

# HOW TO BEST SAMPLE A SOLUTION MANIFOLD?

WOLFGANG DAHMEN

ABSTRACT. Model reduction attempts to guarantee a desired “model quality”, e.g. given in terms of accuracy requirements, with as small a model size as possible. This article highlights some recent developments concerning this issue for the so called Reduced Basis Method (RBM) for models based on parameter dependent families of PDEs. In this context the key task is to sample the *solution manifold* at judiciously chosen parameter values usually determined in a *greedy fashion*. The corresponding *space growth* concepts are closely related to so called *weak greedy* algorithms in Hilbert and Banach spaces which can be shown to give rise to convergence rates comparable to the best possible rates, namely the *Kolmogorov  $n$ -widths* rates. Such algorithms can be interpreted as *adaptive sampling* strategies for approximating compact sets in Hilbert spaces. We briefly discuss the results most relevant for the present RBM context. The applicability of the results for weak greedy algorithms has however been confined so far essentially to well-conditioned coercive problems. A critical issue is therefore an extension of these concepts to a wider range of problem classes for which the conventional methods do not work well. A second main topic of this article is therefore to outline recent developments of RBMs that do realize  $n$ -width rates for a much wider class of variational problems covering indefinite or singularly perturbed unsymmetric problems. A key element in this context is the design of *well-conditioned variational formulations* and their numerical treatment via saddle point formulations. We conclude with some remarks concerning the relevance of uniformly approximating the whole solution manifold also when the *quantity of interest* is only of a *functional* of the parameter dependent solutions.

## 1. INTRODUCTION

Many engineering applications revolve around the task of identifying a configuration that in some sense best fits certain objective criteria under certain constraints. Such design or optimization problems typically involve (sometimes many) *parameters* that need to be chosen so as to satisfy given optimality criteria. An optimization over such a parameter domain usually requires a frequent evaluation of the states under consideration which typically means to frequently solve a *parameter dependent* family of operator equations

$$(1.1) \quad B_y u = f, \quad y \in \mathcal{Y}.$$

In what follows the parameter set  $\mathcal{Y}$  is always assumed to be a compact subset of  $\mathbb{R}^p$  for some fixed  $p \in \mathbb{N}$  and  $B_y$  should be thought of as a (linear) partial differential operator whose coefficients depend on the parameters  $y \in \mathcal{Y}$ . Moreover,  $B_y$  is viewed as an operator taking some Hilbert space  $U$  one-to-one and onto the *normed dual*  $V'$  of some (appropriate) Hilbert space  $V$  where  $U$  and  $V$  are identified through a variational formulation of (1.1) as detailed

---

1991 *Mathematics Subject Classification.* 65J10, 65N12, 65N15, 35B30.

*Key words and phrases.* Tight surrogates, stable variational formulations, saddle point problems, double greedy schemes, greedy stabilization, rate-optimality, transport equations, convection-diffusion equations.

This work has been supported in part by the DFG SFB-Transregio 40, and by the DFG Research Group 1779, the Excellence Initiative of the German Federal and State Governments, and and NSF grant DMS 1222390.

later, see for instance (4.2). Recall also that the normed dual  $V'$  is endowed with the norm

$$(1.2) \quad \|w\|_{V'} := \sup_{v \in V, v \neq 0} \frac{\langle w, v \rangle}{\|v\|_V},$$

where  $\langle \cdot, \cdot \rangle$  denotes the dual pairing between  $V$  and  $V'$ .

Given a parametric model (1.1) the above mentioned design or optimization problems concern now the *states*  $u(y) \in U$  which, as a function of the parameters  $y \in \mathcal{Y}$ , form what we refer to as the *solution manifold*

$$(1.3) \quad \mathcal{M} := \{u(y) := B_y^{-1} f : y \in \mathcal{Y}\}.$$

Examples of (1.1) arise, for instance, in geometry optimization when a transformation of a variable finitely parametrized domain to a reference domain introduces parameter dependent coefficients of the underlying partial differential equation (PDE) over such domains, see e.g. [14]. Parameters could describe conductivity, viscosity or convection directions, see e.g. [10, 23, 24]. As an extreme case, parametrizing the random diffusion coefficients in a stochastic PDE e.g., by Karhunen-Loew or polynomial chaos expansions, leads to a deterministic parametric PDE involving, in principle, even *infinitely* many parameters,  $p = \infty$ , see e.g. [8] and the literature cited there. We will, however, not treat this particular problem class here any further since, as will be explained later, it poses different conceptual obstructions than those in the focus of this paper, namely the absence of ellipticity which makes conventional strategies fail. In particular, we shall explain why for other relevant problem classes, e.g. those dominated by transport processes,  $\mathcal{M}$  is not “as visible” as for elliptic problems and how to restore “full visibility”.

**1.1. General Context - Reduced Basis Method.** A conventional way of searching for a specific state in  $\mathcal{M}$  or optimize over  $\mathcal{M}$  is to compute approximate solutions of (1.1) possibly for a large number of parameters  $y$ . Such approximations would then reside in a sufficiently large trial space  $U_{\mathcal{N}} \subset U$  of dimension  $\mathcal{N}$ , typically a finite element space. Ideally one would try to assure that  $U_{\mathcal{N}}$  is large enough to warrant sufficient accuracy of whatever conclusions are to be drawn from such a discretization. A common terminology in reduced order modeling refers to  $U_{\mathcal{N}}$  as the *truth space* providing accurate computable information. Of course, each such parameter query in  $U_{\mathcal{N}}$  is a computationally expensive task so that many such queries, especially in an online context, would be practically infeasible. On the other hand, solving for each  $y \in \mathcal{Y}$  a problem in  $U_{\mathcal{N}}$  would just treat each solution  $u(y)$  as some “point” in the infinite dimensional space  $U$ , viz. in the very large finite dimensional space  $U_{\mathcal{N}}$ . This disregards the fact that all these points actually belong to a possibly much thinner and more coherent set, namely the low dimensional manifold  $\mathcal{M}$  which, for compact  $\mathcal{Y}$  and well posed problems (1.1), is compact. Moreover, if the solutions  $u(y)$ , as functions of  $y \in \mathcal{Y}$ , depend smoothly on  $y$  there is hope that one can approximate all elements of  $\mathcal{M}$  uniformly over  $\mathcal{Y}$  with respect to the Hilbert space norm  $\|\cdot\|_U$  by a relatively small but judiciously chosen linear space  $U_n$ . Here “small” means that  $n = \dim U_n$  is significantly smaller than  $\mathcal{N} = \dim U_{\mathcal{N}}$ , often by orders of magnitude. As detailed later the classical notion of *Kolmogorov  $n$ -widths* quantifies how well a compact set in a Banach space can be approximated in the corresponding Banach norm by a linear space and therefore can be used as a *benchmark* for the effectiveness of a model reduction strategy.

Specifically, the core objective of the *Reduced Basis Method* (RBM) is to find for a given *target accuracy*  $\varepsilon$  a possibly small number  $n = n(\varepsilon)$  of basis functions  $\phi_j, j = 0, \dots, n$ , whose linear combinations approximate each  $u \in \mathcal{M}$  within accuracy at least  $\varepsilon$ . This means that

ideally for each  $y \in \mathcal{Y}$  one can find coefficients  $c_j(y)$  such that the expansion

$$(1.4) \quad u_n(x, y) := \sum_{j=0}^{n(\varepsilon)} c_j(y) \phi_j(x)$$

satisfies

$$(1.5) \quad \|u(y) - u_n(y)\|_U \leq \varepsilon, \quad y \in \mathcal{Y}.$$

Thus, projecting (1.1) into the small space  $U_n := \text{span}\{\phi_0, \dots, \phi_n\}$  reduces each parameter query to solving a small  $n \times n$  system of equations where typically  $n \ll \mathcal{N}$ .

**1.2. Goal Orientation.** Recall that the actual goal of reduced modeling is often not to recover the full fields  $u(y) \in \mathcal{M}$  but only some *quantity of interest*  $I(y)$  typically given as a *functional*  $I(y) := \ell(u(y))$  of  $u(y)$  where  $\ell \in U'$ . Asking just the value of such a functional is possibly a weaker request than approximating all of  $u(y)$  in the norm  $\|\cdot\|_U$ . In other words, one may have  $|\ell(u(y)) - \ell(u_n(y))| \leq \varepsilon$  without insisting on the validity of (1.5) for a tolerance of roughly the same size. Of course, one would like to exploit this in favor of online efficiency. Duality methods as used in the context of *goal-oriented* finite element methods [3] are indeed known to offer ways of economizing the approximate evaluation of functionals. Such concepts apply in the RBM context as well, see e.g. [21, 16]. However, as we shall point out later, guaranteeing that  $|\ell(u(y)) - \ell(u_n(y))| \leq \varepsilon$  holds for  $y \in \mathcal{Y}$ , ultimately reduces to tasks of the type (1.5) as well. So, in summary, understanding how to ensure (1.5) for possibly small  $n(\varepsilon)$  remains the core issue and therefore guides the subsequent discussions.

Postponing for a moment the issue of how to actually compute the  $\phi_j$ , it is clear that they should intrinsically depend on  $\mathcal{M}$  rendering the whole process highly nonlinear. To put the above approach first into perspective, viewing  $u(x, y)$  as a function of the spatial variables  $x$  and of the parameters  $y$ , (1.4) is just *separation of variables* where the factors  $c_j(y)$ ,  $\phi_j(x)$  are a priori unknown. It is perhaps worth stressing though that, in contrast to other attempts to find good *tensor approximation*, in the RBM context explicit representations are only computed for the spatial factors  $\phi_j$  while for each  $y$  the weight  $c_j(y)$  has to be *computed* by solving a small system in the reduced space  $U_n$ . Thus, the computation of  $\{\phi_0, \dots, \phi_{n(\varepsilon)}\}$  could be interpreted as *dictionary learning* and, loosely speaking,  $n = n(\varepsilon)$  being relatively small for a given target accuracy, means that all elements in  $\mathcal{M}$  are *approximately sparse* with respect to the dictionary  $\{\phi_0, \dots, \phi_n, \dots\}$ .

The methodology just outlined has been pioneered by Y. Maday, T.A. Patera and collaborators, see e.g. [6, 21, 24, 23]. As indicated before, RBM is one variant of a *model order reduction* paradigm that is specially tailored to parameter dependent problems. Among its distinguishing constituents one can name the following. There is usually a careful division of the overall computational work into an *offline phase*, which could be computationally intense but should remain manageable, and an *online phase* which should be executable with highest efficiency taking advantage of a precomputed basis and matrix assemblations during the offline phase. It is important to note that while the offline phase is accepted to be computationally expensive it should remain *offline-feasible* in the sense that a possibly extensive search over the parameter domain  $\mathcal{Y}$  in the offline phase requires for each query solving only problems in the small reduced space. Under which circumstances this is possible and how to realize such division concepts has been worked out in the literature, see e.g. [24, 23]. Here we are content with stressing that an important role is played by the way how the operator  $B_y$  depends on the parameter  $y$ , namely in an *affine* way as stated in (2.13) later below. Second, and this is

perhaps the most distinguishing constituent, along with each solution in the reduced model one strives to provide a *certificate* of accuracy, i.e., computed bounds for incurred error tolerances [24, 23].

**1.3. Central Objectives.** When trying to quantify the performance of such methods aside from the above mentioned structural and data organization aspects, among others, the following questions come to mind:

(i) for which type of problems do such methods work very well in the sense that  $n(\varepsilon)$  in (1.5) grows only slowly when  $\varepsilon$  decreases? This concerns quantifying the sparsity of solutions.

(ii) How can one compute reduced bases  $\{\phi_0, \dots, \phi_{n(\varepsilon)}\}$  for which  $n(\varepsilon)$  is *nearly minimal* in a sense to be made precise below?

Of course, the better the sparsity quantified by (i) the better could be the pay-off of an RBM. However, as one may expect, an answer to (i) depends strongly on the problem under consideration. This is illustrated also by the example presented in §5.4. Question (ii), instead, can be addressed independently of (i) in the sense that, no matter how many basis functions have to be computed in order to meet a given target accuracy, can one come up with methods that guarantee generating a *nearly minimal number* of such basis functions? This has to do with *how to sample* the solution manifold and is the central theme in this paper.

The most prominent way of generating the reduced bases is a certain *greedy sampling* of the manifold  $\mathcal{M}$ . Contriving *greedy sampling strategies* that give rise to reduced bases of nearly minimal length, in a sense to be made precise below, also for *non-coercive or unsymmetric singularly perturbed problems* is the central objective in this paper. We remark though that a greedy parameter search in its standard form is perhaps not suitable for very high dimensional parameter spaces without taking additional structural features of the problem into account. The subsequent discussions do therefore not target specifically the large amount of recent work on stochastic elliptic PDEs, since while greedy concepts are in principle well understood for elliptic problems they are per se not necessarily adequate for infinitely many parameters without exploiting specific problem dependent structural information.

First, we recall in §2 a *greedy space growth* paradigm commonly used in all established RBMs. To measure its performance in the sense of (ii) we follow [6] and compare the corresponding distances  $\text{dist}_U(\mathcal{M}, U_n)$  to the smallest possible distances achievable by linear spaces of dimension  $n$ , called *Kolmogorov  $n$ -widths*. The fact that for *elliptic problems* the convergence rates for the greedy errors are essentially those of the  $n$ -width, and hence *rate-optimal*, is shown in §3 to be ultimately reduced to analyzing so called *weak greedy algorithms* in Hilbert spaces, see also [4, 13]. However, for indefinite or strongly unsymmetric and singularly perturbed problems this method usually operates far from optimality. We explain why this is the case and describe in §4 a remedy proposed in [10]. A pivotal role is played by certain *well-conditioned variational formulations* of (1.1) which are then shown to lead again to an optimal *outer greedy* sampling strategy also for non-elliptic problems. An essential additional ingredient consists of certain stabilizing *inner greedy loops*, see §5. The obtained rate-optimal scheme is illustrated by a numerical example addressing convection dominated convection-diffusion problems in §5.4. We conclude in §6 with applying these concepts to the efficient evaluation of quantities of interest.

## 2. THE GREEDY PARADIGM

The by far most prominent strategy for constructing reduced bases for a given parameter dependent problem (1.1) is the following greedy procedure, see e.g. [23]. The basic idea is that, having already constructed a reduced space  $U_n$  of dimension  $n$ , find an element  $u_{n+1} =$

$u(y_{n+1})$  in  $\mathcal{M}$  that is farthest away from the current space  $U_n$ , i.e., that maximizes the best approximation error from  $U_n$  and then grow  $U_n$  by setting  $U_{n+1} := U_n + \text{span}\{u_{n+1}\}$ . Hence, denoting by  $P_{U,U_n}$  the  $U$ -orthogonal projection onto  $U_n$ ,

$$(2.1) \quad y_{n+1} := \operatorname{argmax}_{y \in \mathcal{Y}} \|u(y) - P_{U,U_n} u(y)\|_U, \quad u_{n+1} := u(y_{n+1}).$$

Unfortunately, determining such an exact maximizer is computationally way too expensive even in an offline phase because one would have to compute for a sufficiently dense sampling of  $\mathcal{Y}$  the exact solution  $u(y)$  of (1.1) in  $U$  (in practice in  $U_{\mathcal{N}}$ ). Instead one tries to construct more efficiently computable *surrogates*  $R(y, U_n)$  satisfying

$$(2.2) \quad \|u(y) - P_{U,U_n} u(y)\|_U \leq R(y, U_n), \quad y \in \mathcal{Y}.$$

Recall that “efficiently computable” in the sense of offline-feasibility means that for each  $y \in \mathcal{Y}$ , the surrogate  $R(y, U_n)$  can be evaluated by solving only a problem of size  $n$  in the reduced space  $U_n$ . Deferring an explanation of the nature of such surrogates, Algorithm 1 described below is a typical offline-feasible *surrogate based greedy algorithm* (SGA). Clearly, the maximizer in (2.3) below is not necessarily unique. In case several maximizers exist it does not matter which one is selected.

---

**Algorithm 1** surrogate based greedy algorithm

---

```

1: function SGA
2:   Set  $U_0 := \{0\}$ ,  $n = 0$ ,
3:   while  $\operatorname{argmax}_{y \in \mathcal{Y}} R(y, U_n) \geq \text{tol}$  do
4:      $y_{n+1} := \operatorname{argmax}_{y \in \mathcal{Y}} R(y, U_n)$ ,
(2.3)    $u_{n+1} := u(y_{n+1})$ ,
        $U_{n+1} := \text{span}\{U_n, \{u(y_{n+1})\}\} = \text{span}\{u_1, \dots, u_{n+1}\}$ 
5:   end while
6: end function

```

---

Strictly speaking, the scheme SGA is still idealized since:

- (a) computations cannot be carried out in  $U$ ;
- (b) one cannot parse through all of a continuum  $\mathcal{Y}$  to maximize  $R(y, U_n)$ .

Concerning (a), as mentioned earlier computations in  $U$  are to be understood as synonymous to computations in a sufficiently large truth space  $U_{\mathcal{N}}$  satisfying all targeted accuracy tolerances for the underlying application. Solving problems in  $U_{\mathcal{N}}$  is strictly confined to the offline phase and the number of such solves should remain of the order of  $n = \dim U_n$ . We will not distinguish in what follows between  $U$  and  $U_{\mathcal{N}}$  unless such a distinction matters.

As for (b), the maximization is usually performed with the aid of a complete search over a *discrete* subset of  $\mathcal{Y}$ . Again, we will not distinguish between a possibly continuous parameter set and a suitable training subset. In fact, continuous optimization methods that would avoid a complete search have so far not proven to work well since each greedy step increases the number of local maxima of the objective functional. Now, how fine such a discretization for a complete search should be depends on how smoothly the  $u(y)$  depend on  $y$ . But even when

such a dependence is very smooth a coarse discretization of a *high-dimensional* parameter set  $\mathcal{Y}$  would render a complete search infeasible so that, depending on the problem at hand, one has to resort to alternate strategies such as, for instance, random sampling. However, since it seems that (b) can only be answered for a specific problem class we will not address this issue in this paper any further.

Instead, we focus on general principles which guarantee the following. Loosely speaking the reduced spaces based on sampling  $\mathcal{M}$  should perform *optimally* in the sense that the resulting spaces  $U_n$  have the (near) “smallest dimension” needed to satisfy a given target tolerance while the involved offline and online cost remains feasible in the sense indicated above. To explain first what is meant by “optimal” let us denote the *greedy error* produced by SGA as

$$(2.4) \quad \sigma_n(\mathcal{M})_U := \max_{v \in \mathcal{M}} \inf_{\bar{u} \in U_n} \|v - \bar{u}\|_U = \max_{y \in \mathcal{Y}} \|u(y) - P_{U,U_n} u(y)\|_U.$$

Note that if we replaced in (2.4) the space  $U_n$  by *any* linear subspace  $W_n \subset U$  and infimize the resulting distortion over *all* subspaces of  $U$  of dimension at most  $n$ , one obtains the classical *Kolmogorov  $n$ -widths*  $d_n(\mathcal{M})_U$  quantifying the “thickness” of a compact set, see (3.2). One trivially has

$$(2.5) \quad d_n(\mathcal{M})_U \leq \sigma_n(\mathcal{M})_U, \quad n \in \mathbb{N}.$$

Of course, it would be best if one could reverse the above inequality. We will discuss in the next section to what extent this is possible.

To prepare for such a discussion we need more information about how the surrogate  $R(y, U_n)$  relates to the actual error  $\|u(y) - P_{U,U_n} u(y)\|_U$  because the surrogate drives the greedy search and one expects that the quality of the snapshots found in SGA depends on how “tight” the upper bound in (2.2) is.

To identify next the essential conditions on a “good” surrogate it is instructive to consider the case of *elliptic* problems. To this end, suppose that

$$\langle B_y u, v \rangle = b_y(u, v) = \langle f, v \rangle, \quad u, v \in U,$$

is a uniformly  $U$ -coercive bounded bilinear form and  $f \in U'$ , i.e., there exist constants  $0 < c_1 \leq C_1 < \infty$  such that

$$(2.6) \quad c_1 \|v\|_U^2 \leq b_y(v, v), \quad |b_y(u, v)| \leq C_1 \|u\|_U \|v\|_U, \quad u, v \in U, \quad y \in \mathcal{Y},$$

holds uniformly in  $y \in \mathcal{Y}$ . The operator equation (1.1) is then equivalent to: given  $f \in U'$  and a  $y \in \mathcal{Y}$ , find  $u(y) \in U$  such that

$$(2.7) \quad b_y(u(y), v) = \langle f, v \rangle, \quad v \in U.$$

Ellipticity has two important well-known consequences. First, since (2.6) implies  $\|B_y\|_{U \rightarrow U'} \leq C_1$ ,  $\|B_y^{-1}\|_{U' \rightarrow U} \leq c_1^{-1}$  the operator  $B_y : U \rightarrow U'$  has a finite condition number

$$(2.8) \quad \kappa_{U,U'}(B_y) := \|B_y\|_{U \rightarrow U'} \|B_y^{-1}\|_{U' \rightarrow U} \leq C_1/c_1$$

which, in particular, means that residuals in  $U'$  are uniformly comparable to errors in  $U$

$$(2.9) \quad C_1^{-1} \|f - B_y \bar{u}\|_{U'} \leq \|u(y) - \bar{u}\|_U \leq c_1^{-1} \|f - B_y \bar{u}\|_{U'}, \quad \bar{u} \in U, \quad y \in \mathcal{Y}.$$

Second, by Céa’s Lemma, the Galerkin projection  $\Pi_{y,U_n}$  onto  $U_n$  is up to a constant as good as the *best approximation*, i.e., under the assumption (2.6)

$$(2.10) \quad \|u(y) - \Pi_{y,U_n} u(y)\|_U \leq \frac{C_1}{c_1} \inf_{v \in U_n} \|u(y) - v\|_U.$$

(When  $b_y(\cdot, \cdot)$  is in addition symmetric  $C_1/c_1$  can be replaced by  $(C_1/c_1)^{1/2}$ .) Hence, by (2.9) and (2.10),

$$(2.11) \quad R(y, U_n) := c_1^{-1} \sup_{v \in U} \frac{\langle f, v \rangle - b_y(\Pi_{y, U_n} u(y), v)}{\|v\|_U}$$

satisfies more than just (2.2), namely it provides also a uniform *lower bound*

$$(2.12) \quad \frac{c_1}{C_1} R(y, U_n) \leq \|u(y) - P_{U, U_n} u(y)\|_U, \quad y \in \mathcal{Y}.$$

Finally, suppose that  $b_y(\cdot, \cdot)$  depends *affinely* on the parameters in the sense that

$$(2.13) \quad b_y(u, v) = \sum_{k=1}^M \theta_k(y) b_k(u, v),$$

where the  $\theta_k$  are smooth functions of  $y \in \mathcal{Y}$  and the bilinear forms  $b_k(\cdot, \cdot)$  are independent of  $y$ . Then, based on suitable precomputations (in  $U_N$ ) in the offline phase, the computation of  $\Pi_{y, U_n} u(y)$  reduces for each  $y \in \mathcal{Y}$  to the solution of a rapidly assembled  $(n \times n)$ -system, and  $R(y, U_n)$  can indeed be computed very efficiently, see [23, 16, 24].

An essential consequence of (2.2) and (2.12) can be formulated as follows.

**Proposition 2.1.** *Given  $U_n \subset U$ , the function  $u_{n+1}$  generated by (2.3) for  $R(y, U_n)$  defined by (2.11), has the property that*

$$(2.14) \quad \|u_{n+1} - P_{U, U_n} u_{n+1}\|_U \geq \frac{c_1}{C_1} \max_{v \in \mathcal{M}} \min_{\bar{u} \in U_n} \|v - \bar{u}\|_U.$$

Hence, maximizing the residual based surrogate  $R(y, U_n)$  (over a suitable discretization of  $\mathcal{Y}$ ) is a computationally feasible way of determining, up to a fixed factor  $\gamma := c_1/C_1 \leq 1$ , the maximal distance between  $\mathcal{M}$  and  $U_n$  and performs in this sense almost as well as the “ideal” but computationally infeasible surrogate  $R^*(\mu, U_n) := \|u(y) - P_{U, U_n} u(y)\|_U$ .

**Proof of Proposition 2.1:** Suppose that  $\bar{y} = \operatorname{argmax}_{y \in \mathcal{Y}} R(y, U_n)$ ,  $y^* := \operatorname{argmax}_{y \in \mathcal{Y}} \|u(y) - P_{U, U_n} u(y)\|_U$  so that  $u_{n+1} = u(\bar{y})$ . Then, keeping (2.12) and (2.10) in mind, we have

$$\begin{aligned} \|u_{n+1} - P_{U, U_n} u_{n+1}\|_U &= \|u(\bar{y}) - P_{U, U_n} u(\bar{y})\|_U \geq \frac{c_1}{C_1} R(\bar{y}, U_n) \geq \frac{c_1}{C_1} R(y^*, U_n) \\ &\geq \frac{c_1}{C_1} \|u(y^*) - P_{U, U_n} u(y^*)\|_U = \frac{c_1}{C_1} \max_{y \in \mathcal{Y}} \|u(y) - P_{U, U_n} u(y)\|_U, \end{aligned}$$

where we have used (2.2) in the second but last step. This confirms the claim.  $\square$

Property (2.14) turns out to play a key role in the analysis of the performance of the scheme SGA.

### 3. GREEDY SPACE GROWTH

Proposition 2.1 allows us to view the algorithm SGA as a special instance of the following scenario. Given a compact subset  $\mathcal{K}$  of a Hilbert space  $H$  with inner product  $(\cdot, \cdot)$  inducing the norm  $\|\cdot\|^2 = (\cdot, \cdot)$ , consider the *weak greedy* Algorithm 2 (WGA) below.

Note that again the choice of  $u_{n+1}$  is not necessarily unique and what follows holds for *any* choice satisfying (3.1).

Greedy strategies have been used in numerous contexts and variants. The current version is not to be confused though with the *weak orthogonal greedy algorithm* introduced in [25] for

---

**Algorithm 2** weak greedy algorithm
 

---

1: **function** WGA2: Set  $H_0 := \{0\}$ ,  $n = 0$ ,  $u_0 := 0$ , fix any  $0 < \gamma \leq 1$ ,3: given  $H_n$ , choose some  $u_{n+1} \in \mathcal{K}$  for which

$$(3.1) \quad \min_{v_n \in H_n} \|v_n - u_{n+1}\| \geq \gamma \max_{v \in \mathcal{K}} \min_{v_n \in U_n} \|v - v_n\| =: \gamma \sigma_n(\mathcal{K})_H,$$

and set  $H_{n+1} := H_n + \text{span}\{u_{n+1}\}$ .4: **end function**


---

*approximating a function* by a linear combination of  $n$  terms from a *given* dictionary. In contrast, the scheme WGA described in Algorithm 2 aims at *constructing* a (problem dependent) dictionary with the aid of a PDE model. While greedy function approximation is naturally compared with the *best  $n$ -term approximation* from the underlying dictionary (see [2, 25] for related results), a natural question here is to compare the corresponding greedy errors

$$\sigma_n(\mathcal{K})_H := \max_{v \in \mathcal{K}} \min_{v_n \in U_n} \|v - v_n\| =: \max \text{dist}_H(\mathcal{K}, U_n)$$

incurred when approximating a compact set  $\mathcal{K}$  with the smallest possible deviation of  $\mathcal{K}$  from any  $n$ -dimensional linear space, given by the Kolmogorov  $n$ -widths

$$(3.2) \quad d_n(\mathcal{K})_H := \inf_{\dim V = n} \sup_{v \in \mathcal{K}} \inf_{v_n \in V} \|v - v_n\| = \inf_{\dim V = n} \max \text{dist}_H(\mathcal{K}, V),$$

mentioned earlier in the preceding section. One trivially has  $d_n(\mathcal{K})_H \leq \sigma_n(\mathcal{K})_H$  for all  $n \in \mathbb{N}$  and the question arises whether there actually exists a constant  $C$  such that

$$(3.3) \quad \sigma_n(\mathcal{K})_H \leq C d_n(\mathcal{K})_H, \quad n \in \mathbb{N}.$$

One may doubt such a relation to hold for several reasons. First, orthogonal greedy *function approximation* performs in a way comparable to best  $n$ -term approximation only under rather strong assumptions on the underlying given dictionary. Intuitively, one expects that errors made early on in the iteration are generally hard to correct later although this intuition turns out to be misleading in the case of the present *set approximation*. Second, the spaces  $U_n$  generated by the greedy growth are restricted by being generated only from snapshots in  $\mathcal{K}$  while the best spaces can be chosen freely, see the related discussion in [4].

The comparison (3.3) was addressed first in [6] for the ideal case  $\gamma = 1$ . In this case a bound of the form  $\sigma_n(\mathcal{K})_H \leq C n 2^n d_n(\mathcal{K})_H$  could be established for some absolute constant  $C$ . This is useful only for cases where the  $n$ -widths decay faster than  $n^{-1} 2^{-n}$  which indeed turns out to be possible for elliptic problems (2.7) with a sufficiently smooth affine parameter dependence (2.13). In fact, in such a case the  $u(y)$  can be even shown to be *analytic* as a function of  $y$ , see [8] and the literature cited there. It was then shown in [4] that the slightly better bound

$$(3.4) \quad \sigma_n(\mathcal{K})_H \leq \frac{2^{n+1}}{\sqrt{3}} d_n(\mathcal{K})_H, \quad n \in \mathbb{N},$$

holds. More importantly, these bounds cannot be improved in general. Moreover, the possible exponential loss in accuracy is not due to the fact the greedy spaces are generated by snapshots from  $\mathcal{K}$ . In fact, denoting by  $\bar{d}_n(\mathcal{K})_H$  the restricted “inner” widths, obtained by allowing only subspaces spanned by snapshots of  $\mathcal{K}$  in the competition, one can prove that  $\bar{d}_n(\mathcal{K})_H \leq n d_n(\mathcal{K})_H$ ,  $n \in \mathbb{N}$ , which is also sharp in general [4].

While these findings may be interpreted as limiting the use of reduced bases generated in a greedy fashion to problems where the  $n$ -widths decay exponentially fast the situation turns out

to be far less dim if one does *not* insist on a *direct comparison* of the type (3.3) with  $n$  being *the same* on both sides of the inequality. In [4, 13] the question is addressed whether a certain *convergence rate* of the  $n$ -widths  $d_n(\mathcal{K})_H$  implies some convergence rate of the greedy errors  $\sigma_n(\mathcal{K})_H$ . The following result from [4] gave a first affirmative answer.

**Theorem 3.1.** *Let  $0 < \gamma \leq 1$  be the parameter in (3.1) and assume that  $d_0(\mathcal{K})_H \leq M$  for some  $M > 0$ . Then*

$$d_n(\mathcal{K})_H \leq Mn^{-\alpha}, \quad n \in \mathbb{N},$$

for some  $\alpha > 0$ , implies

$$(3.5) \quad \sigma_n(\mathcal{K})_H \leq CMn^{-\alpha}, \quad n > 0,$$

where  $C := q^{\frac{1}{2}}(4q)^\alpha$  and  $q := \lceil 2^{\alpha+1}\gamma^{-1} \rceil^2$ .

This means that the weak greedy scheme may still be highly profitable even when the  $n$ -widths do not decay exponentially. Moreover, as expected, the closer the weakness parameter  $\gamma$  is to one, the better, which will later guide the sampling strategies for constructing reduced bases.

Results of the above type are not confined to polynomial rates. A sub-exponential decay of the  $d_n(\mathcal{K})_H$  with a rate  $e^{-cn^\alpha}$ ,  $\alpha \leq 1$  is shown in [4] to imply a rate

$$(3.6) \quad \sigma_n(\mathcal{K})_H \leq C(\alpha, \gamma)e^{-\tilde{c}n^{\tilde{\alpha}}}, \quad \tilde{\alpha} = \alpha/(1 + \alpha), \quad n \in \mathbb{N}.$$

The principle behind the estimates (3.5), (3.6) is to exploit a “flatness” effect or what one may call “conditional delayed comparison”. More precisely, given any  $\theta \in (0, 1)$  and defining  $q := \lceil 2(\gamma\theta) \rceil^2$ , one can show that ([4, Lemma 2.2])

$$\sigma_{n+qm}(\mathcal{K})_H \geq \theta\sigma_n(\mathcal{K})_H \quad \Rightarrow \quad \sigma_n(\mathcal{K})_H \leq q^{1/2}d_m(\mathcal{K})_H, \quad n \in \mathbb{N}.$$

Thus, a comparison between greedy errors and  $n$ -widths is possible when the greedy errors do not decay too quickly. This is behind the diminished exponent  $\tilde{\alpha}$  in (3.6).

These results have been improved upon in [13] in several ways employing different techniques yielding improved comparisons. Abbreviating  $\sigma_n := \sigma_n(\mathcal{K})_H$ ,  $d_n := d_n(\mathcal{K})_H$ , a central result in the present general Hilbert space context states that for any  $N \geq 0$ ,  $K \geq 1$ ,  $1 \leq m < K$  one has

$$(3.7) \quad \prod_{i=1}^K \sigma_{N+i}^2 \leq \gamma^{-2K} \left(\frac{K}{M}\right)^m \left(\frac{K}{K-m}\right)^{K-m} \sigma_{N+1}^{2m} d_m^{2(K-m)}.$$

As a first important consequence, one derives from these inequalities a nearly direct comparison between  $\sigma_n$  and  $d_n$  without any constraint on the decay of  $\sigma_n$  or  $d_n$ . In fact, taking  $N = 0$ ,  $K = n$ , and any  $1 \leq m < n$  in (3.7), using the monotonicity of the  $\sigma_n$ , one shows that  $\sigma_n^{2n} \leq \gamma^{-2n} \left(\frac{n}{m}\right)^m \left(\frac{n}{n-m}\right)^{n-m} d_m^{2(n-m)}$  from which one deduces

$$(3.8) \quad \sigma_n \leq \sqrt{2}\gamma^{-1} \min_{1 \leq m < n} d_m^{\frac{n-m}{n}}, \quad n \in \mathbb{N}.$$

This, in particular, gives the direct unconditional comparison

$$\sigma_{2n}(\mathcal{K})_H \leq \gamma^{-1} \sqrt{2d_n(\mathcal{K})_H}, \quad n \in \mathbb{N}.$$

The estimate (3.8) is then used in [13] to improve on (3.6) establishing the bounds

$$(3.9) \quad d_n(\mathcal{K})_H \leq C_0 e^{-c_0 n^\alpha}, \quad \Rightarrow \quad \sigma_n(\mathcal{K})_H \leq \sqrt{2C_0} \gamma^{-1} e^{-c_1 n^\alpha}, \quad n \in \mathbb{N},$$

i.e., the exponent  $\alpha$  is preserved by the rate for the greedy errors. Moreover, one can recover (3.5) from (3.7) (with different constants).

Although not needed in the present context the second group of results in [13] should be mentioned that concerns the extension of the weak greedy algorithm WGA to *Banach* spaces  $X$  in place of the Hilbert space  $H$ . Remarkably, a direct comparison between  $\sigma_n(\mathcal{K})_X$  and  $d_n(\mathcal{K})_X$  similar to (3.7) is also established in [13]. The counterpart to (3.8) reads  $\sigma_{2n} \leq 2\gamma^{-1}\sqrt{nd_n}$  i.e., one loses a factor  $\sqrt{n}$  which is shown, however, to be necessary in general.

All the above results show that the smaller the weakness parameter  $\gamma$  the stronger the derogation of the rate of the greedy errors in comparison with the  $n$ -widths.

#### 4. WHAT ARE THE RIGHT PROJECTIONS?

As shown by (3.5) and (3.9), the weak greedy algorithm WGA realizes optimal rates for essentially all ranges of interest. A natural question is under which circumstances a surrogate based greedy algorithm SGA is in this sense also *rate-optimal*, namely ensures the validity of (3.5) and (3.9). Obviously, this is precisely the case when new snapshots generated through maximizing the surrogate have the *weak greedy property* (3.1). Note that Proposition 2.1 says that the *residual based surrogate* (2.11) in the case of *coercive problems* does ensure the weak-greedy property so that SGA is indeed rate-optimal for coercive problems. Note also that the weakness parameter  $\gamma = c_1/C_1$  is in this case the larger the smaller the condition number of the operator  $B_y$  is, see (2.8). Obviously, the key is that the surrogate not only yields an upper bound for the best approximation error but also, up to a constant, a lower bound (2.12), and the more tightly the best approximation error is sandwiched by the surrogate the better the performance of SGA. Therefore, even if the problem is coercive for a very small  $\gamma = c_1/C_1$ , as is the case for convection dominated *convection-diffusion problems*, in view of the dependence of the bounds in (3.5) and (3.9) on  $\gamma^{-1}$ , one expects that the performance of a greedy search based on (2.11) degrades significantly.

In summary, as long as algorithm SGA employs a *tight surrogate* in the sense that

$$(4.1) \quad c_S R(y, U_n) \leq \inf_{v \in U_n} \|u(y) - v\|_U \leq R(y, U_n), \quad y \in \mathcal{Y},$$

holds for some constant  $c_S > 0$ , independent of  $y \in \mathcal{Y}$ , algorithm SGA is *rate-optimal* in the sense of (3.5), (3.9), i.e., it essentially realizes the  $n$ -width rates over all ranges of interest, see [10]. We refer to  $c_S^{-1} := \kappa_n(R)$  as the *condition* of the surrogate  $R(\cdot, U_n)$ . In the RBM community the constant  $c_S^{-1}$  is essentially the *stability factor* which is usually computed along with an approximate reduced solution. Clearly, the bounds in §3 also show that the quantitative performance of SGA is expected to be the better the smaller the condition of the surrogate, i.e., the larger  $c_S$ .

As shown so far, coercive problems with a small condition number  $\kappa_{U,U'}(B_y)$  represent an ideal setting for RBM and standard Galerkin projection combined with the *symmetric surrogate* (2.11), based on measuring the residual in the dual norm  $\|\cdot\|_{U'}$  of the “error-norm”  $\|\cdot\|_U$ , identifies rate-optimal snapshots for a greedy space growth. Of course, this marks a small segment of relevant problems. Formally, one can still apply these projections and surrogates for any variational problem (2.7) for which a residual can be computed. However, in general, for indefinite or unsymmetric singularly perturbed problems, the tightness relation (4.1) may no longer hold for surrogates of the form (2.11) or, if it holds the condition  $\kappa_n(R)$  becomes prohibitively large. In this latter case, the upper bound of the best approximation error is too loose to direct the search for proper snapshots. A simple example is the *convection-diffusion*

problem: for  $f \in (H_0^1(\Omega))'$  find  $u \in H_0^1(\Omega)$ ,  $\Omega \subset \mathbb{R}^d$ , such that

$$(4.2) \quad \varepsilon(\nabla u, \nabla v) + (\vec{b} \cdot \nabla u, v) + (cu, v) =: b_y(u, v) = \langle f, v \rangle, \quad v \in H_0^1(\Omega),$$

where, for instance,  $y = (\varepsilon, \vec{b}) \in \mathcal{Y} := [\varepsilon_0, 1] \times S^{d-1}$ ,  $S^{d-1}$  the  $(d-1)$ -sphere.

**Remark 4.1.** *It is well known that when  $c - \frac{1}{2} \operatorname{div} \vec{b} \geq 0$  the problem (4.2) has for any  $f \in H^{-1}(\Omega) := (H_0^1(\Omega))'$  a unique solution. Thus, for  $U := H_0^1(\Omega)$  (2.6) is still valid but with  $\kappa_{U,U'}(B_y) \sim \varepsilon^{-1}$  which becomes arbitrarily large for a correspondingly small diffusion lower bound  $\varepsilon_0$ .*

The standard scheme SGA indeed no longer performs nearly as well as in the well conditioned case. The situation is even less clear when  $\varepsilon = 0$  (with modified boundary conditions) where no “natural” variational formulation suggests itself (we refer to [10] for a detailed discussion of these examples). Moreover, for *indefinite problems* the Galerkin projection does generally perform like the best approximation which also adversely affects tightness of the standard symmetric residual based surrogate (2.11).

Hence, to retain rate-optimality of SGA also for the above mentioned extended range of problems one has to find a better surrogate than the one based on the symmetric residual bound in (2.11). We indicate in the next section that such better surrogates can indeed be obtained at affordable computational cost for a wide range of problems through combining *Petrov-Galerkin projections* with appropriate *unsymmetric* residual bounds. The approach can be viewed as *preconditioning* the continuous problem already on the infinite dimensional level.

**4.1. Modifying the Variational Formulation.** We consider now a wider class of (not necessarily coercive) variational problems

$$(4.3) \quad b(u, v) = \langle f, v \rangle, \quad v \in V,$$

where we assume at this point only for each  $f \in V'$  the existence of a unique solution  $u \in U$ , i.e., the operator  $B : U \rightarrow V'$ , induced by  $b(\cdot, \cdot)$ , is bijective. This is well known to be equivalent to the validity of

$$(4.4) \quad \begin{cases} \inf_{w \in W} \sup_{v \in V} \frac{b(w, v)}{\|w\|_U \|v\|_V} \geq \beta, & \sup_{v \in V} \sup_{w \in U} \frac{b(w, v)}{\|w\|_U \|v\|_V} \leq C_b, \\ \text{for } v \in V \exists w \in W, \text{ such that } b(w, v) \neq 0, \end{cases}$$

for some constants  $\beta, C_b$ . However, one then faces two principal obstructions regarding an RBM based on the scheme SGA:

(a) first, as in the case of (4.2) for small diffusion,  $\kappa_{U,V'}(B) \leq C_b/\beta$  could be very large so that the corresponding error-residual relation

$$(4.5) \quad \|u - v\|_U \leq \beta^{-1} \|f - Bv\|_{V'}, \quad v \in U,$$

renders a corresponding residual based surrogate ill-conditioned.

(b) When  $b(\cdot, \cdot)$  is not coercive, the Galerkin projection does, in general, not perform as well as the best approximation.

The following approach has been used in [10] to address both (a) and (b). The underlying basic principle is not new, see [1], and variants of it have been used for different purposes in different contexts such as least squares finite element methods [18] and, more recently, in connection with *discontinuous Petrov Galerkin methods* [9, 11, 12]. In the context of RBMs the concepts of *natural norms* goes sort of half way by sticking in the end to Galerkin projections [24]. This marks an essential distinction from the approach in [10] discussed later below.

The idea is to change the topology of one of the spaces so as to (ideally) make the corresponding induced operator an *isometry*, see also [9]. Following [10], fixing for instance,  $\|\cdot\|_V$ , one can define

$$(4.6) \quad \|w\|_{\hat{U}} := \sup_{v \in V} \frac{b(w, v)}{\|v\|_V} = \|Bw\|_{V'}, \quad w \in U,$$

which means that one has for  $Bu = f$

$$(4.7) \quad \|u - w\|_{\hat{U}} = \|f - Bw\|_{V'}, \quad w \in U,$$

a perfect error-residual relation. It also means that, replacing  $\|\cdot\|_U$  in (4.4) by  $\|\cdot\|_{\hat{U}}$ , yields the inf-sup constant  $\hat{\beta} = 1$ . Alternatively, fixing  $\|\cdot\|_U$ , one may set

$$(4.8) \quad \|v\|_{\hat{V}} := \sup_{w \in U} \frac{b(w, v)}{\|w\|_U} = \|B^*v\|_{U'}, \quad v \in V,$$

to again arrive at an isometry  $B : U \rightarrow \hat{V}'$ , meaning

$$(4.9) \quad \|u - w\|_U = \|f - Bw\|_{\hat{V}'}, \quad w \in U.$$

Whether the norm for  $U$  or for  $V$  is prescribed depends on the problem at hand and we refer to [7, 10, 9] for examples of both type.

Next note that for any subspace  $W \subset U$  one has

$$(4.10) \quad u_W = \operatorname{argmin}_{w \in W} \|u - w\|_{\hat{U}} = \operatorname{argmin}_{w \in W} \|f - Bw\|_{V'},$$

and analogously for the pair  $(U, \hat{V})$ , i.e., the best approximation in the  $\hat{U}$ -norm is a *minimum residual solution* in the  $V'$ -norm.

To use residuals in  $V'$  as surrogates requires fixing a suitable discrete projection for a given trial space. In general, in particular when  $V \neq U$ , the Galerkin projection is no longer appropriate since inf-sup stability of the *infinite dimensional* problem is no longer inherited by an arbitrary pair of *finite dimensional* trial and test spaces. To see which type of projection would be ideal, denote by  $R_U : U' \rightarrow U$  the Riesz map defined for any linear functional  $\ell \in U'$  by

$$\langle \ell, w \rangle = (R_U \ell, w)_U, \quad w \in U.$$

Then, by (4.6), for any  $w \in W \subset U$ , taking  $v := R_V Bw \in V$  one has

$$b(w, v) = \langle Bw, v \rangle = \langle Bw, R_V Bw \rangle = (Bw, Bw)_{V'} = (w, w)_{\hat{U}}.$$

Thus, in particular,

$$b(u - u_h, R_V b w) = (u - u_h, w)_{\hat{U}},$$

i.e., given  $W \subset U$ , using  $V(W) := R_V B(W)$  as a test space in the *Petrov-Galerkin* scheme

$$(4.11) \quad b(u_h, v) = \langle f, v \rangle, \quad v \in V(W) := R_V B(W),$$

is equivalent to computing the  $\hat{U}$ -orthogonal projection of the exact solution  $u$  of (4.3) and hence the best  $\hat{U}$ -approximation to  $u$ . One readily sees that this also means

$$(4.12) \quad \inf_{w \in W} \sup_{v \in V(W)} \frac{b(w, v)}{\|w\|_{\hat{U}} \|v\|_V} = 1,$$

i.e., we have a Petrov-Galerkin scheme for the pair of spaces  $W, V(W)$  with perfect stability and the Petrov-Galerkin projection is the best  $\hat{U}$ -projection. Unfortunately, this is not of much help yet, because computing the *ideal test space*  $V(W) = R_V B(W) = B^{-*} R_{\hat{U}}^{-1}(W)$  is

not numerically feasible. Nevertheless, it provides a useful orientation for finding good and practically realisable pairs of trial and test spaces, as explained next.

**4.2. A Saddle Point Formulation.** We briefly recall now from [9, 10] an approach to deriving from the preceding observations a practically feasible numerical scheme which, in particular, fits into the context of RBMs. Taking (4.10) as point of departure we notice that the minimization of  $\|f - Bw\|_{V'}$  over  $W$  is a least squares problem whose normal equations read: find  $u_W \in W$  such that (with  $R_{V'} = R_V^{-1}$ )

$$(4.13) \quad 0 = (f - Bu_W, Bw)_{V'} = \langle R_V^{-1}(f - Bu_W), Bw \rangle, \quad w \in W.$$

Introducing the auxiliary variable  $r := R_V^{-1}(f - Bu_W)$  which is equivalent to

$$(4.14) \quad \langle R_V r, v \rangle = (r, v)_V = \langle f - Bu_w, v \rangle, \quad v \in V,$$

the two relations (4.13) and (4.14) can be rewritten in form of the *saddle point problem*

$$(4.15) \quad \begin{aligned} (r, v)_V + b(u_W, v) &= \langle f, v \rangle, & v \in V. \\ b(w, r) &= 0, & w \in W. \end{aligned}$$

The corresponding inf-sup constant is still one (since the supremum is taken over all of  $V$ ) and  $(\cdot, \cdot)_V$  is a scalar product so that (4.15) has a unique solution  $u_W$ , see e.g. [5]. Taking for any  $w \in W$  the test function  $v = R_V Bw \in V$  in the first line of (4.15), one obtains

$$(r, v)_V = (r, R_V Bw)_V = \langle r, Bw \rangle = b(w, r) = 0,$$

by the second line in (4.15) so we see that  $\langle f, R_V Bw \rangle = b(u_W, R_V Bw)$  holds for all  $w \in W$  which means that  $u_W$  solves the ideal Petrov-Galerkin problem (4.11). Thus (4.15) is equivalent to the ideal Petrov Galerkin scheme (4.11).

Of course, (4.15) is still not realizable since the space  $V$  is still the full infinite dimensional space. One more step to arrive at a realizable scheme is the following: given the finite dimensional space  $W$  find a finite dimensional space  $Z \subset V$  so that when  $V$  in (4.15) is replaced by  $Z$ , one obtains a *stable* finite dimensional saddle point problem which is the same as saying that its inf-sup constant is safely bounded away from zero. Since  $Z = V$  would yield perfect stability the choice of  $Z \subset V$  can be viewed as a *stabilization*. To quantify this we follow [10] and say that for some  $\delta \in (0, 1)$ ,  $Z \subset V$  is  $\delta$ -proximal for  $W \subset U$  if  $Z$  is sufficiently close to the ideal test space  $V(W) = R_V B(W)$  in the sense that

$$(4.16) \quad \|(I - P_{V,Z})R_V Bw\|_V \leq \delta \|R_V Bw\|_V, \quad w \in W.$$

The related main findings from [10] can be summarized as follows.

**Theorem 4.2.** (i) *The pair  $(u_{W,Z}, r_{W,Z}) \in W \times Z \subset U \times V$  solves the saddle point problem*

$$(4.17) \quad \begin{aligned} (r_{W,Z}, v)_V + b(u_{W,Z}, v) &= \langle f, v \rangle, & v \in Z, \\ b(w, u_{W,Z}) &= 0, & w \in W, \end{aligned}$$

*if and only if  $u_{W,Z}$  solves the Petrov-Galerkin problem*

$$(4.18) \quad b(u_{W,Z}, v) = \langle f, v \rangle, \quad v \in P_{V,Z}(R_V B(W)).$$

(ii) *If  $Z$  is  $\delta$ -proximal for  $W$ , (4.18) is solvable and one has*

$$(4.19) \quad \begin{aligned} \|u - u_{W,Z}\|_{\hat{U}} &\leq \frac{1}{1-\delta} \inf_{w \in W} \|u_{W,Z} - w\|_{\hat{U}}, \\ \|u - u_{W,Z}\|_{\hat{U}} + \|r_{W,Z}\|_V &\leq \frac{2}{1-\delta} \inf_{w \in W} \|u_{W,Z} - w\|_{\hat{U}}. \end{aligned}$$

(iii)  $Z$  is  $\delta$ -proximal for  $W$  if and only if

$$(4.20) \quad \inf_{w \in W} \sup_{v \in Z} \frac{b(w, v)}{\|w\|_{\hat{U}} \|v\|_V} \geq \sqrt{1 - \delta^2}.$$

Note that (4.17) involves ordinary bilinear forms and finite dimensional spaces  $W, Z$  and (iii) says that the  $V$ -projection of the ideal test space  $R_V B(W)$  onto  $Z$  is a good test space if and only if  $Z$  is  $\delta$ -proximal for  $W$ . Loosely speaking,  $Z$  is large enough to “see” a substantial part of the ideal test space  $R_V B(W)$  under projection. The perhaps most important messages to be taken home regarding the RBM context read as follows.

**Remark 4.3.** (i) *The Petrov-Galerkin scheme (4.18) is realized through the saddlepoint problem (4.17) without explicitly computing the test space  $P_{V,Z}(R_V B(W))$ .*

(ii) *Moreover, given  $W$ , by compactness and (4.16), one can in principle enlarge  $Z$  so as to make  $\delta$  as small as possible, a fact that will be exploited later.*

(iii) *The solution component  $u_{W,Z}$  is a near-best approximation to the exact solution  $u$  in the  $\hat{U}$ -norm.*

(iv)  *$r_{W,Z}$  can be viewed as a lifted residual which tends to zero in  $V$  when  $W$  grows and can be used for a-posteriori error estimation, see [9]. In the Reduced Basis context this can be exploited for certifying the accuracy of the truth solutions and for constructing computationally feasible surrogates for the construction of the reduced bases.*

## 5. THE REDUCED BASIS CONSTRUCTION

We point out next how to use the preceding results for sampling the solution manifold  $\mathcal{M}$  of a given *parametric family of variational problems*: given  $y \in \mathcal{Y}$ ,  $f \in V'_y$ , find  $u(y) \in U_y$  such that

$$(5.1) \quad b_y(u(y), v) = \langle f, v \rangle, \quad v \in V_y,$$

in a way that the corresponding subspaces are rate-optimal. We will always assume that the dependence of the bilinear form  $b_y(\cdot, \cdot)$  on  $y \in \mathcal{Y}$  is affine in the sense of (2.13).

As indicated by the notation the spaces  $U_y, V_y$  for which the variational problems are well-posed in the sense that the induced operator  $B_y : U_y \rightarrow V'_y$  is bijective, could depend on  $y$  through  $y$ -dependent norms. However, to be able to speak of a “solution manifold”  $\mathcal{M}$  as a compact subset of some “reference Hilbert space”, the norms  $\|\cdot\|_{U_y}$  should be *uniformly* equivalent to some *reference* norm  $\|\cdot\|_U$  which has to be taken into account when formulating (5.1). In fact, under this condition, as shown in [10], for well-posed variational formulations of pure transport problems the dependence of the test spaces  $V_y$  on  $y \in \mathcal{Y}$  is essential, in that

$$(5.2) \quad V := \bigcap_{y \in \mathcal{Y}} V_y,$$

is a strict subset of each individual  $V_y$ . This complicates the construction of a tight surrogate. We refer to [10] for ways of dealing with this obstruction and confine the subsequent discussion for simplicity to cases where the test norms  $\|\cdot\|_{V_y}$  are also uniformly equivalent to a single reference norm  $\|\cdot\|_V$ , see the example later below.

Under the above assumptions, the findings of the preceding section will be used next to contrive a well-conditioned tight surrogate even for non-coercive or severely ill-conditioned variational problems which is then in general unsymmetric, i.e.,  $V_y \neq U_y$ . These surrogates will then be used in SGA. To obtain such a residual based well-conditioned surrogate in the

sense of (4.1), we first *renorm* the pairs of spaces  $U_y$  or  $V_y$  according to (4.6) or (4.8). In anticipation of the example below, for definiteness we concentrate on (4.6) and refer to [10] for a discussion of (4.8). As indicated above, we assume further that the norms  $\|\cdot\|_{\hat{U}_y}, \|\cdot\|_{V_y}$  are equivalent to reference norms  $\|\cdot\|_{\hat{U}}, \|\cdot\|_V$ , respectively.

**5.1. The Strategy.** Suppose that we have already constructed a pair of spaces  $U_n \subset U_y, V_n \subset V_y, y \in \mathcal{Y}$ , such that for a given  $\delta < 1$

$$(5.3) \quad \inf_{w \in U_n} \sup_{v \in V_n} \frac{b_y(w, v)}{\|w\|_{\hat{U}_y} \|v\|_{V_y}} \geq \sqrt{1 - \delta^2}, \quad y \in \mathcal{Y},$$

i.e.,  $V_n \subset V$  is  $\delta$ -proximal for  $U_n \subset U$ . Thus, by Theorem 4.2, the parametric saddle point problem

$$(5.4) \quad \begin{aligned} (r_n(y), v)_{V_y} + b_y(u_n(y), v) &= \langle f, v \rangle, & v \in V_n, \\ b(w, r_n(y)) &= 0, & w \in W, \end{aligned}$$

has for each  $y \in \mathcal{Y}$  a unique solution  $(u_n(y), r_n(y)) \in U_n \times V_n$ . By the choice of norms we know that

$$(5.5) \quad \|u(y) - u_n(y)\|_{\hat{U}_y} = \|f - B_\mu u_n(y)\|_{V'_y}, \quad y \in \mathcal{Y},$$

i.e.,

$$(5.6) \quad R(y, U_n \times V_n) := \|f - B_\mu u_n(y)\|_{V'_y}, \quad y \in \mathcal{Y}$$

suggests itself as a surrogate. There are some subtle issues about how to evaluate  $R(y, U_n \times V_n)$  in the dual  $V'_N$  of a sufficiently large truth space  $V_N \subset V_y, y \in \mathcal{Y}$ , so as to faithfully reflect errors in  $\hat{U}_\mu$ , not only in the truth space  $U_N \subset U_y$  but in  $\hat{U}$ , and how these quantities are actually related to the auxiliary variable  $\|r_n(y)\|_{V_y}$  which is computed anyway. As indicated before, these issues are aggravated when the norms  $\|\cdot\|_{V_y}$  are *not* all equivalent to a single reference norm. We refer to a corresponding detailed discussion in [10, §5.1] and continue working here for simplicity with the idealized version (5.6) and assume its offline feasibility.

Thus, we can evaluate the errors  $\|u(y) - u_n(y)\|_{\hat{U}_y}$  and can determine a maximizing parameter  $y_{n+1}$  for which

$$(5.7) \quad \|u(y_{n+1}) - u_n(y_{n+1})\|_{\hat{U}_y} = \max_{y \in \mathcal{Y}} \|f - B_\mu u_n(y)\|_{V'_y}.$$

Now relation (4.19) in Theorem 4.2 tells us that for each  $y \in \mathcal{Y}$

$$(5.8) \quad \|u(y) - u_n(y)\|_{\hat{U}_y} \leq (1 - \delta)^{-1} \inf_{w \in U_n} \|u(y) - w\|_{\hat{U}_y},$$

i.e.,  $u_n(y)$  is a near best approximation to  $u$  from  $U_n$  which is, in fact, the nearer to the best approximation the smaller  $\delta$ . By (5.5) and (5.8), the surrogate (5.6) is indeed well-conditioned with condition number close to one for small  $\delta$ .

A natural strategy is now to enlarge  $U_n$  to  $U_{n+1} := U_n + \text{span}\{u(y_{n+1})\}$ . In fact, this complies with the *weak greedy* step (3.1) in §3 with weakness parameter  $\gamma = (1 - \delta)$  as close to one as one wishes, when  $\delta$  is chosen accordingly small, provided that the pair of spaces  $U_n, V_n$  satisfies (5.3). A repetition would therefore, in principle, be a realization of Algorithm 1, SGA, establishing rate-optimality of this RBM. Obviously, the critical condition for such a procedure to work is to ensure at each stage the validity of the weak-greedy condition (3.1) which in the present situation means that the companion space  $V_n$  is at each stage  $\delta$ -proximal for  $U_n$ . So far we have not explained yet how to grow  $V_n$  along with  $U_n$  so as to ensure  $\delta$ -proximality. This is explained in the subsequent section.

**Remark 5.1.** *One should note that, due to the possible parameter dependence of the norms  $\|\cdot\|_{\hat{U}_y}, \|\cdot\|_{V_y}$  on  $y$ , obtaining tight surrogates with the aid of an explicit Petrov-Galerkin formulation, would be infeasible in an RBM context because one would have to recompute the corresponding (parameter dependent) test basis for each parameter query which is not online-feasible. It is therefore actually crucial to employ the saddle point formulation in the context of RBMs since this allows us to determine a space  $V_n$  of somewhat larger dimension than  $U_n$  but stabilizes the saddle point problem for all  $y$  simultaneously.*

**5.2. A Greedy Stabilization.** A natural option is to enlarge  $V_n$  by the second component  $r_n(y_{n+1})$  of (5.4). Note though that the lifted residuals  $r_n$  tend to zero as  $n \rightarrow \infty$ . Hence, the solution manifold of the ( $y$ -dependent version of the) saddle point formulation (4.15) has the form

$$\mathcal{M} \times \{0\},$$

where  $\mathcal{M}$  is the solution manifold of (5.1) (since  $r(y) = 0$  for  $y \in \mathcal{Y}$ ). Thus, the spaces  $V_n$  are *not* needed to approximate the solution manifold. Instead the sole purpose of the space  $V_n$  is to guarantee stability. At any rate, the grown pair  $U_{n+1}, V_n + \text{span}\{r_n(y_{n+1})\} =: V_{n+1}^0$  may fail to satisfy now (5.3).

Therefore, in general one has to further enrich  $V_{n+1}^0$  by additional *stabilizing* elements again in a greedy fashion until (5.3) holds for the desired  $\delta$ . For problems that initially arise as natural saddle point problems such as the Stokes system, enrichments by so called *supremizers* (to be defined in a moment) have been proposed already in [14, 15, 22]. In these cases it is possible to enrich  $V_{n+1}^0$  by a *fixed* a priori known number of such supremizers to guarantee inf-sup stability. As shown in [10], this is generally possible when using fixed (parameter independent) reference norms  $\|\cdot\|_{\hat{U}}, \|\cdot\|_V$  for  $U$  and  $V$ . For the above more general scope of problems a greedy strategy was proposed and analyzed in [10], a special case of which is also considered in [15] without analysis. The strategy in [10] adds only as many stabilizing elements as are actually needed to ensure stability and works for a much wider range of problems including singularly perturbed ones. In cases where not all parameter dependent norms  $\|\cdot\|_{V_y}$  are equivalent such a strategy is actually necessary and its convergence analysis is then more involved, see [10].

To explain the procedure, suppose that after growing  $U_n$  to  $U_{n+1}$  we have already generated an enrichment  $V_{n+1}^k$  of  $V_{n+1}^0$  (which could be, for instance, either  $V_{n+1}^0 := V_n + \text{span}\{r_n(y_{n+1})\}$  or  $V_{n+1}^0 := V_n$ ) but the pair  $U_{n+1}, V_{n+1}^k$  still fails to satisfy (5.3) for the given  $\delta < 1$ . To describe the next enrichment from  $V_{n+1}^k$  to  $V_{n+1}^{k+1}$  we first search for a parameter  $\bar{y} \in \mathcal{Y}$  and a function  $\bar{w} \in U_n$  for which the inf-sup condition (5.3) is worst, i.e.,

$$(5.9) \quad \sup_{v \in V_{n+1}^k} \frac{b_{\bar{y}}(\bar{w}, v)}{\|v\|_{V_{\bar{y}}}\|\bar{w}\|_{\hat{U}_{\bar{y}}}} = \inf_{y \in \mathcal{Y}} \left( \inf_{w \in U_{n+1}} \sup_{v \in V_{n+1}^k} \frac{b_y(w, v)}{\|v\|_{V_y}\|w\|_{\hat{U}_y}} \right).$$

If this worst case inf-sup constant does not exceed yet  $\sqrt{1 - \delta^2}$ , the current space  $V_{n+1}^k$  does not contain an effective supremizer for  $\bar{y}, \bar{w}$ , yet. However, since the truth space satisfies the uniform inf-sup condition (5.3) there *must exist* a good supremizer in the truth space which can be seen to be given by

$$(5.10) \quad \bar{v} = \operatorname{argmax}_{v \in V_{\bar{y}}} \frac{b_{\bar{y}}(\bar{w}, v)}{\|v\|_{V_{\bar{y}}}\|\bar{w}\|_{\hat{U}_{\bar{y}}}},$$

providing the next enrichment

$$(5.11) \quad V_{n+1}^{k+1} := \text{span}\{V_{n+1}^k, \bar{v}\}.$$

We defer some comments on the numerical realization of finding  $\bar{y}, \bar{v}$  in (5.9), (5.10) to the next section.

This strategy can now be applied recursively until one reaches a satisfactory uniform inf-sup condition for the reduced spaces. Again, the termination of this stabilization loop is easily ensured when (2.13) holds and the norms  $\|\cdot\|_{\hat{U}_y}, \|\cdot\|_{V_y}$  are uniformly equivalent to reference norms  $\|\cdot\|_{\hat{U}}, \|\cdot\|_V$ , respectively, but is more involved in the general case [10].

**5.3. The Double Greedy Scheme and Main Result.** Thus, in summary, to ensure that the greedy scheme SGA with the particular surrogate (5.6), based on the corresponding *outer greedy* step for extending  $U_n$  to  $U_{n+1}$ , has the *weak greedy property* (3.1), one can employ an *inner stabilizing greedy* loop producing a space  $V_{n+1} = V_{n+1}^{k^*}$  which is  $\delta$ -proximal for  $U_{n+1}$ . Here  $k^* = k^*(\delta)$  is the number of enrichment steps needed to guarantee the validity of (5.3) for the given  $\delta$ . A sketchy version of the corresponding “enriched” SGA, developed in [10], looks as follows:

---

**Algorithm 3** double greedy algorithm

---

```

1: function SGA-DOU
2:   Initialize  $U_1, V_1^0, \delta \in (0, 1)$ , target accuracy tol,  $n \leftarrow 1$ ,
3:   while  $\sigma_n(\mathcal{M}) > \text{tol}$  do
4:     while  $U_n, V_n^0$  fail to satisfy (5.3) do
5:       compute  $V_n$  with the aid of the inner stabilizing greedy loop,
6:     end while
7:     given  $U_n, V_n$ , satisfying (5.3), compute  $U_{n+1}, V_{n+1}^0$  with the aid of the outer greedy
       step 4, (2.3) in algorithm SGA for the surrogate (5.6),
8:   end while
9: end function

```

---

As indicated above, both Algorithm 1, SGA and Algorithm 3, SGA-DOU are surrogate based greedy algorithms. The essential difference is that for non-coercive problems or problems with an originally large variational condition number in SGA-DOU an additional interior greedy loop provides a tight well-conditioned (unsymmetric) surrogate which guarantees the desired weak greedy property (with weakness constant  $\gamma$  as close to one as one wishes) needed for rate-optimality.

Of course, the viability of Algorithm SGA-DOU hinges mainly on two questions:

(a) how to find the worst inf-sup constant in (5.9) and how to compute the supremizer in (5.10)?

(b) does the inner greedy loop terminate (early enough)?

As for (a), it is well known that, fixing bases for  $U_n, V_n^k$ , finding the worst inf-sup constant amounts to determine for  $y \in \mathcal{Y}$  the cross-Gramian with respect to  $b_y(\cdot, \cdot)$  and compute its smallest singular value. Since these matrices are of size  $n \times (n + k)$  and hence (presumably) of “small” size, a search over  $\mathcal{Y}$  requires solving only problems in the reduced spaces and are under the assumption (2.13) therefore offline-feasible. The determination of the corresponding supremizer  $\bar{v}$  in (5.10), in turn, is based on the well-known observation that

$$\operatorname{argmax}_{v \in V_{\bar{y}}} \frac{b_{\bar{y}}(\bar{w}, v)}{\|v\|_{V_{\bar{y}}}} = R_{V_{\bar{y}}} B_{\bar{y}} \bar{w},$$

which is equivalent to solving the Galerkin problem

$$(\bar{v}, z)_{V_{\bar{y}}} = b_{\bar{y}}(\bar{w}, z), \quad z \in V_{\bar{y}}.$$

Thus, each enrichment step requires one offline-Galerkin solve in the truth space.

A quantitative answer to question (b) is more involved. We are content here with a few related remarks and we refer to a detailed discussion of this issue in [10]. As mentioned before, when all the norms  $\|\cdot\|_{\hat{U}_y}, \|\cdot\|_{V_y}, y \in \mathcal{Y}$ , are equivalent to reference norms  $\|\cdot\|_{\hat{U}}, \|\cdot\|_V$ , respectively, the inner loop terminates after at most the number of terms in (2.13). When the norms  $\|\cdot\|_{V_y}$  are no longer uniformly equivalent to a single reference norm termination is less clear. Of course, since all computations are done in a truth space which is finite dimensional, compactness guarantees termination after finitely many steps. However, the issue is that the number of steps should not depend on the truth space dimension. The reasoning in [10] used to show that (under mild assumptions) termination happens after a finite number of steps which is *independent* of the truth space dimension, is based on the following fact. Defining  $U_n^1(y) := \{w \in U_n : \|w\|_{\hat{U}_y} = 1\}$ , solving the problem

$$(5.12) \quad (\bar{y}, \bar{w}) := \operatorname{argmax}_{y \in \mathcal{Y}; w \in U_n^1(y)} \inf_{\phi \in V_n^k} \|R_{V_y} B_y w - \phi\|_{V_y},$$

when all the  $\|\cdot\|_{\hat{U}_y}$ -norms are equivalent to a single reference norm, can be shown to be equivalent to a greedy step of the type (5.9) and can hence again be reduced to similar small eigenvalue problems in the reduced space. Note, however, that (5.12) is similar to a greedy space growth used in the outer greedy loop and for which some understanding of convergence is available. Therefore, successive enrichments based on (5.12) are studied in [10] regarding their convergence. The connection with the inner stabilizing loop based on (5.9) is that

$$\operatorname{argmax}_{y \in \mathcal{Y}; w \in U_n^1(y)} \inf_{\phi \in V_n^k} \|R_{V_{\bar{y}}} B_{\bar{y}} \bar{w} - \phi\|_{V_{\bar{y}}} \leq \delta,$$

just means

$$\inf_{\phi \in V_n^k} \|R_{V_y} B_y w - \phi\|_{V_y} \leq \delta \|R_{V_y} B_y\|_{V_y} = \delta \|w\|_{\hat{U}_y}, \quad w \in U_n, y \in \mathcal{Y},$$

which is a statement on  $\delta$ -proximality known to be equivalent to inf-sup stability, see Theorem 4.2, and (4.16).

A central result from [10] can be formulated as follows, see [10, Theorem 5.5].

**Theorem 5.2.** *If (2.13) holds and the norms  $\|\cdot\|_{\hat{U}_y}, \|\cdot\|_{V_y}$  are all equivalent to a single reference norm  $\|\cdot\|_{\hat{U}}, \|\cdot\|_V$ , respectively, and the surrogates (5.6) are used, then the scheme SGA-DOU is rate optimal, i.e., the greedy errors  $\sigma_n(\mathcal{M})_{\hat{U}}$  decay at the same rate as the  $n$ -widths  $d_n(\mathcal{M})_{\hat{U}}$ ,  $n \rightarrow \infty$ .*

Recall that the quantitative behavior of the greedy error rates are directly related to those of the  $n$ -widths by  $\gamma^{-1} = c_S^{-1}$ , see Theorem 3.1. This suggests that a fast decay of  $d_n(\mathcal{M})_{\hat{U}}$  is reflected by the corresponding greedy errors already for moderate values of  $n$  which is in the very interest of reduced order modeling. This will be confirmed by the expamples below. In this context an important feature of SGA-DOU is that through the choice of the  $\delta$ -proximity parameter the weakness parameter  $\gamma$  can be driven towards one, of course, at the expense of somewhat larger spaces  $V_{n+1}$ . Hence, stability constants close to one are built into the method. This is to be contrasted by the conventional use of SGA based on surrogates that are not ensured to be well conditioned and for which the computation of the certifying stability constants tends to be computationally expensive.

**5.4. A Numerical Example.** The preceding theoretical results are illustrated next by a numerical example that brings out some of the main features of the scheme. While the double greedy scheme applies to non-coercive or indefinite problems (e.g. see [10] for pure transport) we focus here on a classical *singularly perturbed* problem because it addresses also some principal issues for RBMs regarding problems with *small scales*. Specifically, we consider the *convection-diffusion* problem (4.2) on  $\Omega = (0, 1)^2$  for a simple *parameter dependent convection* field

$$\vec{b}(y) := \begin{pmatrix} \cos y \\ \sin y \end{pmatrix}, \quad y \in [0, 2\pi), \quad c = 1,$$

keeping for simplicity the diffusion level  $\varepsilon$  fixed but allowing it to be arbitrarily small. All considerations apply as well to variable and parameter dependent diffusion with any arbitrarily small but strictly positive lower bound. The “transition” to a pure transport problem is discussed in detail in [10, 28]. Parameter dependent convection directions mark actually the more difficult case and are, for instance, of interest with regard to kinetic models.

Let us first briefly recall the main challenges posed by (4.2) for very small diffusion  $\varepsilon$ . The problem becomes obviously dominantly unsymmetric and singularly perturbed. Recall that for each positive  $\varepsilon$  the problem possesses for each  $y \in \mathcal{Y}$  a unique solution  $u(y)$  in  $U = H_0^1(\Omega)$  that has a zero trace on the boundary  $\partial\Omega$ . However, as indicated earlier, the condition number  $\kappa_{U,U'}(B_y)$  of the underlying convection-diffusion operator  $B_y$ , viewed as an operator from  $U = H_0^1(\Omega)$  onto  $U' = H^{-1}(\Omega)$ , behaves like  $\varepsilon^{-1}$ , that is, it becomes increasingly *ill conditioned*. This has well known consequences for the performance of numerical solvers but above all for the stability of corresponding discretizations.

We emphasize that the conventional mesh dependent stabilizations like SUPG (cf. [17]) do *not* offer a definitive remedy because the corresponding condition, although improved, remains very large for very small  $\varepsilon$ . In [19] SUPG-stabilization for the offline truth calculations as well as for the low-dimensional online Galerkin projections are discussed for moderate Peclét-numbers of the order of up to  $10^3$ . In particular, comparisons are presented when only the offline phase uses stabilization while the un-stabilized bilinear form is used in the online phase, see also the references in [19] for further related work.

As indicated earlier, we also remark in passing that the singularly perturbed nature of the problem poses an additional difficulty concerning the choice of the truth space  $U_{\mathcal{N}}$ . In fact, when  $\varepsilon$  becomes very small one may not be able to afford resolving correspondingly thin layers in the truth space which increases the difficulty of capturing essential features of the solution by the reduced model.

This problem is addressed in [10] by resorting to a weak formulation that does not use  $H_0^1(\Omega)$  (or a renormed version of it) as a trial space but builds on the results from [7]. A central idea is to enforce the boundary conditions on the outflow boundary  $\Gamma_+(y)$  only weakly. Here  $\Gamma_+(y)$  is that portion of  $\partial\Omega$  for which the inner product of the outward normal and the convection direction is positive. Thus, solutions are initially sought in the larger space  $H_{0,\Gamma^-(y)}^1(\Omega) =: U_-(y)$  enforcing homogeneous boundary conditions only on the *inflow* boundary  $\Gamma_-(y)$ . Since the outflow boundary, and hence also the inflow boundary depend on the parameter  $y$ , this requires subdividing the parameter set into smaller sectors, here four, for which the outflow boundary  $\Gamma_+ = \Gamma_+(y)$  remains unchanged. We refer in what follows for simplicity to one such sector denoted again by  $\mathcal{Y}$ .

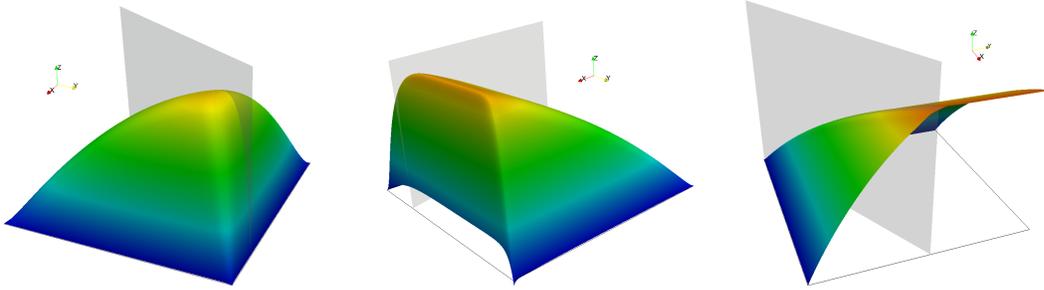


FIGURE 5.1. left:  $\varepsilon = 2^{-5}, n = 6, n_V = 13$ ; middle:  $\varepsilon = 2^{-7}, n = 7, n_V = 20$ ; right:  $\varepsilon = 2^{-26}, n = 20, n_V = 57$ .

The following prescription of the *test space* falls into the category (4.6) where the norm for  $U$  is adapted. Specifically, choosing

$$s_y(u, v) := \frac{1}{2} (\langle B_y u, v \rangle + \langle B_y v, u \rangle),$$

$$\|v\|_{\bar{V}_y}^2 := s_y(v, v) = \varepsilon |v|_{H^1(\Omega)}^2 + \left\| \left( c - \frac{1}{2} \operatorname{div} \vec{b}(y) \right)^{1/2} v \right\|_{L_2(\Omega)}^2,$$

in combination with a boundary penalization on  $\Gamma_+$ , we follow [7, 28] and define

$$\|u\|_{\bar{U}_y}^2 := \|\bar{B}_y u\|_{\bar{V}_y}^2 = \|\bar{B}_y u\|_{\bar{V}_y'}^2 + \lambda \|u\|_{H_b(\mu)}^2,$$

where  $H_b(y) = H_{00}^{1/2}(\Gamma_+(y))$ ,  $\bar{V}_y := V_y \times H_b(y)'$  and  $\bar{B}_y$  denotes the operator induced by this weak formulation over  $\bar{U}_y := H_{0, \Gamma_-(y)}^1(\Omega) \times H_b(y)$ . The corresponding variational formulation is of minimum residual type (cf. (4.10)) and reads

$$(5.13) \quad u(y) = \operatorname{argmin}_{w \in U_-(y)} \{ \|\bar{B}_y w - f\|_{\bar{V}_y'}^2 + \lambda \|w\|_{H_b(y)}^2 \}.$$

One can show that its (infinite dimensional) solution, whenever being sufficiently regular, solves also the strong form of the convection diffusion problem (4.2). Figure 5.1 illustrates the effect of this formulation where we set  $n = \dim U_n, n_V := \dim V_n$

The shaded planes shown in Figure 5.1 indicate the convection direction for which the snapshot is taken. For moderately large diffusion the boundary layer at  $\Gamma_+$  is resolved by the truth space discretization and the boundary conditions at the outflow boundary are satisfied exactly. For smaller diffusion in the middle example the truth space discretization can no longer resolve the boundary layer and for very small diffusion (right) the solution is close to the one for pure transport. The rationale of (5.13) is that all norms commonly used for convection diffusion equations resemble the one chosen here, for instance in the form of a mesh dependent “broken norm”, which means that most part of the incurred error of an approximation is concentrated in the layer region, see e.g. [27, 26]. Hence, when the layers are not resolved by the discretization, enforcing the boundary conditions does not improve accuracy and, on the contrary, may degrade accuracy away from the layer by causing oscillations. The present formulation instead avoids any non-physical oscillations and enhances accuracy in those parts of the domain where this is possible for the afforded discretization, see [7, 10, 28] for a detailed discussion. The following table quantifies the results for the case of small diffusion  $\varepsilon = 2^{-26}$  and a truth discretization whose a posteriori error bound is 0.002.

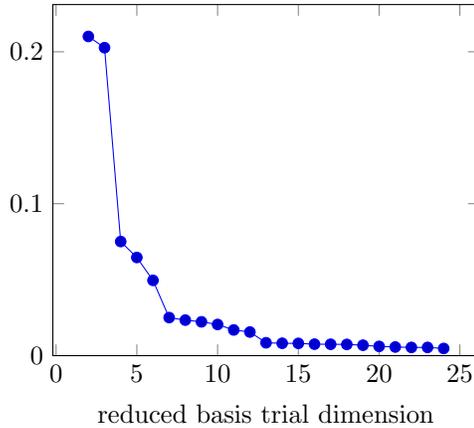


FIGURE 5.2. Convection-diffusion equation,  $\varepsilon = 2^{-26}$ , maximal a-posteriori error 0.00208994

TABLE 1. Convection-diffusion equation,  $\varepsilon = 2^{-26}$ , maximal a-posteriori error 0.00208994

$n$	$n_V$	$\delta$	surrogate	surr/a-post	$n$	$n_V$	$\delta$	surrogate	surr/a-post
2	5	1.36e-03	2.10e-01	1.01e+02	14	39	1.17e-04	8.15e-03	3.90e+00
4	9	1.10e-02	7.51e-02	3.59e+01	16	45	9.79e-05	7.56e-03	3.62e+00
6	15	1.75e-03	4.95e-02	2.37e+01	18	51	6.32e-05	7.40e-03	3.54e+00
8	21	9.16e-04	2.34e-02	1.12e+01	20	57	4.74e-05	6.09e-03	2.92e+00
10	27	3.65e-04	2.05e-02	9.82e+00	22	63	2.36e-05	5.43e-03	2.60e+00
12	33	3.34e-04	1.56e-02	7.45e+00	24	65	2.36e-05	4.73e-03	2.27e+00

The columns 3 and 8 show the  $\delta$  governing the condition of the saddle point problems (and hence of the corresponding Petrov-Galerkin problems), see (5.3), the greedy space growth is based upon. Hence the surrogates are very tight giving rise to weakness parameters very close to one. As indicated in Remark 4.3 one can use also an a posteriori bound for the truth solution based on the corresponding lifted residual. Columns 5 and 10 show therefore the relative accuracy of the current reduced model and the truth model. This corresponds to the stability constants computed by conventional RBMs. Even for elliptic problems these latter ones are significantly larger than the ones for the present singularly perturbed problem which are guaranteed to be close to one by the method itself. Based on the a posteriori bounds for the truth solution (which are also obtained with the aid of tailored well-conditioned variational formulations, see [7]), the greedy space growth is stopped when the surrogates reach the order of the truth accuracy. As illustrated in Figure 5.2, in the present example this is essentially already the case for  $\leq 20$  trial reduced basis functions and almost three times as many test functions. To show this “saturation effect” we have continued the space growth formally up to  $n = 24$  showing no further significant improvement which is in agreement with the resolution provided by the truth space. These relations agree with the theoretical predictions in [10]. Figure 5.2 illustrates

also the rapid gain of accuracy by the first few reduced basis functions which supports the fact that the solution manifold is “well seen” by the Petrov-Galerkin surrogates. More extensive numerical tests shown in [10] show that the achieved stability is independent of the diffusion but the larger the diffusion the smaller become the dimensions  $n = \dim U_n, n_V = \dim V_n$  for the reduced spaces. This indicates the expected fact that the larger the diffusion the smoother is the dependence of  $u(y)$  on the parameter  $y$ . In fact, when  $\varepsilon \rightarrow 0$  one approaches the regime of pure transport where the smoothness of the parameter dependence is merely Hölder continuity requiring for a given target accuracy a larger number of reduced basis functions, see [10].

## 6. IS IT NECESSARY TO RESOLVE ALL OF $\mathcal{M}$ ?

The central focus of the preceding discussion has been to control the maximal deviation

$$(6.1) \quad \sigma_n(\mathcal{M})_U = \max_{y \in \mathcal{Y}} \|u(y) - P_{U,U_n} u(y)\|_U,$$

and to push this deviation below a given tolerance for  $n$  as small as possible. However, in many applications one is not interested in the whole solution field but only in a *quantity of interest*  $I(y)$ , typically of the form  $I(y) = \ell(u(y))$  where  $\ell \in U'$  is a bounded linear functional. Looking then for some desired optimal state  $I^* = \ell(u(y^*))$  one is interested in a guarantee of the form

$$(6.2) \quad |\ell(u_n(y)) - \ell(u(y))| \leq \text{tol}, \quad y \in \mathcal{Y},$$

where the states  $u_n(y)$  belong to a possibly small reduced space  $U_n$  in order to be then able to carry out the optimization over  $y \in \mathcal{Y}$  in the small space  $U_n \subset U$ . Asking only for the values of just a *linear functional* of the solution seems to be much less demanding than asking for the whole solution and one wonders whether this can be exploited in favor of even better online efficiency.

Trying to reduce computational complexity by exploiting the fact that, retrieving only a linear functional of an unknown state - a scalar quantity - may require less information than recovering the whole state, is the central theme of *goal oriented* adaptation in finite element methods, see [3]. Often the desired accuracy is indeed observed to be reached by significantly coarser discretizations than needed to approximate the whole solution within a corresponding accuracy. The underlying effect, sometimes referred to as “squared accuracy” is well understood and exploited in the RBM context as well, see [16, 21]. We briefly sketch the main ideas for the current larger scope of problems and point out that, nevertheless, a guarantee of the form (6.2) ultimately requires controlling the maximal deviation of a reduced space in the sense of (6.1). Hence, an optimal sampling of a solution manifold remains crucial.

First, a trivial estimate gives for  $\ell \in U'$

$$(6.3) \quad |\ell(u_n(y)) - \ell(u(y))| \leq \|\ell\|_{U'} \|u_n(y) - u(y)\|_U$$

so that a control of  $\sigma_n(\mathcal{M})_U$  would indeed yield a guarantee. However, the  $n$  needed to drive  $\|\ell\|_{U'} \sigma_n(\mathcal{M})_U$  below tol is usually larger than necessary.

To explain the principle of improving on (6.3) we consider again a variational problem of the form (4.3) (suppressing any parameter dependence for a moment) for a pair of spaces  $U, V$  where we assume now that  $\kappa_{U,V'}(B) \leq C_b/c_b$  is already small, possibly after renorming an initial less favorable formulation through (4.6) or (4.8). Let  $u \in U$  again denote the exact solution of (4.3). Given a  $\ell \in U'$  we wish to approximate  $\ell(u)$ , using an *approximate solution*  $\bar{u} \in W \subset U$  defined by

$$(6.4) \quad b(\bar{u}, v) = \langle f, v \rangle, \quad v \in \tilde{V}(W) \subset V,$$

where  $\tilde{V}(W)$  is a suitable test space generated by the methods discussed in §4.1. In addition we will use the solution  $z \in V$  of the *dual problem*:

$$(6.5) \quad b(w, z) = -\ell(w), \quad w \in U,$$

together with an approximation  $\bar{z} \in Z \subset V$  defined by

$$(6.6) \quad b(w, \bar{z}) = -\ell(w), \quad w \in \tilde{W}(Z) \subset U,$$

again with a suitable test space  $\tilde{W}(Z)$ . Recall that we need not determine the test spaces  $\tilde{V}(W), \tilde{W}(Z)$  explicitly but rather realize the corresponding Petrov-Galerkin projections through the equivalent saddle point formulations with suitable  $\delta$ -proximal auxiliary spaces generated by a greedy stabilization.

Then, defining the primal residual functional

$$(6.7) \quad r_{\bar{u}}(v) := r(\bar{u}, v) := b(u - \bar{u}, v) = \langle f, v \rangle - b(\bar{u}, v),$$

and adapting the ideas in [16, 21] for the symmetric case  $V = U$  to the present slightly more general setting, we claim that

$$(6.8) \quad \hat{\ell}(\bar{u}) := \ell(\bar{u}) - r(\bar{u}, \bar{z})$$

is an approximation to the true value  $\ell(u)$  satisfying

$$(6.9) \quad |\hat{\ell}(\bar{u}) - \ell(u)| \leq C \inf_{w \in W} \|u - w\|_U \inf_{v \in Z} \|z - v\|_V,$$

where  $C$  depends only on the inf-sup constant of the finite dimensional problems. In fact, since by (6.5),

$$\ell(u) - \ell(\bar{u}) = b(\bar{u} - u, z) = -r(\bar{u}, z),$$

one has  $\ell(u) = \ell(\bar{u}) - r(\bar{u}, z)$  and hence

$$\begin{aligned} |\hat{\ell}(\bar{u}) - \ell(u)| &= |\ell(\bar{u}) - r(\bar{u}, \bar{z}) - \ell(\bar{u}) + r(\bar{u}, z)| = |r(\bar{u}, z - \bar{z})| = |b(u - \bar{u}, z - \bar{z})| \\ &\leq C_b \|u - \bar{u}\|_U \|z - \bar{z}\|_V, \end{aligned}$$

which confirms the claim since  $\bar{u}, \bar{z}$  are near-best approximations due to the asserted inf-sup stability of finite dimensional problems.

Clearly, (6.9) says that in order to approximate  $\ell(u)$  the primal approximation in  $U$  need not resolve  $u$  at all as long as the dual solution  $z$  is approximated well enough. Moreover, when  $\ell$  is a local functional, e.g. a local average approximating a point evaluation,  $z$  is close to the corresponding Green's function with (near) singularity in the support of  $\ell$ . In the elliptic case  $z$  would be very smooth away from the support of  $\ell$  and hence well approximable by a relatively small number of degrees of freedom concentrated around the support of  $\ell$ . Thus, it may very well be more profitable to spend less effort on approximating  $u$  than on approximating  $z$ .

Returning to parameter dependent problems (5.1), the methods in §5 can now be used as follows to construct possibly small reduced spaces for a frequent online evaluation of the quantities  $I(y) = \ell(u(y))$ . We assume that we already have properly renormed families of norms  $\|\cdot\|_{U_y}, \|\cdot\|_{V_y}$ ,  $y \in \mathcal{Y}$ , with uniform inf-sup constants close to one. We also assume now that both families of norms are equivalent (by compactness of  $\mathcal{Y}$  uniformly equivalent) to reference norms  $\|\cdot\|_U, \|\cdot\|_V$ , respectively. Hence we can consider two solution manifolds

$$\mathcal{M}_{\text{pr}} := \{u(y) = B_y^{-1}f, y \in \mathcal{Y}\} \subset U, \quad \mathcal{M}_{\text{dual}} := \{z(y) := B_y^{-*}\ell, y \in \mathcal{Y}\} \subset V,$$

and use Algorithm 3, SGA-DOU to generate (essentially in parallel) two sequences of pairs of reduced spaces

$$(U_n, V_n), (Z_n, W_n), \quad n \in \mathbb{N}.$$

Here  $V_n \subset V, W_n \subset U$  are suitable stabilizing spaces such that for  $m < n$  and for the corresponding reduced solutions  $u_m(y) \in U_m, z_{n-m}(y) \in Z_{n-m}$  the quantity

$$(6.10) \quad I_{n,m}(y) := \ell(u_m(y)) - r(u_m(y), z_{n-m}(y))$$

satisfies

$$(6.11) \quad |I(y) - I_{n,m}(y)| \leq C \sigma_m(\mathcal{M}_{\text{pr}})_U \sigma_{n-m}(\mathcal{M}_{\text{dual}})_V,$$

with a constant  $C$  independent of  $n, m$ . The choice of  $m < n$  determines how to distribute the computational effort for computing the two sequences of reduced bases and their stabilizing companion spaces. By Theorem 5.2, one can see that whichever  $n$ -width rate  $d_n(\mathcal{M}_{\text{pr}})_U$  or  $d_n(\mathcal{M}_{\text{dual}})_V$  decays faster one can choose  $m < n$  to achieve for a total of  $\dim U_m + \dim Z_{n-m} = n$  the smallest error bound. Of course, the rates are not known and one can use the tight surrogates to bound and estimate the respective errors very accurately. For instance, when  $d_n(\mathcal{M}_{\text{pr}})_U \leq Cn^{-\alpha}$ ,  $d_n(\mathcal{M}_{\text{dual}})_V \leq Cn^{-\beta}$ ,  $m = \left\lfloor \left( \frac{\alpha}{\alpha+\beta} \right) n \right\rfloor$  yields an optimal distribution with a bound

$$(6.12) \quad |I(y) - I_{n,m}(y)| \leq C \left( \frac{\alpha + \beta}{\beta} \right)^\beta \left( \frac{\alpha + \beta}{\alpha} \right)^\alpha n^{-(\alpha+\beta)}.$$

In particular, when  $\beta > \alpha$  the dimensions on the reduced bases for the dual problem should be somewhat larger but essentially using the same dimensions for the primal and dual reduced spaces yields the rate  $n^{-(\alpha+\beta)}$  confirming the “squaring” when  $\alpha = \beta$ . In contrast, as soon as either one of the  $n$ -width rates decays exponentially it is best to grow only the reduced spaces for the faster decay while keeping a fixed space for the other side.

## 7. SUMMARY

We have reviewed recent developments concerning reduced basis methods with the following main focus. Using Kolmogorov  $n$ -width as a benchmark for the performance of reduced basis methods in terms of minimizing the dimensions of the reduced models for a given target accuracy, we have shown that this requires essentially to construct tight well-conditioned surrogates for the underlying variational problem. We have explained how *renormation* in combination with *inner stabilization loops* can be used to derive such residual based surrogates even for problem classes not covered by conventional schemes. This includes in a fully robust way indefinite as well as ill-conditioned (singularly perturbed) coercive problems. Greedy strategies based on such surrogates are then shown to constitute an optimal sampling strategy, i.e., the resulting snapshots span reduced spaces whose distances from the solution manifold decay essentially at the same rate as the Kolmogorov  $n$ -widths. This means, in particular, that stability constants need not be determined by additional typically expensive computations but can be pushed by the stabilizing inner greedy loop as close to one as one wishes. Finally, we have explained why the focus on uniform approximation of the entire solution manifold is equally relevant for applications where only functionals of the parameter dependent solutions have to be approximated.

## REFERENCES

- [1] Barrett, J.W., Morton, K.W.: Approximate symmetrization and Petrov-Galerkin methods for diffusion-convection problems. *Comput. Method. Appl. M.* 45 (1984), 97–12
- [2] A. Barron, A. Cohen, W. Dahmen, R. DeVore, Approximation and learning by greedy algorithms, *Annals of Statistics*, 3(No 1)(2008), 64–94.
- [3] R. Becker, R. Rannacher, An optimal error control approach to a-posteriori error estimation, *Acta Numerica* 2001, 1–102.

- [4] P.Binev, A.Cohen, W. Dahmen, R.DeVore, G. Petrova, P. Wojtaszczyk, Convergence Rates for Greedy Algorithms in Reduced Basis Methods, *SIAM J. Math. Anal.*, 43 (2011), 1457–1472.
- [5] F. Brezzi, M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Series in Computational Mathematics, Vol. 15, Springer-Verlag, 1991. ISBN: 978-1-4612-7824-5 (Print) 978-1-4612-3172-1 (Online)
- [6] A. Buffa, Y. Maday, A.T. Patera, C. Prud'homme, and G. Turinici, A Priori convergence of the greedy algorithm for the parameterized reduced basis, *ESAIM: Mathematical Modelling and Numerical Analysis*, 46 (03) (2012), 595-603.
- [7] A. Cohen, W. Dahmen, G. Welper, Adaptivity and Variational Stabilization for Convection-Diffusion Equations, *ESAIM: Mathematical Modelling and Numerical Analysis*, 46 (5) (2012), 1247 - 1273.
- [8] Cohen, A., DeVore, R., Schwab, C.: Convergence rates of best  $N$ -term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.* **10**, 615–646 (2010)
- [9] W.Dahmen, C.Huang, C.Schwab, G.Welper, Adaptive Petrov-Galerkin methods for first order transport equations, *SIAM J. Numer. Anal.*, 50(5) (2012), 2420–2445.
- [10] W. Dahmen, C. Plesken, G. Welper, Double Greedy Algorithms: Reduced Basis Methods for Transport Dominated Problems, *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(3) (2014), 623–663. DOI 10.1051/m2an/2013103
- [11] L. F. Demkowicz and J. Gopalakrishnan, A class of discontinuous Petrov-Galerkin Methods I: The transport equation, *Comput. Methods Appl. Mech. Engrg.*, 199 (No. 23-24) (2010), pp. 1558–1572.
- [12] L. Demkowicz, J. Gopalakrishnan, A class of discontinuous Petrov-Galerkin methods. Part II: Optimal test functions, *Numerical Methods for Partial Differential Equations*, 27 (No.1) (2011), pp. 70 - 105.
- [13] R. DeVore, G. Petrova, P. Wojtaszczyk, Greedy algorithms for reduced bases in Banach spaces, *Constr. Approx.* 37 (2013), 455–466.
- [14] A.-L. Gerner, K. Veroy. Reduced basis a posteriori error bounds for the Stokes equations in parameterized domains: A penalty approach, to appear in *Mathematical Models and Methods in Applied Sciences (M3AS)*.
- [15] A. Gerner, K. Veroy-Grepl, Certified reduced basis methods for parametrized saddle point problems, preprint 2012, to appear in *SIAM J. Sci. Comput.*
- [16] M.A. Grepl, A.T. Patera. A Posteriori Error Bounds for Reduced-Basis Approximations of Parametrized Parabolic Partial Differential Equations, *M2AN Mathematical Modelling and Numerical Analysis*, 2005, 39(1), pp. 157-181.
- [17] T. Hughes, G. Sangalli, Variational Multiscale Analysis: the Fine-scale Green's Function, Projection, Optimization, Localization, and Stabilized Methods, *SIAM Journal of Numerical Analysis*, Vol. 45, (No. 2) (2007), 539–557.
- [18] T. Manteuffel, S. McCormick, J. Ruge, and J. G. Schmidt, First-order system  $\mathcal{LL}^*$  (*FOSLL*)\* for general scalar elliptic problems in the plane, *SIAM Journal on Numerical Analysis*, 43 (2005), pp. 2098-2120.
- [19] P. Pacciarini, G. Rozza, Stabilized reduced basis method for parametrized advection–diffusion PDEs. *Computer Methods in Applied Mechanics and Engineering*, 274 (2014), 1D18.
- [20] A.T. Patera, G. Rozza, Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations, Version 1.0, Copyright MIT 2006–2007, to appear in (tentative rubric) MIT Pappalardo Graduate Monographs in Mechanical Engineering.
- [21] C. Pruh'homme, D.V. Rovas, K. Veroy, L. Machies, Y. Maday, A.T. Patera, G. Turinici, Reliable real-time solution of parametrized partial differential equations: reduced basis output-bound methods, *Transactions of the ASME*, 124 (2002), 70–80.
- [22] G. Rozza, K. Veroy, On the stability of reduced basis techniques for Stokes equations in parametrized domains, *Computer Methods in Applied Mechanics and Engineering*, (196), No.7 ( 2007), 1244–1260.
- [23] G. Rozza, D.B.P. Huynh, A.T. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations, *Arch. Comput. Methods Eng.*, 15 (2008), 229–275.
- [24] S. Sen, K. Veroy, D.B.P. Huynh, S. Deparis, N.C. Nguyen, A.T. Patera, “Natural norm” a-posteriori error estimators for reduced basis approximations, *Journal of Computational Physics*, 217 (2006), 37–62.
- [25] V. Temlyakov, Weak greedy algorithms, *Adv. Comp. Math.*, 12 (2000), 213–227.
- [26] R. Verfürth, Robust a posteriori error estimates for stationary convection-diffusion equations. *SIAM J. Numer. Anal.*, 43, No. 4 (2005), 1766–1782.
- [27] G. Sangalli, A uniform analysis of non-symmetric and coercive linear operators, *SIAM J. Math. Anal.*, 36 (no 6) (2005), 2033-2048.
- [28] G. Welper, Infinite dimensional stabilization of convection-dominated problems, PhD Thesis, RWTH Aachen, Nov. 2012.

