection process took place on the whole of $\mathcal{M}$.

In our numerical test, we took $\delta = 10^{\quad 2}$, that is $\bar{a} = 1.01$. As to the parametric dimension, we considered $d = 16$ and $d = 64$ that correspond to subdivisions of D into 4×4 and 8×8 subdomains, respectively. As to the decay of a_j, we test the values $t = 1$ and $t = 2$. Finally for the growth of the training sample size $N(n)$, we test the values

$$\beta = 1, 1.25, 1.5, 1.75, 2. \tag{5.5}$$

The error curves of the reduced basis approximation, averaged over 20 realizations of the random training sample, for the various choices of d, t and β, are displayed on Figures 1 and 2. The reduced basis approximation error has been computed by using separate test sample of cardinality similar to that of the training sample.

As predicted by the theoretical approximation results, the convergence rate of the reduced basis method improves as t gets larger. The observed convergence rates (for the highest value of β) are closer to the value t than to $t \quad 1/2$ which can be rigourously established from the polynomial approximation rates. This reflects the fact that the reduced basis approximations perform generally better than sparse polynomial approximations.

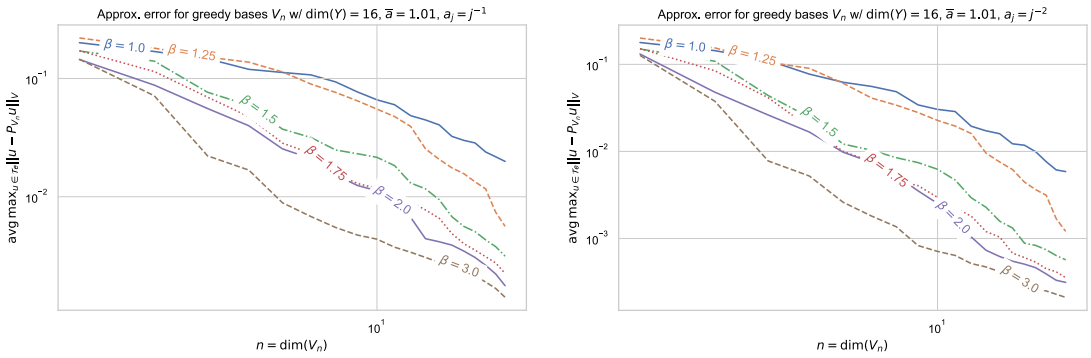

FIGURE 1. Reduced basis approximation with $d = 16$ and decay rate $t = 1$ or $t = 2$.

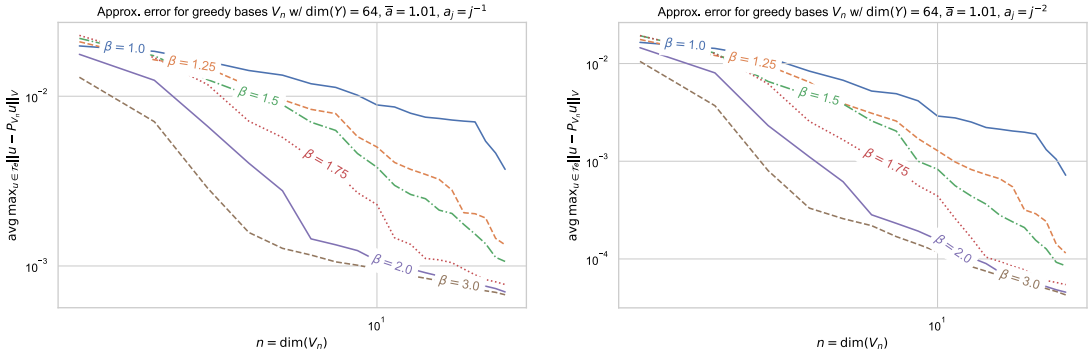

FIGURE 2. Reduced basis approximation with $d = 64$ and decay rate $t = 1$ or $t = 2$.

We also note that, for the same value of t, the errors are smaller in the higher parametric dimension $d = 64$ than for $d = 16$. This apparent paradox can be explained: in both cases, the most active variables are the first ones (y_1, then y_2, ...) yet they are associated to domains D_j of smaller size in the high dimensional case, therefore having less impact on the variation of the solution with these variables.

As expected, we observe that the error curve behaves better as we increase the value of β but we observed that this phenomenon stagnates at $\beta = 2$. This hints that the scaling $N(n) = n^2$ is practically sufficient to ensure in this case the optimal convergence behaviour which would be met with a very rich training set, for example an ε-net. The value $\beta = 2$ is much smaller than the value β^* given by the theoretical analysis which is thus too pessimistic.

Acknowledgements This research was supported by the ONR Contracts N00014-11-1-0712, N00014-12-1-0561, and N00014-15-1-2181; the NSF Grants DMS 1222715, DMS 1720297, DMS 1222390, and DMS 1521067; the Institut Universitaire de France; The European Research Council under grant ERC AdG 338977 BREAD; and by the SmartState and Williams-Hedberg Foundation.

REFERENCES

[1] M. Bachmayr, A. Cohen and G. Migliorati, Sparse polynomial approximation of parametric elliptic PDEs. Part I: affine coefficients. *ESAIM: M2AN* **51** (2017) 321–339.

[2] 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** (2011) 1457–1472.

[3] A. Bu a, Y. Maday, A.T. Patera, C. Prud'homme and G. Turinici, *A priori* convergence of the Greedy algorithm for the parameterized reduced basis. *ESAIM: M2AN* **46** (2012) 595–603.

[4] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile and R. Tempone, Discrete least squares polynomial approximation with random evaluations – application to parametric and stochastic elliptic PDEs. *ESAIM: M2AN* **49** (2015) 815–837.

[5] A. Chkifa, A. Cohen and C. Schwab, Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs. *J. Math. Pures Appl.* **1 3** (2015) 400–428.

[6] A. Cohen and R. DeVore, Approximation of high-dimensional PDEs. *Acta Numer.* **24** (2015) 1–159.

[7] A. Cohen, R. DeVore and C. Schwab, Analytic regularity and polynomial approximation of parametric and stochastic PDEs. *Anal. Appl.* **9** (2011) 11—47.

[8] A. Cohen and G. Migliorati, Multivariate approximation in downward closed polynomial spaces. In: Contemporary Computational Mathematics – A Celebration of the 80th Birthday of Ian Sloan. Springer, Cham (2018) 233–282.

[9] R. DeVore, G. Petrova and P. Wojtaszczyk, Greedy algorithms for reduced bases in Banach spaces. *Constr. Approx.* **37** (2013) 455–466.

[10] Y. Maday, A.T. Patera and G. Turinici, *A priori* convergence theory for reduced-basis approximations of single-parametric elliptic partial di erential equations. *J. Sci. Comput.* **17** (2002) 437–446.

[11] Y. Maday, A.T. Patera and G. Turinici, Global *a priori* convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial di erential equations. *C. R. Acad. Sci. Paris Ser. I Math.* **335** (2002) 289–294.

[12] A.T. Patera, C. Prudhomme, D.V. Rovas and K. Veroy, *A posteriori* error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial di erential equations. In: *Proc. 16th AIAA Computational Fluid Dynamics Conference* (2003).

[13] G. Pisier, The Volume of Convex Bodies and Banach Spaces Geometry. Cambridge University Press (1989).

[14] G. Rozza, D.B.P. Huynh and A.T. Patera, Reduced basis approximation and *a posteriori* error estimation for affinely parametrized elliptic coercive partial di erential equations application to transport and continuum mechanics. *Arch. Comput. Methods Eng.* **15** (2008) 229–275.

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

[16] V.N. Temlyakov, The Marcinkiewicz-type discretization theorems. *Constr. Approx.* **48** (2018) 337–369.