DOUBLE GREEDY ALGORITHMS: REDUCED BASIS METHODS FOR TRANSPORT DOMINATED PROBLEMS *

Wolfgang Dahmen[1], Christian Plesken[1] and Gerrit Welper[1]

Dedicated to Dietrich Braess on the occasion of his 75th birthday.

Abstract. The central objective of this paper is to develop reduced basis methods for parameter dependent transport dominated problems that are rigorously proven to exhibit rate-optimal performance when compared with the Kolmogorov n-widths of the solution sets. The central ingredient is the construction of computationally feasible "tight" surrogates which in turn are based on deriving a suitable well-conditioned variational formulation for the parameter dependent problem. The theoretical results are illustrated by numerical experiments for convection-diffusion and pure transport equations. In particular, the latter example sheds some light on the smoothness of the dependence of the solutions on the parameters.

1. Introduction

Over the past few years *model order reduction* has become an indispensable constituent of large scale design or optimization problems. In particular, the *Reduced Basis Method* (RBM) is perhaps by now one of the most important paradigms for highly complex *frequent query problems* involving parameter dependent PDEs, see *e.g.* [25, 28, 32]. Among other things, at least under certain circumstances, modeling errors are rigorously controlled and can be upgraded if necessary.

While the development of RBMs has been a very active area with impressive success stories in by now a variety of important application fields, it is fair to say that a theoretical underpinning of what one might call "near-optimal performance" – in a sense to be made precise later – is still confined to a relatively narrow problem class. The central purpose of this paper is therefore to extend the scope of problems for which RBMs can be developed and rigorously proven to perform in that near optimal sense. The focus of the present work is on performance in terms of the *accuracy* offered by the reduced model, roughly speaking, centering around the

[1] Institut für Geometrie und Praktische Mathematik, RWTH Aachen, Templergraben 55, 52056 Aachen, Germany.
dahmen@igpm.rwth-aachen.de; plesken@igpm.rwth-aachen.de; welper@igpm.rwth-aachen.de

question how to ensure any certified target tolerance of the reduced model by a possibly small number of reduced basis functions, of course, always insisting on the standard *offline-online* division of the overall computational work.

1.1. General framework

Suppose that $B_\mu : X \to Y'$, $\mu \in \mathcal{P}$, is a family of (linear) operators from a Hilbert space X onto the dual Y' of another Hilbert space Y, depending on parameters μ from a compact set $\mathcal{P} \subset \mathbb{R}^p$. Under appropriate conditions on $\{B_\mu\}_{\mu \in \mathcal{P}}$ the solution set

$$\mathcal{M} := \{p(\mu) = B_\mu^{-1} f : \mu \in \mathcal{P}\} \subset X \tag{1.1}$$

for the family of operator equations

$$B_\mu p(\mu) = f, \quad \mu \in \mathcal{P}, \tag{1.2}$$

is a compact subset of X. In the context of frequent query problems, like steering a functional $\ell(p(\mu))$ of the solution towards a target value, RBMs try to exploit the fact that $\mathcal{M}$ may be a very thin subset of X. In fact, compactness of $\mathcal{M}$ means that the *Kolmogorov n-widths*

$$d_n(\mathcal{M})_X := \inf_{\dim V \leq n} \max \operatorname{dist}(\mathcal{M}, V)_X, \tag{1.3}$$

tend to zero as $n \to \infty$, where V is taken from the set of *all* n-dimensional subspaces of X and

$$\max \operatorname{dist}(\mathcal{M}, X_n)_X := \sup_{p \in \mathcal{M}} \inf_{q \in X_n} \|p - q\|_X.$$

The objective is then to construct (problem dependent) subspaces $X_n \subset X$ of possibly small dimension n such that for a given *target accuracy* tol, say,

$$\max \operatorname{dist}(\mathcal{M}, X_n)_X \leq \operatorname{tol} \tag{1.4}$$

is *guaranteed* to hold. In particular, this implies that for any $p \in \mathcal{M}$ and any bounded linear functional $\ell \in X'$, a trivial estimate immediately gives $|\ell(p) - \ell(P_{X,X_n}p)| \leq \|\ell\|\operatorname{tol}$ (which could even be improved by duality arguments, see *e.g.* [32]), where P_{X,X_n} is the X-orthogonal projection onto X_n.

Of course, a key question is how to practically construct spaces X_n warranting (1.4) for possibly small n. A common strategy of essentially all RBMs is the following. Given X_n, find a *surrogate* $R(\mu, X_n)$, $\mu \in \mathcal{P}$, such that

$$\|p(\mu) - P_{X,X_n}p(\mu)\|_X \leq C_R R(\mu, X_n) \tag{1.5}$$

holds for some constant C_R independent of μ and n. Here it is crucial that the evaluation of $R(\mu, X_n)$ is sufficiently efficient so that the maximization of $R(\mu, X_n)$ over $\mu \in \mathcal{P}$ is computationally feasible. Then perform the *greedy algorithm* **GA** based on this surrogate, as described in Algorithm 1.

We have ignored for the moment the fact that the snapshots $p(\mu_n)$ can, of course, not be computed exactly but only approximately within some sufficiently large but finite dimensional "truth space".

To see whether such a greedy space search produces good reduced models one can compare them with the "best possible" spaces. Clearly, the n-width $d_n(\mathcal{M})_X$ from (1.3) is a *lower bound* for the accuracy attainable by any RBM, *i.e.*,

$$d_n(\mathcal{M})_X \leq \sigma_n(\mathcal{M})_X := \sup_{\mu \in \mathcal{P}} \|p(\mu) - P_{X,X_n}p(\mu)\|_X. \tag{1.7}$$

Unfortunately, in general it seems to be impossible to compute the precise *optimal* subspaces for which the n-width is attained. Nevertheless, the closer $\sigma_n(\mathcal{M})_X$ is to $d_n(\mathcal{M})_X$ the better the choice of X_n.

To see what can be achieved in this regard, recall from [1,3] that even when $R^*(\mu, X_n) := \|p(\mu) - P_{X,X_n}p(\mu)\|_X$ is the *ideal* surrogate, in a direct comparison $\sigma_n(\mathcal{M})_X \leq K_n d_n(\mathcal{M})_X$ the constant K_n can be as large as 2^n. Nevertheless, the following more favorable results in terms of *convergence rates* hold for surrogates that are *tight*, *i.e.*, if in addition to the upper bound (1.5) it uniformly sandwiches the exact distance.

Algorithm 1. Greedy algorithm

1: **function GA**
2:     Set $X_0 := \{0\}$, $n = 0$,
3:     **while** $\mathrm{argmax}_{\mu \in \mathcal{P}} R(\mu, X_n) \geq tol$ **do**
4:

$$\mu_{n+1} := \underset{\mu \in \mathcal{P}}{\mathrm{argmax}}\, R(\mu, X_n),$$

$$p_{n+1} := p(\mu_{n+1}),$$ (1.6)

$$X_{n+1} := \mathrm{span}\left\{X_n, \{p(\mu_{n+1})\}\right\} = \mathrm{span}\left\{p_1, \ldots, p_{n+1}\right\}$$

5:     **end while**
6: **end function**

Definition 1.1. We call the surrogate $R(\mu, X_n)$, *tight* if there exist positive constants c_R, C_R such that

$$c_R R(\mu, X_n) \leq \|p(\mu) - P_{X, X_n} p(\mu)\|_X \leq C_R R(\mu, X_n),$$ (1.8)

uniformly in $\mu \in \mathcal{P}$. Moreover, we call

$$\kappa(R) := \inf \left\{ C_R/c_R : c_R, C_R \text{ satisfiy } (1.8) \text{ for all } \mu \in \mathcal{P}, n \in \mathbb{N} \right\},$$ (1.9)

the *condition* of the surrogate R.

Remark 1.2. As already observed in [1] whenever the surrogate is tight, *i.e.* (1.8) holds, then the snapshots $p_n = p(\mu_n)$ from (1.6) satisfy the *weak greedy condition*

$$\|p_n - P_{X, X_n} p_n\|_X \geq \kappa(R)^{-1} \max \mathrm{dist}\,(\mathcal{M}, X_n)_X, \quad n \in \mathbb{N},$$ (1.10)

where $\kappa(R)$ is given by (1.9).

The following statements are then readily derived from the results in [1, 11].

Theorem 1.3. *Assume that the spaces X_n are obtained through a greedy algorithm* **GA***, (1.6) based on tight surrogates. Then, if $d_n(\mathcal{M})_X = O(n^{-\alpha})$, for some $\alpha > 0$ or if $d_n(\mathcal{M})_X = O(e^{-cn^\alpha})$, for some $c, \alpha > 0$, one has*

$$\max \mathrm{dist}\,(\mathcal{M}, X_n)_X = O(n^{-\alpha}), \qquad \max \mathrm{dist}\,(\mathcal{M}, X_n)_X = O(e^{-\tilde{c}n^\alpha}), \qquad n \to \infty,$$ (1.11)

respectively, where the constants depend on α, c, and $\kappa(R)$ with exact specification given in [1, 11]. Moreover, these bounds remain valid up to the tolerance tol^* when all computations are carried out within this accuracy.

We call an RBM *rate-optimal* if the generated spaces X_n satisfy "Kolmogorov optimal" bounds of the type (1.11).

There are two important points to be drawn from these results that guide the subsequent developments. The first one is: although dispensing with the (infeasible) ideal surrogate $R^*(\mu, X_n) := \|p(\mu) - P_{X, X_n} p(\mu)\|_X$, a tight surrogate still ensures that the accuracy provided by the reduced bases is in terms of rates still essentially as good as that of the "Kolmogorov-best" subspaces. The second point is *quantitative*. It is absolutely vital to make sure that the condition $\kappa(R)$ stays as small as possible. In fact, a look at the dependence of the constants in (1.11) on $\kappa(R)$ (see [1, 11]) reveals that the closer $\kappa(R)$ is kept to one, the better is the accuracy of the reduced spaces, in comparison with the best spaces, already for a *small reduced dimension*, which is at the heart of model reduction.

Hence, the central objective of this paper is to develop a rigorous conceptual framework to obtain practically feasible tight surrogates whose *condition* $\kappa(R) \leq C_R/c_R$ is as close to one as possible, in particular, for problem classes for which this is currently not known.

1.2. Objectives and layout

To provide an orientation for subsequent developments the corresponding *ideal scenario* and the corresponding basic mechanisms are briefly recalled in Section 2.2. It is by and large confined to problems that are uniformly elliptic with respect to the parameters. The perhaps next best understood case is the reduction of a parabolic problem to a sequence of elliptic problems [15–17], where however, the lower bound in (1.8) – and hence tightness – seems to be missing. This has been recently significantly improved in [24] using a space-time variational formulation. Moreover, important progress has been made in [13, 14, 27, 29] developing RBMs for specific saddle point – hence indefinite – problems such as the Stokes system. In particular, in the present terminology stability and, as a consequence, tight surrogates are obtained by enriching the velocity spaces by *supremizers*. More precisely, there are two approaches. For standard affine parameter dependence of the involved bilinear forms one can determine *a priori* an enrichment, depending on the number of terms in the bilinear forms, that ensures that the infinite dimensional inf-sup-constant is preserved, see [14, 27, 29]. Since the number of these supremizers is possibly quite large, as an alternative, it is proposed in [13] to *adaptively* add supremizers until a desired inf-sup-stability is reached. It is observed experimentally that in the tested examples this adaptive enrichment results in an overall much smaller number of stabilizing functions although the actual guaranteed termination of such a procedure has apparently not been discussed. Although termination in the context treated in [13, 14, 27, 29] is apparent, we shall encounter situations where this is no longer the case. Nevertheless, relating also the stabilizing enrichments to greedy approximations allows us to treat this case as well, see Section 4.4.

Although the present paper addresses a rather different problem class the treatment of saddle point problems turns out to be an important point of contact. In fact, the stabilizing enrichment of the reduced velocity spaces by adding supremizers can be viewed as a special instance of the interior loop of what we call here *double greedy schemes*, presented first at a workshop in Paris, 2011 [6]. The central objective of this paper is in fact to develop rate-optimal RBMs – *viz.* identify well-conditioned tight surrogates – for a much wider scope of problems, including indefinite, unsymmetric and singularly perturbed problems, in particular, *transport dominated* problems. The development of RBMs for problems of this latter type, even in nonlinear and time dependent formulations, have been recently addressed in [10, 23]. Resorting to the concept of *natural norms* (see [32]) the deficiencies of standard H^{-1}-based surrogates are shown to be mitigated but apparently not cured.

Although we employ an analog to the notion of "natural norm" we present in this paper a different approach based on deriving *well-conditioned, possibly unsymmetric variational formulations* for such problems. An essential difference from [10, 23, 32] lies in guaranteeing inf-sup constants close to one not only for the infinite dimensional problem but also for the reduced problems. In fact, in contrast to [10, 23, 32] our method can be interpreted as (implicitly) generating for a given reduced space a "near-optimal test space" in a *Petrov−Galerkin* sense which gives rise to inf-sup constants that can be driven towards one, see Section 3.2. In addition to just using natural norms for X these controlled inf-sup constants allow one to approximately realize the projector $P_{X,X_n}p(\mu)$ in the Definition (1.1) of tight surrogates as well as to stably approximate the solution in the online phase. The starting point is an *"optimal pair" of norms* for the infinite dimensional problem, one of which can be prescribed, such that the operator induced by the corresponding variational formulation is even an isometry.

For the sake of orientation, we formulate in Sections 2.3.1, 2.3.2, two simple model problems that nevertheless exhibit increasing levels of obstructions. In particular, the case of *pure transport* plays an extreme role in several respects. There is no viscosity that can be used for stabilization. Moreover, depending on the data, the dependence of the solution on the parameter, here the convection field, may become very unsmooth which, in spite of perfect stability, hints at a principal limitation of the RBM-concepts in such scenarios. We emphasize though that the general methodology presented below is not restricted to those problems at all. The main features of this approach, can be summarized as follows:

(i)   Tight *a posteriori* bounds for the truth spaces as well as reduced spaces warrant certification. In particular, truth and reduced spaces can be upgraded without discarding prior computations, see the robustness results in [1].

(ii) While remaining feasible in the sense of an online/offline decomposition through a built in stabilization loop, the scheme automatically gives rise to stability constants that can, in principle, be made arbitrarily close to one, see also (5.7) and Section 6.

(iii) Viewing time as an additional "spatial" variable, the results can be applied to time dependent problems through corresponding space-time discretizations, which is one reason to focus on transport problems, see [7].

In summary, the particular variational formulations presented in Section 3 combined with certain stabilization techniques optimally inherits the analytic structure of the underlying infinite dimensional problem to the reduced model.

Section 5 is then devoted to the algorithmic development and analysis of a *double greedy* scheme giving rise to rate-optimal RBMs.

The theoretical findings are then applied in Section 6 to the two model problems concerning convection-diffusion and pure transport equations. First numerical experiments quantify the results and highlight several particular obstructions.

In Section 7 we apply the (slightly modified) scheme to other types of saddle point problems not necessarily stemming from the generation of well-conditioned variational formulations. As a simple consequence we obtain rate-optimality also for the problems considered in [13, 14, 29].

To simplify the exposition we write $a \lesssim b$ to express that a is bounded by some constant multiple of b independent on any parameters a, b may depend on. Likewise $a \sim b$ means $a \lesssim b$ and $b \lesssim a$.

2. Conceptual preview

2.1. Feasibility

In all subsequent developments we will be dealing exclusively with *affine* parameter dependence, see *e.g.* [28]. Under this assumption we insist on the usual division of the computational work into an *offline* and *online* mode. Solving a problem in the full space X, which is typically computationally very intense, happens only in offline mode where it is understood that actual computations take place in some sufficiently large but finite dimensional subspace $X_\mathcal{N}$ of X which is commonly referred to as the "truth space". Typically $X_\mathcal{N}$ is chosen so as to guarantee

$$\sup_{p \in \mathcal{M}} \inf_{v \in X_\mathcal{N}} \|p - v\|_X \leq \mathrm{tol}^*, \tag{2.1}$$

for some tolerance tol^* that is sufficiently small for the application at hand. The subscript $\mathcal{N}$ refers to the truth space and is sometimes suppressed when there is no risk of confusion. The greedy search for the reduced basis functions falls therefore into the offline mode. This requires evaluating the surrogate for a sufficiently large *training set* of parameters which for simplicity we also denote by $\mathcal{P}$. In what follows, we call the surrogate *feasible* if each evaluation of the surrogate requires solving only a problem in the small current reduced space X_n. We sometimes say then that the offline mode is (computationally offline) feasible.

Likewise, the *online evaluation* is called *feasible* if each reduced basis approximation of some $p(\mu)$ requires solving only a "small" problem of dimension n in the reduced space X_n. In this mode solving a "large" problem in $X_\mathcal{N}$ is prohibited.

Note that a feasible surrogate is not allowed to explicitly contain the true solution $p(\mu)$ (in the truth space). This is why one is essentially forced to resort to *residuals* to estimate the true error, which in turn requires a tight *error-residual* relation.

2.2. The ideal setting

To clearly identify the mechanisms leading to *residual based* tight surrogates, we briefly revisit in this subsection elliptic problems. This will later guide the realization of the essential ingredients also for unsymmetric and transport dominated problems.

To this end, let $b_\mu(\cdot, \cdot) : X \times X \to \mathbb{R}$ be a symmetric uniformly X-elliptic bilinear form and $\ell \in X'$, *i.e.*

$$c_a \|q\|_X^2 \leq b_\mu(q, q), \quad b_\mu(p, q) \leq C_a \|p\|_X \|q\|_X, \quad p, q \in X, \ \mu \in \mathcal{P}, \tag{2.2}$$

holds uniformly in $\mu \in \mathcal{P}$. For compact $\mathcal{P}$ one obtains a compact solution set $\mathcal{M} \subset X$ for: given $\ell \in X'$, find $p(\mu) \in X$ such that

$$b_\mu(p(\mu), q) = \langle \ell, q \rangle, \quad q \in X.$$

There are two key properties that ensure rate-optimality in this setting:

(MP) Mapping property. The operator B_μ, defined by $\langle B_\mu p, q \rangle = b_\mu(p, q)$, $p, q \in X$, is for each $\mu \in \mathcal{P}$ an isomorphism from X onto X', *i.e.*

$$c_a \|p\|_X \leq \|B_\mu p\|_{X'} \leq C_a \|p\|_X, \quad \text{uniformly in } \mu \in \mathcal{P}. \tag{2.3}$$

In other words, errors measured in the "energy norm" $\|\cdot\|_X$ are equivalent to residuals in the dual norm $\|\cdot\|_{X'}$.

(BAP) Best Approximation Property. The Galerkin projection to the current reduced space, which can be done in online mode, produces, up to constants, a best approximation with respect to the X-norm, *i.e.*

$$\|p(\mu) - \Pi_{\mu, X_n} p(\mu)\|_X \leq \sqrt{\frac{C_a}{c_a}} \inf_{\varphi \in X_n} \|p(\mu) - \varphi\|_X$$

where Π_{μ, X_n} denotes the Galerkin-projector onto X_n defined by

$$b_\mu(p(\mu), q) = b_\mu(\Pi_{\mu, X_n} p(\mu), q), \quad q \in X_n.$$

Note that **(BAP)** is a direct consequence of Cea's Lemma for symmetric problems. Now the mapping property **MP** yields for $p_n(\mu) := \Pi_{\mu, X_n} p(\mu)$

$$\|p(\mu) - P_{X, X_n} p(\mu)\|_X \sim \sup_{q \in X} \frac{\langle \ell, q \rangle - b_\mu(p_n(\mu), q)}{\|q\|_X} := R(\mu, X_n). \tag{2.4}$$

Thus, **MP** and **BAP** imply that the residual based surrogate, defined by (2.4), is tight, while the computation of $p(\mu)$ is completely avoided but traded against the cheap computation of the Galerkin projection in X_n. However, the condition $\kappa(R)$ of the surrogate (see (1.9)) depends on the *condition number* $\kappa_{X,X}(B_\mu) \leq C_a/c_a$ (see (2.2)) of the operator B_μ, which should therefore be of moderate size.

Finally, feasibility of the surrogate in (2.4) is well-known (see *e.g.* [28]) to be ensured when the parameter dependence of $b_\mu(\cdot, \cdot)$ is *affine*, see (3.43) below.

Note that for unsymmetric problems, treated later, the validity of **(BAP)** is *in general not* automatic even when **MP** holds and therefore plays a pivotal role in subsequent discussions.

2.3. Two model problems

As soon as one leaves the elliptic setting **MP**, **BAP**, and tightness of residual based surrogates, are no longer for free. In particular, so far *well-conditioned* tight surrogates do *not* seem to be available yet for many unsymmetric PDEs like convection dominated or pure transport problems. We shall discuss two model problems that bring out several principal obstructions. The first example concerns convection-diffusion equations for which, in principle, classical variational formulations are available. The second example concerns pure transport for which a "natural" variational formulation is less obvious and for which the parameter dependence of the solutions turns out to be less regular. Perhaps more importantly, the two examples represent two different scenarios regarding the *spaces* associated with the bilinear form $b_\mu(\cdot, \cdot)$, an issue that has apparently not been addressed in the RBM context.

2.3.1. Convection-diffusion equations

As a first example we consider the linear *convection-diffusion* equation

$$-\mathrm{div}(\epsilon \nabla u) + b(\mu) \cdot \nabla u + cu = f, \quad \text{in } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{2.5}$$

where for simplicity we assume for now that only the convection $b(\mu)$ depends on a parameter μ while ϵ could be *arbitrarily small*. We could as well include the viscosity and the reaction term varying in suitable regimes. Its classical weak formulation is

$$b_\mu(p, q) := \epsilon(\nabla p, \nabla q) + (b(\mu) \cdot \nabla p, q) + (cp, q) = \langle f, q \rangle, \quad q \in X = H_0^1(\Omega). \tag{2.6}$$

It is well-known that $b(\mu) \in W^{1,\infty}(\Omega)^d$, $c \in L_\infty(\Omega)$, $\mu \in \mathcal{P}$, such that

$$-\frac{1}{2}\,\mathrm{div}\,b(\mu) + c \geq 0, \tag{2.7}$$

implies *well-posedness* of (2.6) in the sense that the induced operator $B_\mu : H_0^1(\Omega) \to (H_0^1(\Omega))'$ is an isomorphism, *i.e.*, there exists for each $\mu \in \mathcal{P}$ a unique solution $p(\mu)$ to (2.6) in $H_0^1(\Omega)$. However, although (2.2) is still valid, the condition number $\kappa_{H_0^1(\Omega), H_0^1(\Omega)}(B_\mu)$ behaves like the Péclet number $|b(\mu)|_\infty/\epsilon$ and hence is unacceptably large for strongly dominating convection. As a consequence, in this case the condition $\kappa(R)$ of the corresponding surrogate (2.4) based on the $H^{-1}(\Omega)$-residual grows with the Péclet number. Hence, although such a surrogate is theoretically tight, as long as $\epsilon \geq \epsilon_0$ where $\epsilon_0 > 0$ is fixed, the condition $\kappa(R)$ (see (1.9)) is so large, that, due to the constants in (1.11), one can expect essentially no control of the quality of the reduced spaces for very small ϵ_0 and moderate n.

Therefore, we are mainly interested here in a robust treatment of *arbitrarily* large Péclet numbers $|b(\mu)|_\infty/\epsilon$ which to our knowledge is currently not well covered by RBM methodology.

Unfortunately, an easy cure based on the standard (mesh-dependent) stabilization methods such as SUPG (see *e.g.* [26] for a survey) does not give rise to an error-residual relation that stays independent of the Péclet number $|b|_\infty/\epsilon$ either.

Instead we pursue here a different line based on stabilizing the problem on the *infinite dimensional* level which, in particular, involves *unsymmetric* variational formulations, *i.e.*, $b_\mu(\cdot, \cdot)$ is viewed as a bilinear form on a pair of (possibly) *different* and *parameter dependent* Hilbert spaces X_μ, Y_μ, $\mu \in \mathcal{P}$.

2.3.2. Linear transport equations

In some sense the situation is even aggravated when the diffusion vanishes completely as in pure *parametric transport equations* forming the core ingredient of Boltzmann equations and related kinetic models as well as kinetic formulations of conservation laws. Already the simplest version of a (stationary) linear transport equation

$$\mu \cdot \nabla p + cp = f, \quad \text{in } \Omega, \quad p = p_b, \quad \text{on } \Gamma_-(\mu), \tag{2.8}$$

will be seen to represent the "worst scenario" from the RBM perspective, where, denoting by $n(x)$ the outward normal at the point x,

$$\Gamma_-(\mu) := \{x \in \partial\Omega : n(x) \cdot \mu < 0\},$$

is the inflow boundary for the given convection vector μ. An example of a parameter domain would be the sphere S^{d-1} appearing in radiative transfer models, see [19,31]. It will be seen that the two examples differ in a subtle but essential way, in particular, regarding smoothness of the dependence of the solutions on the parameter.

A possible variational formulation of (2.8) can be found in [12]. In order to eventually apply the n-width benchmark, it is preferable to measure all parameter dependent solutions in a single reference norm. Therefore we employ here a slightly different variational formulation from [7]: multiplying (2.8) by a test function and integrating by parts, yields

$$(p, -\mu \cdot \nabla q + cq) + \int_{\partial\Omega \setminus \Gamma_-} n \cdot \mu pq = \langle f, q \rangle - \int_{\Gamma_-} n \cdot \mu pq.$$

If we now take test functions q that vanish on $\partial\Omega \setminus \Gamma_-$ the boundary integral on the left hand side is zero. Furthermore, we may replace the function p in the boundary integral on the right hand side by the boundary condition p_b so that we obtain

$$b_\mu(p,q) := \langle p, -\mu \cdot \nabla q + cq \rangle = \langle f, q \rangle - \int_{\Gamma_-} n \cdot \mu p_b q. \tag{2.9}$$

For this variational formulation it is natural to define the function spaces

$$Y_\mu := \mathrm{clos}_{\|\cdot\|_{Y_\mu}} \left\{ q \in C^\infty(\Omega) : q|_{\partial\Omega \setminus \Gamma_-} = 0 \right\}, \quad X_\mu := L_2(\Omega) \tag{2.10}$$

endowed with the norms

$$\|q\|_{Y_\mu} := \|B_\mu^* q\|_{L_2}, \quad \|p\|_{X_\mu} := \|p\|_{L_2}. \tag{2.11}$$

It is shown in [7] that the operator B_μ induced by $b_\mu(\cdot,\cdot)$, is an isomorphism $B_\mu : X_\mu \to Y_\mu'$ so that (2.11) indeed defines a norm.

Remark 2.1. Notice that the spaces Y_μ differ even as sets for different μ. Moreover, in contrast to the previous example we must have $X_\mu \neq Y_\mu$ here.

3. ROBUST ERROR-RESIDUAL MAPPINGS FOR UNSYMMETRIC PROBLEMS

3.1. The basic principle for MP

In the following, we consider general bilinear forms $b_\mu(\cdot,\cdot) : X_\mu \times Y_\mu \to \mathbb{R}$ for possibly *parameter dependent* Hilbert spaces X_μ and Y_μ giving rise to what one may call an infinite dimensional *Petrov−Galerkin* formulation where the *trial space* X_μ generally differs from the *test space* Y_μ. Thus, the operator B_μ given by $\langle B_\mu q, v \rangle = b_\mu(q,v)$, $q \in X_\mu$, $v \in Y_\mu$, is now viewed as a mapping from X_μ to Y_μ'. In accordance with the preceding examples, we shall assume that this operator is actually an *isomorphism*, *i.e.*, the operator equation

$$B_\mu p(\mu) = f, \tag{3.1}$$

has for any $f \in Y_\mu'$ a unique solution in X_μ. It is well-known that the mapping properties of B_μ are quantified by Babuska's Theorem: if there exist constants $0 < \beta(\mu), C_b(\mu) < \infty$ such that

$$\inf_{q \in X_\mu} \sup_{v \in Y_\mu} \frac{b_\mu(q,v)}{\|q\|_{X_\mu}\|v\|_{Y_\mu}} \geq \beta(\mu), \quad \sup_{q \in X_\mu} \sup_{v \in Y_\mu} \frac{|b_\mu(q,v)|}{\|q\|_{X_\mu}\|v\|_{Y_\mu}} \leq C_b(\mu) \tag{3.2}$$

and for every $v \in Y_\mu$ there exists a $q \in X_\mu$ such that $b_\mu(q,v) \neq 0$, then one has $\kappa_{X_\mu,Y_\mu}(B_\mu) \leq C_b(\mu)/\beta(\mu)$.

Let us stress that, as mentioned before, the subscript μ indicates that spaces and norms *may* depend on the parameter but need not do so. Specifically, the generic notational convention adopted here covers the following scenarios.

(s1) The norms $\|\cdot\|_{X_\mu} \|\cdot\|_{Y_\mu}$ (possibly independent of μ) correspond to a given variational formulation for which $C_b(\mu)/\beta(\mu)$ is *known* to have moderate size uniformly in $\mu \in \mathcal{P}$. Examples are (2.6) with small Peclet number or saddle point problems like the Stokes system. In the latter case, while (MP) may be satisfactory, still additional care must be taken to ensure **BAP** with small constants.

(s2) As in the case of convection dominated convection-diffusion equations $\kappa_{X_\mu,Y_\mu}(B_\mu)$ could be very large, severely degrading a greedy selection of snapshots in an RBM. In this case the pair of norms $\|\cdot\|_{X_\mu} \|\cdot\|_{Y_\mu}$ has to be modified. The first goal of this subsection is to describe how to "stabilize" the problem on the *infinite dimensional* level which could be viewed as preconditioning (3.1). The underlying basic principle has been used before in several works for different purposes, see *e.g.* [8, 9, 22, 34] and is also a point of contact with the concept of natural norms from [32]. Here our main orientation is taken from [5, 7, 34]. We briefly rehash the essential facts in order to bring in an additional *new* element, namely the interrelation of *Petrov−Galerkin* schemes and *associated saddle point problems*, which plays an essential role for eventually constructing uniformly well-conditioned tight surrogates.

(s3) There may initially be no "natural" initial variational formulation as in the example of the transport equation. The procedure applied in (s2) can be used to *create* for each μ a pair of spaces X_μ, Y_μ with norms $\| \cdot \|_{X_\mu} \| \cdot \|_{Y_\mu}$ so that $\kappa_{X_\mu, Y_\mu}(B_\mu) = 1$ and no need for further modifying the norm arises.

This section is devoted to developing the conceptual tools for coping with (s1)–(s3). We proceed with collecting a few useful preliminaries. It will be useful to identify for a given $q \in X_\mu$ the *supremizer* v_q for which $\sup_{v \in Y_\mu} b_\mu(q, v)/\|v\|_{Y_\mu}$ is attained, see *e.g.* [14, 25].

Remark 3.1. For every $q \in X_\mu$ the optimal test function is given by

$$v_q := \underset{v \in Y_\mu}{\mathrm{argmax}} \frac{b_\mu(q, v)}{\|v\|_{Y_\mu}} = R_{Y_\mu}^{-1} B_\mu q, \tag{3.3}$$

where $R_{Y_\mu} : Y_\mu \to Y_\mu'$ is the Riesz-map defined by

$$\langle R_{Y_\mu} v, w \rangle = (v, w)_{Y_\mu}, \quad v, w \in Y_\mu, \qquad \| \cdot \|_{Y_\mu}^2 = (\cdot, \cdot)_{Y_\mu}. \tag{3.4}$$

Hence, in particular, one has

$$\inf_{q \in X_\mu} \frac{\|R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}}{\|q\|_{X_\mu}} = \inf_{q \in X_\mu} \sup_{v \in Y_\mu} \frac{b_\mu(q, v)}{\|q\|_{X_\mu} \|v\|_{Y_\mu}}.$$

For convenience we recall the simple argument. Written in variational form, the supremizer is defined by $(v_q, w)_{Y_\mu} = b_\mu(q, w)$ for all $w \in Y_\mu$, which yields

$$\sup_{v \in Y_\mu} \frac{b_\mu(q, v)}{\|v\|_{Y_\mu}} = \sup_{v \in Y_\mu} \frac{(v_q, v)_{Y_\mu}}{\|v\|_{Y_\mu}} = \|v_q\|_{Y_\mu},$$

which readily confirms the claim.

Although for most of the following considerations the dependence of the involved bilinear forms on the parameter $\mu \in \mathcal{P}$ is irrelevant it will be convenient for later purposes to retain the generic parameter dependence in the notation.

Renormation. The possible ill-conditioning reflected by a very large $\kappa_{X,Y}(B) \leq C_b/\beta$ in (3.2) can be remedied by properly modifying one of the two norms $\| \cdot \|_{Y_\mu}$ or $\| \cdot \|_{X_\mu}$ while keeping the other one fixed. Specifically, we wish to choose an equivalent but possibly different norm $\| \cdot \|_{\hat{X}_\mu}$ for X so that ideally $C_b(\mu) = \beta(\mu) = 1$, which then means that

$$\|p(\mu) - q\|_{\hat{X}_\mu} = \|f - B_\mu q\|_{Y_\mu'}, \quad q \in X_\mu, \ \mu \in \mathcal{P}. \tag{3.5}$$

In this event the *residual* of a best approximation would be an *ideal* surrogate even sparing one the computation of stability constants for the error certification.

Our starting point is exactly this latter *ideal error-residual relation*. Specifically, given $\| \cdot \|_{Y_\mu}$, we endow now X_μ with a new norm $\| \cdot \|_{\hat{X}_\mu}$, defined by

$$\|p\|_{\hat{X}_\mu} := \sup_{v \in Y_\mu} \frac{b_\mu(p, v)}{\|v\|_{Y_\mu}} = \|B_\mu p\|_{Y_\mu'} = \|R_{Y_\mu}^{-1} B_\mu p\|_{Y_\mu}, \quad p \in X_\mu, \ \mu \in \mathcal{P}, \tag{3.6}$$

which corresponds to the concept of natural norms in [32]. Note that this is indeed a well-defined norm because $R_{Y_\mu}^{-1} B_\mu : X_\mu \to Y_\mu$ is an isomorphism, hence injective, and that because of $\|R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}^2 = \langle B_\mu^* R_{Y_\mu}^{-1} B_\mu q, q \rangle$,

the corresponding Riesz map $R_{\hat{X}_\mu} : X_\mu \to X'_\mu$ is given by $R_{\hat{X}_\mu} := B_\mu^* R_{Y_\mu}^{-1} B_\mu$. In addition, this shows that the $\hat{X}_\mu$-norm is equivalent to the original norm, *i.e.*, there are $c_M, C_M > 0$ such that

$$c_M \|q\|_{X_\mu} \le \|q\|_{\hat{X}_\mu} \le C_M \|q\|_{X_\mu}, \quad q \in X_\mu, \ \mu \in \mathcal{P}. \tag{3.7}$$

Note also that (2.11) is a special case of (3.6), where

$$R_{Y_\mu} = B_\mu B_\mu^*, \tag{3.8}$$

and thus $\| \cdot \|_{\hat{X}_\mu} = \|(B_\mu B_\mu^*)^{-1} B_\mu \cdot \|_{Y_\mu} = \| \cdot \|_{L_2} = \| \cdot \|_{X_\mu}$.

Remark 3.2. For the $\| \cdot \|_{\hat{X}_\mu}$ norm one has optimal continuity and stability constants $C_b(\mu) = \beta(\mu) = 1$, $\mu \in \mathcal{P}$, *i.e.*

$$\sup_{q \in \hat{X}_\mu} \sup_{v \in Y_\mu} \frac{b_\mu(q, v)}{\|v\|_{Y_\mu} \|q\|_{\hat{X}_\mu}} = \inf_{q \in \hat{X}_\mu} \sup_{v \in Y_\mu} \frac{b_\mu(q, v)}{\|v\|_{Y_\mu} \|q\|_{\hat{X}_\mu}} = 1. \tag{3.9}$$

Hence, $\kappa_{\hat{X}_\mu, Y_\mu}(B_\mu) = 1$, *i.e.*, B_μ is an isometry for these norms, which is the desired robust – in fact optimal – error-residual relation (3.5) **MP**. We call a pair of norms for which (3.9) holds *optimal*.

Proof. The first relation follows from

$$|b_\mu(q, v)| = |\langle R_{Y_\mu}^{-1} B_\mu q, R_{Y_\mu} v \rangle| \le \|R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu} \|R_{Y_\mu} v\|_{Y'_\mu} = \|q\|_{\hat{X}_\mu} \|v\|_{Y_\mu}.$$

On the other hand, note that for any $q \in X_\mu$ its supremizer $v_q := R_{Y_\mu}^{-1} B_\mu q \in Y_\mu$ gives by (3.6), (3.4), $b_\mu(q, v_q) = \langle B_\mu q, R_{Y_\mu}^{-1} B_\mu q \rangle = \|q\|_{\hat{X}_\mu}^2$ and $\|v_q\|_{Y_\mu} = \|q\|_{\hat{X}_\mu}$ so that

$$\inf_{q \in X_\mu} \sup_{v \in Y_\mu} \frac{b_\mu(q, v)}{\|q\|_{\hat{X}_\mu} \|v\|_{Y_\mu}} \ge \inf_{q \in X_\mu} \frac{\langle B_\mu q, R_{Y_\mu}^{-1} B_\mu q \rangle}{\|q\|_{\hat{X}_\mu}^2} = 1, \tag{3.10}$$

which completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 3.3. Likewise one can choose a particular test-norm and adapt the trial norm for obtaining an isometry by setting

$$\|v\|_{\hat{Y}_\mu} := \sup_{q \in X_\mu} \frac{b_\mu(q, v)}{\|q\|_{X_\mu}}. \tag{3.11}$$

Hence, one can also *prescribe* a desirable X_μ-norm while now a properly adapted $\hat{Y}_\mu$-norm ensures that B_μ is an isometry. Since both versions are mathematically equivalent (see [5,34]) we continue to express renormation by (3.6).

3.2. Petrov−Galerkin and saddle point problems

Even when **MP** holds the validity of **BAP** is no longer automatic for unsymmetric or indefinite problems. In principle, it can be guaranteed through contriving suitable *Petrov Galerkin* discretizations. This is where our approach differs in an essential way from the use of the natural norms in [10,32]. A central issue in this section is to show that such Petrov−Galerkin schemes are *equivalent* to certain saddle-point problems. On the one hand, this avoids the explicit computation of the respective test spaces which could be parameter dependent and hence render the scheme infeasible in the RBM sense. On the other hand, the stability of the Petrov−Galerkin scheme is tantamount to the inf-sup stability of the saddle point formulation which in the end turns out to be constructively realizable in a feasible way.

To this end, let $W \subset X_\mu$ be a "generic" trial space which will play several different roles. It may stand for the full infinite dimensional space, or for the truth space, or eventually for the reduced space. Notice first that the *best approximation* $p_W(\mu) \in W$ for $p(\mu) = B_\mu^{-1} f$ is

$$p_W(\mu) := \operatorname*{argmin}_{q \in W} \|p(\mu) - q\|_{\hat{X}_\mu} = \operatorname*{argmin}_{q \in W} \|f - B_\mu q\|_{Y'_\mu}, \tag{3.12}$$

which is therefore given by the *normal equation*: find $p_W(\mu) \in W$ such that

$$(f - B_\mu p_W(\mu), B_\mu q)_{Y'_\mu} = 0, \quad q \in W. \tag{3.13}$$

What keeps us from using this as the basis for a variational discretization, is the fact that the Y'_μ-scalar product is usually hard to evaluate numerically. Noting that $R_{Y'_\mu} = R_{Y_\mu}^{-1}$ the last equation is equivalent to

$$\left\langle R_{Y_\mu}^{-1} \left(f - B_\mu p_W(\mu) \right), B_\mu q \right\rangle = 0, \quad q \in W. \tag{3.14}$$

Introducing the auxiliary variable $u(\mu) := R_{Y_\mu}^{-1} \left(f - B_\mu p_W(\mu) \right)$, or rather

$$\left\langle R_{Y_\mu} u(\mu), v \right\rangle = \left\langle f - B_\mu p_W(\mu), v \right\rangle, \quad v \in Y_\mu, \tag{3.15}$$

in weak form, the relation (3.14) and hence (3.12) can be equivalently written as

$$\begin{aligned} \langle R_{Y_\mu} u(\mu), v \rangle + b_\mu(p_W(\mu), v) &= \langle f, v \rangle, \ v \in Y_\mu, \\ b_\mu(q, u(\mu)) &= 0, \quad q \in W, \end{aligned} \tag{3.16}$$

which now just involves standard L_2-inner products. Of course, in particular for $W = X_\mu$

$$\begin{aligned} \langle R_{Y_\mu} u(\mu), v \rangle + b_\mu(p(\mu), v) &= \langle f, v \rangle, \ v \in Y_\mu, \\ b_\mu(q, u(\mu)) &= 0, \quad q \in X_\mu, \end{aligned} \tag{3.17}$$

is equivalent to the original problem (3.1), which now takes the form of a *saddle point problem*. Bijectivity of B_μ readily shows that

$$u(\mu) = 0, \quad \mu \in \mathcal{P}. \tag{3.18}$$

Hence, the solution manifold of the saddle point problem (3.17) in $\bigcup_{\mu \in \mathcal{P}} X_\mu \times Y_\mu$ can be identified according to

$$\mathcal{M} = \mathcal{M}_X \times \{0\}, \quad \mathcal{M}_X := \{p(\mu) : p(\mu) \text{ solves } (3.1)\}, \tag{3.19}$$

as it should, with the one for the original problem (1.1).

Remark 3.4. Even when the spaces Y_μ differ as sets when μ varies, as *e.g.* in (2.11) for the transport equation, the solution manifold is still compact as long as the norms $\| \cdot \|_{X_\mu}$ are all equivalent to a reference norm. Hence, the greedy errors are guaranteed to tend to zero and the n-widths benchmark is applicable. The issue of parameter dependence of the involved spaces will be taken up in Section 3.3 again.

Now given a finite dimensional subspace W, we cannot treat (3.16) yet, since we cannot test by all $v \in Y_\mu$. The following interpretation of this idealized situation is immediate from the normal equation (3.14).

Remark 3.5. The problem (3.16) is equivalent to the Petrov−Galerkin scheme: find $p_W(\mu)$ such that

$$b_\mu(p_W(\mu), v) = \langle f, v \rangle, \quad v \in Y_W, \tag{3.20}$$

where

$$Y_W := R_{Y_\mu}^{-1} B_\mu W, \tag{3.21}$$

is the *optimal test space* associated with W and $p_W(\mu)$ is the best X_μ-approximation to $p(\mu)$ in X_μ.

Since (3.20) is practically infeasible a natural strategy is to replace Y_μ by a sufficiently large finite dimensional subspace $V \subset Y_\mu$ that inherits "sufficient" stability. The following observation, which plays a crucial role in what follows, explains the interrelation between a practically feasible version of (3.20) and a fully finite dimensional version of (3.16).

Proposition 3.6. *The solution component $p_{W,V}(\mu)$ of the saddle point problem*

$$\begin{array}{ll} \langle R_{Y_\mu} u_{V,W}(\mu), v \rangle + b_\mu(p_{W,V}(\mu), v) = \langle f, v \rangle, \ v \in V, \\ b_\mu(q, u_{V,W}(\mu)) \qquad\qquad\qquad = \quad 0, \quad q \in W. \end{array} \tag{3.22}$$

solves the Petrov–Galerkin problem (3.20) *with the optimal test space Y_W replaced by $\tilde{Y}_W = P_{Y_\mu,V}(R_{Y_\mu}^{-1} B_\mu(W))$ where $P_{Y_\mu,V}$ denotes the Y_μ-orthogonal projection.*

Proof. For any $q \in W$, consider $v_q := P_{Y_\mu,V}(R_{Y_\mu}^{-1} B_\mu q) \in V$ and note that, by the first equation (3.22),

$$b_\mu(p_{W,V}(\mu), v_q) = \langle B_\mu p_{W,V}(\mu), v_q \rangle = \langle f, v_q \rangle - \langle R_{Y_\mu} u_{V,W}, v_q \rangle.$$

Since

$$\langle R_{Y_\mu} u_{V,W}(\mu), v_q \rangle = (u_{V,W}(\mu), v_q)_{Y_\mu} = (u_{V,W}(\mu), R_{Y_\mu}^{-1} B_\mu q)_{Y_\mu} = b_\mu(q, u_{V,W}(\mu)) = 0,$$

where we have used the second equation in (3.22). $\qquad\square$

Clearly, the larger V the closer $\tilde{Y}_W$ is to Y_W so that the choice of V can be viewed as a *stabilization*. To quantify this observation, we call V *δ-proximal for W* if

$$\|(I - P_{Y_\mu,V}) R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu} \le \delta \|R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}, \quad q \in W, \tag{3.23}$$

holds for some fixed $0 \le \delta < 1$, see [7, 34].

Proposition 3.7. *Assume that for given $W \times V \subset X_\mu \times Y_\mu$ the test space V is δ-proximal for W, i.e.* (3.23) *is satisfied. Then, one has*

$$\|p(\mu) - p_{W,V}(\mu)\|_{\hat{X}_\mu} \le \frac{1}{1-\delta} \inf_{q \in W} \|p(\mu) - q\|_{\hat{X}_\mu}. \tag{3.24}$$

and

$$\|p(\mu) - p_{W,V}(\mu)\|_{\hat{X}_\mu} + \|u(\mu) - u_{V,W}(\mu)\|_{Y_\mu} \le \frac{2}{1-\delta} \inf_{q \in W} \|p(\mu) - q\|_{\hat{X}_\mu}. \tag{3.25}$$

Moreover, one has

$$\inf_{q \in W} \sup_{v \in V} \frac{b_\mu(q, v)}{\|v\|_{Y_\mu} \|q\|_{\hat{X}_\mu}} \ge \sqrt{1 - \delta^2}. \tag{3.26}$$

Proof. Let $p_W(\mu)$ denote the best $\hat{X}_\mu$-approximation to the exact solution $p(\mu)$ of (3.17). Then, for any $q \in W$ one has, on account of Remark 3.2,

$$\begin{aligned} (p_W(\mu) - p_{W,V}(\mu), q)_{\hat{X}_\mu} &= (p(\mu) - p_{W,V}(\mu), q)_{\hat{X}_\mu} = (B_\mu(p(\mu) - p_{W,V}(\mu)), B_\mu q)_{Y'_\mu} \\ &= \langle B_\mu(p(\mu) - p_{W,V}(\mu)), R_{Y_\mu}^{-1}(B_\mu q) \rangle = b_\mu(p(\mu) - p_{W,V}(\mu), R_{Y_\mu}^{-1}(B_\mu q)) \\ &= b_\mu(p(\mu) - p_{W,V}(\mu), (I - P_{Y_\mu,V}) R_{Y_\mu}^{-1}(B_\mu q)), \end{aligned}$$

where we have used Petrov−Galerkin orthogonality, asserted by Proposition 3.6, in the last step. By duality, Remark 3.2, (3.5), respectively (3.6), and (3.23), we conclude that

$$\|p_W(\mu) - p_{W,V}(\mu)\|_{\hat{X}_\mu} = \sup_{q \in W} \frac{b_\mu(p(\mu) - p_{W,V}(\mu), (I - P_{Y_\mu,V})R_{Y_\mu}^{-1}(B_\mu q))}{\|q\|_{\hat{X}_\mu}}$$

$$\leq \frac{\|p(\mu) - p_{W,V}(\mu)\|_{\hat{X}_\mu} \delta \|R_{Y_\mu}^{-1}(B_\mu q)\|_{Y_\mu}}{\|q\|_{\hat{X}_\mu}} = \delta \|p(\mu) - p_{W,V}(\mu)\|_{\hat{X}_\mu},$$

from which (3.24) follows by triangle inequality.

Next recall from (3.18) that, in view of the first relation in (3.22),

$$\|u(\mu) - u_{V,W}(\mu)\|_{Y_\mu}^2 = \|u_{V,W}(\mu)\|_{Y_\mu}^2 = (u_{V,W}(\mu), u_{V,W}(\mu))_{Y_\mu} = \langle f - B_\mu p_{W,V}(\mu), u_{V,W}(\mu) \rangle$$

$$\leq \|f - B_\mu p_{W,V}(\mu)\|_{Y_\mu'} \|u_{V,W}(\mu)\|_{Y_\mu} = \|p(\mu) - p_{W,V}(\mu)\|_{\hat{X}_\mu} \|u(\mu) - u_{V,W}(\mu)\|_{Y_\mu},$$

which together with (3.24) confirms (3.25).

Finally, the inf-sup estimate (3.26) is an immediate consequence of the more general Proposition 3.8 below. $\square$

We shall use the saddle point formulations to contrive rate-optimal RBMs, namely, on one hand, for computing *truth snapshots* in $W = X_\mathcal{N}$ with a suitable δ-proximal test space $Y_\mathcal{N}$, and on the other hand, for computing Galerkin projections in reduced spaces $W = X_n$ again with an associated δ-proximal test space Y_n, whose construction will be discussed in the next section.

To put this into proper perspective, given any $W \subset X_\mu$, the condition (3.23) on a $V \subset Y_\mu$ implies the best approximation property **BAP** for the Galerkin solution component $p_{W,V}(\mu)$ of (3.22) with a constant that becomes the closer to one the smaller the relative error becomes in (3.23). Moreover, (3.25) says that the accuracy of the second "auxiliary" component $u_{V,W}(\mu)$ is automatically completely governed by the accuracy of the first component $p_{W,V}(\mu)$. Finally, (3.23) *implies inf-sup stability* of (3.22). It will be shown below (for later purposes in a little more generality) that conversely inf-sup stability (3.26) *implies* δ-proximality. In fact, since the bilinear form $a_\mu(v,w) := \langle R_{Y_\mu}v, w \rangle = (v,w)_{Y_\mu}$ is trivially Y_μ-elliptic with coercivity and continuity constants $c_a(\mu) = C_a(\mu) = 1$ (see (2.2)), we could have derived the best approximation property **BAP** (3.25) *directly* from a uniform inf-sup condition from standard facts about general saddle point problems, see *e.g.* [2]. We have presented the relatively short self-contained derivation in order to identify the precise constants and to bring out the particular role of the δ-proximality condition (3.23). As we shall show later both conditions (3.23) and (3.26) can be used algorithmically to ensure stability of the saddle point problem and hence tightness of corresponding residual based surrogates.

The above discussion, especially the notion of δ-proximality, draws essentially on the use of an optimal pair of norms given by (3.6) or (3.11). In the scenario (s1) when there is no need for renormation from a stability point of view as in the context of classical saddle point problems, such as the Stokes system, it is more convenient to work with the "original" X_μ-norm related to $\|\cdot\|_{\hat{X}_\mu}$ by (3.7). The following proposition clarifies the announced interrelation between an inf-sup condition and δ-proximality.

Proposition 3.8. *As before assume that $V \subset Y_\mu$ and let $W \subset X_\mu$, $0 \leq \delta < 1$, and $\lambda > 0$. Consider the two conditions:*

$$\|(I - P_{Y_\mu,Y_n})R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu} \leq \delta \|R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu}, \quad \forall\, q \in W, \tag{3.27}$$

$$\inf_{q \in W} \sup_{v \in V} \frac{b_\mu(q,v)}{\|v\|_{Y_\mu} \|q\|_{X_\mu}} \geq \lambda. \tag{3.28}$$

Then (3.27) implies (3.28) with constant $\lambda = c_M\sqrt{1-\delta^2}$. Conversely, (3.28) implies (3.27) with constant $\delta = \sqrt{1 - C_M^2\lambda^2}$ i.e., $\lambda = C_M^{-1}\sqrt{1-\delta^2}$, where c_M and C_M are the constants from the norm equivalence (3.7).

Note that when $c_M = C_M = 1$, *e.g.* in case we use the $\hat{X}_\mu$-norm for X_μ, both stability conditions are equivalent.

Proof. We reformulate (3.27) and (3.28) in terms of equivalent conditions that can be more easily compared. First, squaring (3.27) and using that $P_{Y_\mu,V}$ is the Y_μ-orthogonal projector, we obtain

$$\|R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu}^2 - \|P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu}^2 \leq \delta^2\|R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu}^2, \qquad\qquad \forall q \in W,$$

which is equivalent to

$$\sqrt{1-\delta^2}\|R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu} \leq \|P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu}, \qquad\qquad \forall q \in W.$$

By the definition (3.6) of the graph norm $\|\cdot\|_{\hat{X}_\mu}$, this is equivalent to

$$\|P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu} \geq \sqrt{1-\delta^2}\|q\|_{\hat{X}_\mu}, \qquad\qquad \forall q \in W. \qquad (3.29)$$

Next, to reformulate (3.28), obviously the inf-sup condition is equivalent to

$$\sup_{v \in V} \frac{b_\mu(q,v)}{\|v\|_{Y_\mu}} \geq \lambda\|q\|_{X_\mu}, \qquad\qquad \forall q \in W. \qquad (3.30)$$

From (3.3) in Remark 3.1 we know that the left hand side is maximized by the function $v = P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q$ which yields

$$\sup_{v \in V} \frac{b_\mu(q,v)}{\|v\|_{Y_\mu}} = \frac{\langle P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q, B_\mu q\rangle}{\|P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu}} = \frac{\langle R_{Y_\mu}P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q, R_{Y_\mu}^{-1}B_\mu q\rangle}{\|P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu}} = \|P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu}. \quad (3.31)$$

Substituting the right hand side in the left hand side of the condition (3.30), yields

$$\|P_{Y_\mu,V}R_{Y_\mu}^{-1}B_\mu q\|_{Y_\mu} \geq \lambda\|q\|_{X_\mu}, \qquad\qquad \forall q \in W.$$

We see that this condition is identical to (3.29) up to an equivalence of the $\|\cdot\|_{X_\mu}$ and $\|\cdot\|_{\hat{X}_\mu}$ norms, which proves the assertion. $\qquad\square$

In summary, given a trial space $W \subset X_\mu$, a suitable $V \subset Y_\mu$ such that the Galerkin problem (3.22) has the best approximation property **BAP**, thereby warranting tight residual based surrogates, can be obtained by realizing

$$\inf_{q \in W} \sup_{v \in V} \frac{b_\mu(q,v)}{\|v\|_{Y_\mu}\|q\|_{X_\mu}} \geq \beta, \qquad\qquad (3.32)$$

where $\beta := \min_{\mu \in \mathcal{P}} \beta(\mu) > 0$, see (3.2).

3.3. Parameter dependence, truth spaces, and feasibility

Before applying the above findings to the construction of well-conditioned tight surrogates, we need to be a bit more precise about the *parameter dependence* in order to distinguish eventually several relevant scenarios. Notice that the spaces Y_μ, X_μ are allowed to depend on $\mu \in \mathcal{P}$ in a way that they even differ as sets and no parameter independent reference norm may exist, see Remark 2.1. Let

$$Y := \bigcap_{\mu \in \mathcal{P}} Y_\mu, \qquad\qquad\qquad\qquad X := \bigcap_{\mu \in \mathcal{P}} X_\mu, \qquad\qquad (3.33)$$

where the intersection is understood in the sense of sets. It is clear that Y and X are linear spaces. Although in general, we do *not* insist, that Y and X are endowed with norms that are equivalent to all $\|\cdot\|_{Y_\mu}$ and $\|\cdot\|_{X_\mu}$,

respectively. However, we do assume in what follows that Y, X are dense in Y_μ, X_μ, respectively, for all $\mu \in \mathcal{P}$. Moreover, on account of the compactness of $\mathcal{P}$, we can always define (possibly stronger) norms

$$\|v\|_Y := \sup_{\mu \in \mathcal{P}} \|v\|_{Y_\mu}, \quad \|q\|_X := \sup_{\mu \in \mathcal{P}} \|q\|_{X_\mu} \tag{3.34}$$

for Y, X, respectively. Moreover, since $Y = \bigcap_{\mu \in \mathcal{P}} Y_\mu$ is assumed to be dense in Y_μ, for the inf-sup condition (3.2) it suffices to take for $V = Y_\mu$ the supremum over Y instead of Y_μ, i.e., there exist subspaces $V \subset Y$ for which the discrete inf-sup condition (3.28) holds uniformly in the parameter μ.

Of course, this setting covers, in particular, the special situation – usually considered in the RBM context – that all the spaces Y_μ, X_μ, $\mu \in \mathcal{P}$, agree as sets, respectively, and where the respective norms are uniformly equivalent, i.e., there exist constants $0 < c_\circ, C_\circ < \infty$ such that

$$c_\circ \|v\|_Y \le \|v\|_{Y_\mu} \le C_\circ \|v\|_Y, \quad \mu \in \mathcal{P}, \ v \in Y, \tag{3.35}$$

and

$$c_\circ \|q\|_X \le \|q\|_{X_\mu} \le C_\circ \|q\|_X, \quad \mu \in \mathcal{P}, \ q \in X. \tag{3.36}$$

Recall from Remark 2.1 that for parametric transport equations (3.36) is valid but (3.35) does *not* hold.

At any rate, due to the denseness of X and Y, we can find sufficiently large but *finite dimensional truth spaces* $Y_\mathcal{N} \subset Y, X_\mathcal{N} \subset X$, typically finite element spaces, that can provide a desired target accuracy of the *truth model*. Since we are dealing here with problems for which standard tight *a posteriori* bounds are not available, we comment first on the *truth certification*. Note that this is particularly important for convection dominated convection diffusion equations when a complete resolution of very steep layers is prohibitively expensive even for the truth solution. We know that $\|p(\mu) - q\|_{\hat{X}_\mu} = \|f - B_\mu q\|_{Y'_\mu}$. In order to be able to accurately evaluate the residual in the dual norm $\|\cdot\|_{Y'_\mu}$ one needs in *any setting* suitable assumptions on *data oscillation*, see e.g. [4, 7, 33]. One way to express this is to require that the projection of $R_{Y_\mu}^{-1} f$ into the test space $Y_\mathcal{N}$ captures enough of $R_{Y_\mu}^{-1} f$. To this end, we make use of the following simple observation.

Remark 3.9. Assume that (3.36) holds. Given $W \subseteq X_\mathcal{N}$ and any $\delta \in (0, 1)$, there exists a finite dimensional test space $V \subset Y$ such that

$$\inf_{v \in V} \|q - R_{\hat{X}_\mu}^{-1} B_\mu^* v\|_{\hat{X}_\mu} \le \delta \|q\|_{\hat{X}_\mu}, \quad q \in \mathcal{M}_X + W, \ \mu \in \mathcal{P}, \tag{3.37}$$

which implies

$$(1 - \delta^2)^{1/2} \|f - B_\mu p\|_{Y'_\mu} \le \|P_{Y_\mu, V} R_{Y_\mu}^{-1} (f - B_\mu p)\|_{Y_\mu} \le \|f - B_\mu p\|_{Y'_\mu}, \quad p \in W. \tag{3.38}$$

In the following, we denote by $\mathcal{V}(W, \delta)$ all test spaces in $Y_\mathcal{N}$ which satisfy the stability condition (3.37).

Proof. Since $\mathcal{M}_X$ is compact there is a linear space $V_\mathcal{M}$ such that

$$\inf_{v \in V_\mathcal{M}} \|q - R_{\hat{X}_\mu}^{-1} B_\mu^* v\|_{\hat{X}_\mu} \le \delta \inf_{w \in W} \|q + w\|_{\hat{X}_\mu}, \quad q \in \mathcal{M}_X, \ \mu \in \mathcal{P}.$$

It follows that the space $V_\mathcal{M} + W$ satisfies (3.37). Furthermore, since

$$\inf_{v \in V} \|q - R_{\hat{X}_\mu}^{-1} B_\mu^* v\|_{\hat{X}_\mu} \le \delta \|q\|_{\hat{X}_\mu} \iff \inf_{v \in V} \|R_{Y_\mu}^{-1} B_\mu q - v\|_{Y_\mu} \le \delta \|R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}, \tag{3.39}$$

and since $\inf_{v \in V} \|R_{Y_\mu}^{-1} B_\mu q - v\|_{Y_\mu} = \|(I - P_{Y_\mu, V}) R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}$, the assertion follows. $\square$

We shall comment later how (3.37) can be realized, see also [7, 34] for a more detailed discussion. Since $r_{V,W}(p,f) := P_{Y_\mu,V} R_{Y_\mu}^{-1}(f - B_\mu p)$ is given by

$$\langle R_{Y_\mu} r_{V,W}(p,f), z \rangle = \langle f - B_\mu p, z \rangle, \quad z \in Y_\mathcal{N}, \tag{3.40}$$

the middle term in (3.38) is computable.

Remark 3.10. In what follows we shall always assume that for some fixed $\delta_\mathcal{N} < 1$ and any given $X_\mathcal{N} \subset X$, the finite dimensional space $Y_\mathcal{N}$ is contained in $\mathcal{V}(X_\mathcal{N}, \delta_\mathcal{N})$ satisfying (3.37) which is solely a matter of the fixed truth approximation. Therefore, abbreviating the solution of (3.22) for $W = X_\mathcal{N}, V = Y_\mathcal{N}$, as $p_{X_\mathcal{N}, Y_\mathcal{N}} =: p_\mathcal{N}(\mu) \in X_\mathcal{N}$, $u_\mathcal{N}(\mu) := u_{Y_\mathcal{N}, X_\mathcal{N}}(\mu)$, we immediately conclude that

$$\|p(\mu) - p_\mathcal{N}(\mu)\|_{\hat{X}_\mu} \le (1 - \delta^2)^{-1/2} \|u_\mathcal{N}(\mu)\|_{Y_\mu}, \quad \mu \in \mathcal{P}. \tag{3.41}$$

Remark 3.11. (i) For any desired target tolerance τ, as soon as the computable quantity $\|u_\mathcal{N}(\mu)\|_{Y_\mu}$ drops below $(1 - \delta^2)^{1/2}\tau$ we know that the truth solution has guaranteed accuracy $\le \tau$ which can be achieved by the refinement scheme in [7].
(ii) The above choice of $Y_\mathcal{N}$ guarantees, by Propositions 3.7, 3.8, in particular, that

$$\inf_{q \in X_\mathcal{N}} \sup_{v \in Y_\mathcal{N}} \frac{b_\mu(q,v)}{\|v\|_{Y_\mu}\|q\|_{\bar{X}_\mu}} \ge \xi\sqrt{1 - \delta_\mathcal{N}^2} =: \beta_\mathcal{N} > 0, \quad \mu \in \mathcal{P}, \text{ where } \xi := \begin{cases} 1, \ \| \cdot \|_{\bar{X}_\mu} = \| \cdot \|_{\hat{X}_\mu}, \\ \beta, \ \| \cdot \|_{\bar{X}_\mu} = \| \cdot \|_{X_\mu}, \end{cases} \tag{3.42}$$

and where β is the inf-sup constant from (3.32) in the case of scenario (s1), see Section 3.1. Hence, $\beta_\mathcal{N}$ can, in principle be driven as close as one wishes to one or β, depending on the choice of norm for X_μ.

Note that the above statements do not contradict the possible case that the norms $\| \cdot \|_{Y_\mu}$ or $\| \cdot \|_{X_\mu}$, $\mu \in \mathcal{P}$, are *not* equivalent to a single reference norm.

In the following, we shall often not distinguish for simplicity of exposition between truth and full spaces unless explicitly stated. In particular, whenever we speak of a computation in Y_μ, X_μ we refer to the truth spaces endowed with the norms $\| \cdot \|_{Y_\mu}, \| \cdot \|_{\hat{X}_\mu}$, respectively.

Finally, the way how the bilinear forms depend on μ is important for practical *feasibility*. We assume that the dependence of the bilinear forms on μ is *affine* in the usual sense, *i.e.*

$$b_\mu(\cdot, \cdot) = \sum_{k=1}^{m_B} \Theta_k^b(\mu) b_k(\cdot, \cdot), \tag{3.43}$$

with parameter independent bilinear forms $b_k(\cdot, \cdot)$, $k = 1, \ldots, m_B$, and smooth functions Θ_k^b.

4. STABILIZATION

Suppose we are given a pair $W = X_n \subset X_\mathcal{N}$, $Y_n \subset Y_\mathcal{N}$ of finite dimensional spaces with bases $\Phi_n = \{\phi_j\}_{j=1}^n$ and $\Psi_n = \{\psi_j\}_{j=1}^{m(n)}$, respectively. Our convention will always be that the index n reflects the dimension of X_n while generally $\dim Y_n = m(n) \ge n$. While the purpose of X_n is to approximate $\mathcal{M}_X$ the role of Y_n is, in view of Proposition 3.7, to guarantee uniform inf-sup stability. More precisely, whenever Y_n is δ-proximal for X_n (3.23) for some $\delta < 1$, one has

$$\inf_{q \in X_n} \sup_{v \in Y_n} \frac{b_\mu(q,v)}{\|v\|_{Y_\mu}\|q\|_{\hat{X}_\mu}} \ge \sqrt{1 - \delta^2}, \quad \mu \in \mathcal{P}. \tag{4.1}$$

Hence, a natural strategy is to choose a constant $0 < \zeta < 1$, replace the right hand side of (4.1) by $\zeta\sqrt{1 - \delta^2}$ and *enrich* the space Y_n until this relaxed inf-sup condition is valid. The closer one wishes $p_{X_n, Y_n}(\mu)$ to be to the best $\hat{X}_\mu$-approximation $P_{\hat{X}_\mu, X_n} p(\mu)$, the closer ζ should be chosen to one, see (3.24). In particular, *any* $\zeta < 1$ is in principle feasible.

We shall formulate actually two variants of such a *stabilization scheme* which apply under slightly different assumptions.

4.1. Inf-sup stabilization

The first natural idea which has already been used in [13, 14, 29] is to enrich Y_n by the supremizer for the infimizing parameter $\bar{\mu}$. More precisely, we first search for a parameter $\bar{\mu} \in \mathcal{P}$ and a function $\bar{q} \in X_n$ for which the inf-sup condition (3.28) is worst, *i.e.*

$$\sup_{v \in Y_n} \frac{b_{\bar{\mu}}(\bar{q}, v)}{\|v\|_{Y_{\bar{\mu}}} \|\bar{q}\|_{X_{\bar{\mu}}}} = \inf_{\mu \in \mathcal{P}} \left(\inf_{q \in X_n} \sup_{v \in Y_n} \frac{b_{\mu}(q, v)}{\|v\|_{Y_{\mu}} \|q\|_{X_{\mu}}} \right). \tag{4.2}$$

If this worst case inf-sup constant does not exceed yet a desired uniform lower bound, Y_n does not contain an effective supremizer for $\bar{\mu}, \bar{q}$, yet. However, since the truth space satisfies the uniform inf-sup condition (3.42) there exists a good supremizer in the truth space which, on account of Remark 3.1, is given by

$$\bar{v} = R_{Y_{\bar{\mu}}}^{-1} B_{\bar{\mu}} \bar{q} = \operatorname*{argmax}_{v \in Y_{\bar{\mu}}} \frac{b_{\bar{\mu}}(\bar{q}, v)}{\|v\|_{Y_{\bar{\mu}}} \|\bar{q}\|_{X_{\bar{\mu}}}},$$

and provides the enrichment

$$Y_n \to \operatorname{span}\{Y_n, R_{Y_{\mu}}^{-1} B_{\mu} \bar{q}\}. \tag{4.3}$$

This strategy can now be applied recursively until we reach a satisfactory uniform inf-sup condition for the reduced spaces.

Of course, three questions immediately arise:

(i)   Is the computation of $\bar{\mu}$ and $\bar{q}$ feasible?
(ii) Does this process terminate after finitely many steps?
(iii) If so, is the number of necessary stabilization steps affordable?

Assuming for the moment to have positive answers to (ii) and (iii), we first derive a suitable *offline/online* strategy for an efficient implementation of (4.3). First note that for given $\bar{\mu}$ and $\bar{q}$ the new test function $\bar{v} := R_{Y_{\bar{\mu}}}^{-1} B_{\bar{\mu}} \bar{q}$ can be computed by a standard Galerkin scheme

$$(\bar{v}, v)_{Y_{\bar{\mu}}} = b_{\bar{\mu}}(\bar{q}, v), \quad v \in Y_{\bar{\mu}},$$

so that it remains to solve the optimization problem (4.2) to find $\bar{\mu}$ and $\bar{q}$. To this end, we first rewrite the inf-sup condition in terms of the coefficient vectors with respect to the reduced bases. To describe this, we denote the corresponding Gramians, respectively cross-Gramians as

$$\begin{aligned}
\mathbf{R}_{\mathbf{Y}_{\mu}} &:= (\boldsymbol{\Psi}, \boldsymbol{\Psi})_{Y_{\mu}} := \left((\psi_i, \psi_j)_{Y_{\mu}} \right)_{i,j=1}^{m}, \\
\mathbf{R}_{\mathbf{X}_{\mu}} &:= (\boldsymbol{\Phi}, \boldsymbol{\Phi})_{X_{\mu}} := \left((\phi_i, \phi_j)_{X_{\mu}} \right)_{i,j=1}^{n}, \\
\boldsymbol{B}_{\mu} &:= b_{\mu}(\boldsymbol{\Phi}, \boldsymbol{\Psi}) := \left(b_{\mu}(\phi_i, \psi_j) \right)_{j,i=1}^{m,n}.
\end{aligned} \tag{4.4}$$

Practical feasibility relies on the following

Assumption 4.1. *In addition to B_{μ} the Riesz maps $R_{Y_{\mu}}$ and $R_{X_{\mu}}$ depend affinely on the parameter μ.*

Remark 4.2. Under Assumption 4.1 all the matrices in (4.4) can be computed online. Thus, by rewriting the left hand side of the inf-sup condition as

$$\inf_{\boldsymbol{q} \in \mathbb{R}^n} \sup_{\boldsymbol{v} \in \mathbb{R}^{m(n)}} \frac{\boldsymbol{v}^T \boldsymbol{B}_{\mu} \boldsymbol{q}}{\left(\boldsymbol{v}^T \mathbf{R}_{\mathbf{Y}_{\mu}} \boldsymbol{v} \right)^{1/2} \left(\boldsymbol{q}^T \mathbf{R}_{\mathbf{X}_{\mu}} \boldsymbol{q} \right)^{1/2}} \tag{4.5}$$

we are left for each parameter μ with an optimization problem only of the size of the dimensions $m(n), n$ of Y_n and X_n, respectively.

In order to find an infimizing $q \in \mathbb{R}^n$ we eliminate the discrete Riesz maps in the denominator by factoring them as

$$\mathbf{R}_{\mathbf{Y}_\mu} = \boldsymbol{L}_{Y_\mu}^T \boldsymbol{L}_{Y_\mu} \qquad\qquad\qquad \mathbf{R}_{\mathbf{X}_\mu} = \boldsymbol{L}_{X_\mu}^T \boldsymbol{L}_{X_\mu}. \qquad (4.6)$$

Here, one can think of a Cholesky factorization or of a spectral decomposition $\boldsymbol{L}_{Y_\mu} = \Lambda_{X_\mu}^{1/2} \mathbf{Q}_{\mathbf{X}_\mu}$ where the columns of $\mathbf{Q}_{\mathbf{X}_\mu}$ form an eigenbasis and Λ_{X_μ} is the diagonal matrix with the eigenvalues (in descending order) on the diagonal. Replacing $\boldsymbol{v}$ by $\boldsymbol{L}_{Y_\mu} \boldsymbol{v}$ and $\boldsymbol{q}$ by $\boldsymbol{L}_{X_\mu} \boldsymbol{q}$ and defining $\boldsymbol{D}_\mu := \boldsymbol{L}_{Y_\mu}^{-T} \boldsymbol{B}_\mu \boldsymbol{L}_{X_\mu}^{-1}$ we find that

$$\inf_{\boldsymbol{q} \in \mathbb{R}^n} \sup_{\boldsymbol{v} \in \mathbb{R}^{m(n)}} \frac{\boldsymbol{v}^T \boldsymbol{B}_\mu \boldsymbol{q}}{\left(\boldsymbol{v}^T \mathbf{R}_{\mathbf{Y}_\mu} \boldsymbol{v}\right)^{1/2} \left(\boldsymbol{q}^T \mathbf{R}_{\mathbf{X}_\mu} \boldsymbol{q}\right)^{1/2}} = \inf_{\boldsymbol{q} \in \mathbb{R}^n} \sup_{\boldsymbol{v} \in \mathbb{R}^{m(n)}} \frac{\boldsymbol{v}^T \boldsymbol{D}_\mu \boldsymbol{q}}{\|\boldsymbol{v}\|_{\ell_2} \|\boldsymbol{q}\|_{\ell_2}}. \qquad (4.7)$$

and hence, one easily verifies the following fact.

Remark 4.3. For any given μ the corresponding inf-sup constant is the smallest singular value of $\boldsymbol{D}_\mu$ and the optimal $\boldsymbol{q}$ is the corresponding right singular vector. Since the computational cost of the singular value decomposition is polynomial in the dimensions of the reduced bases, we can afford to compute all the inf-sup constants for a sufficiently large sample set $\mathcal{S} \subset \mathcal{P}$ of parameters, yielding the optimal $\bar{\mu}$.

The complete scheme is summarized in Algorithm 2 which we formulate for the general norms X_μ in (3.2) and the inf-sup constant $\beta_\mathcal{N}$ from (3.42).

Algorithm 2. Update to achieve inf-sup stability.
_____

1: **function** UPDATE-INF-SUP(Y_n, X_n)
2:     Choose $0 < \zeta < 1$.
3:     Select a sufficiently large sample $\mathcal{S} \subset \mathcal{P}$.
4:     **repeat**
5:         **for** $\mu \in \mathcal{S}$ **do**
6:             Assemble the Gramians and cross-Gramians $\mathbf{R}_{\mathbf{Y}_\mu}$, $\mathbf{R}_{\mathbf{X}_\mu}$, $\boldsymbol{B}_\mu$.
7:             Compute the Cholesky decompositions

$$\mathbf{R}_{\mathbf{Y}_\mu} = \boldsymbol{L}_{Y_\mu}^T \boldsymbol{L}_{Y_\mu} \qquad\qquad\qquad \mathbf{R}_{\mathbf{X}_\mu} = \boldsymbol{L}_{X_\mu}^T \boldsymbol{L}_{X_\mu}.$$

8:             Determine the smallest singular value $\sigma(\mu)$ and corresponding
                right singular vector $\bar{\boldsymbol{q}}_\mu$ of the matrix $\boldsymbol{D}_\mu = \boldsymbol{L}_{Y_\mu}^{-T} \boldsymbol{B}_\mu \boldsymbol{L}_{X_\mu}$.
9:         **end for**
10:         Set $\bar{\mu} = \min\{\sigma(\mu) : \mu \in \mathcal{S}\}$
11:         Update $Y_n \leftarrow \mathrm{span}\{Y_n, R_{Y_{\bar{\mu}}}^{-1} B_{\bar{\mu}} \bar{q}_{\bar{\mu}}\}$ with $\bar{q}_{\bar{\mu}} = \sum_{i=1}^n (\bar{\boldsymbol{q}}_{\bar{\mu}})_i \phi_i$.
12:     **until** $\sigma(\bar{\mu}) \geq \zeta \beta_\mathcal{N}$
13:     **return** Y_n
14: **end function**

4.2. Stabilization based on δ-proximality

We shall now formulate an alternative stabilizing scheme. It is related to greedy approximation and will shed some light on the above stabilization algorithm regarding the questions (ii), (iii). The idea is to enrich the space Y_n to obtain stability based on the equivalent criterion (3.27) which can be rephrased as

$$\inf_{\phi \in Y_n} \|R_{Y_\mu}^{-1} B_\mu q - \phi\|_{Y_\mu} \leq \delta \|R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}, \quad \forall\, q \in X_n,\, \mu \in \mathcal{P}. \qquad (4.8)$$

Defining

$$X_n^1(\mu) := \left\{ q \in X_n : \|q\|_{\hat{X}_\mu} = \|R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu} = 1 \right\}, \tag{4.9}$$

this is equivalent to

$$\sup_{\mu \in \mathcal{P}} \sup_{q \in X_n^1(\mu)} \inf_{\phi \in Y_n} \|R_{Y_\mu}^{-1} B_\mu q - \phi\|_{Y_\mu} \le \delta. \tag{4.10}$$

We can again employ a greedy strategy to search for the parameter $\mu \in \mathcal{P}$ and the element in $X_n^1 := \bigcup_{\mu \in \mathcal{P}} X_n^1(\mu)$ for which the error is worst:

$$(\bar{\mu}, \bar{q}) = \operatorname*{argmax}_{\mu \in \mathcal{P}; q \in X_n^1(\mu)} \inf_{\phi \in Y_n} \|R_{Y_\mu}^{-1} B_\mu q - \phi\|_{Y_\mu}. \tag{4.11}$$

As long as the approximation error for $\bar{\mu}$ and $\bar{q}$ exceeds some fixed $\delta \in (0, 1)$, we add the best approximation from the full truth space to the reduced basis:

$$Y_n \to \operatorname{span}\{Y_n, R_{Y_{\bar{\mu}}}^{-1} B_{\bar{\mu}} \bar{q}\}.$$

Since, as pointed out below (3.6), $R_{Y_\mu}^{-1} = B_\mu^{-*} R_{\hat{X}_\mu} B_\mu^{-1}$, we see that $(R_{Y_\mu}^{-1})^{-1} = B_\mu R_{\hat{X}_\mu}^{-1} B_\mu^*$. Hence, in view of (3.6), we conclude that

$$(\bar{\mu}, \bar{q}) = \operatorname*{argmax}_{\mu \in \mathcal{P}; q \in X_n^1(\mu)} \left(\inf_{\phi \in Y_n} \|q - R_{\hat{X}_\mu}^{-1} B_\mu^* \phi\|_{\hat{X}_\mu} \right), \tag{4.12}$$

which implies the following observation.

Remark 4.4. If (3.36) holds so that all the spaces X_μ agree with a parameter independent reference space X, the output $(\bar{\mu}, \bar{q})$ is the result of a greedy approximation to the compact set $X_n^1 := \bigcup_{\mu \in \mathcal{P}} X_n^1(\mu)$. Therefore, in principle, the scheme fits into the standard greedy theory in [1, 3, 11]. In fact, by (3.36), (3.43), and the fact that $B_\mu^* : Y_\mu \to (\hat{X}_\mu)'$ is an isometry, the set $R_X^{-1} B_\mu^* Y_n$ is a finite dimensional subspace of X.

It remains to find a fast algorithm for the solution of the maximization problem (4.11) which will make use of the $\|\cdot\|_{\hat{X}_\mu}$-norm (3.6) for X_μ.

Lemma 4.5. Let $q = \sum_{j=1}^n q_j \phi_j =: \mathbf{q^T \Phi}$. Referring to the matrices $\boldsymbol{B}_\mu, \mathbf{R_{Y_\mu}}$ from (4.4), and defining $\mathbf{R_{\hat{X}_\mu}} := (\boldsymbol{\Phi}, \boldsymbol{\Phi})_{\hat{X}_\mu}$, one has

$$\|(I - P_{Y_\mu, Y_n}) R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}^2 = \mathbf{q^T} \left(\mathbf{R_{\hat{X}_\mu}} - \boldsymbol{B}_\mu^{\mathbf{T}} \mathbf{R_{Y_\mu}^{-1}} \boldsymbol{B}_\mu \right) \mathbf{q}. \tag{4.13}$$

Proof. By orthogonality of P_{Y_μ, Y_n} and (3.6), we have

$$\|(I - P_{Y_\mu, Y_n}) R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}^2 = \|q\|_{\hat{X}_\mu}^2 - \|P_{Y_\mu, Y_n} R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}^2.$$

By definition, we have $\|q\|_{\hat{X}_\mu}^2 = \mathbf{q^T R_{\hat{X}_\mu} q}$. As for the second term, note that $P_{Y_\mu, Y_n} R_{Y_\mu}^{-1} B_\mu q$ is the Galerkin solution of $R_{Y_\mu} z = B_\mu q$. Since for any $w \in Y_\mu'$ the coefficient vector $\mathbf{z}$ of $P_{Y_\mu, Y_n} R_{Y_\mu}^{-1} w$ is given by

$$\mathbf{R_{Y_\mu} z} = \langle \mathbf{w}, \boldsymbol{\Psi} \rangle =: \left(\langle \mathbf{w}, \psi_\mathbf{j} \rangle \right)_{\mathbf{j=1}}^{\mathbf{m}},$$

we conclude that for $w := \mathbf{q^T B_\mu \Phi}$ one has

$$\mathbf{z} = \mathbf{R_{Y_\mu}^{-1}} \langle \boldsymbol{\Psi}, \boldsymbol{B}_\mu \boldsymbol{\Phi} \rangle \mathbf{q} = \mathbf{R_{Y_\mu}^{-1}} \boldsymbol{B}_\mu \mathbf{q}.$$

Hence

$$\|P_{Y_\mu, Y_n} R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}^2 = \langle B_\mu q, P_{Y_\mu, Y_n} R_{Y_\mu}^{-1} B_\mu q \rangle_{Y_\mu} = \mathbf{q^T} \boldsymbol{B}_\mu^{\mathbf{T}} \mathbf{R_{Y_\mu}^{-1}} \boldsymbol{B}_\mu \mathbf{q}.$$

which confirms the claim. $\square$

Similarly, by the definition (3.6) of the $\hat{X}_\mu$-norm, we have $\|q\|_{\hat{X}_\mu} = \|R_{Y_\mu}^{-1} B_\mu q\|_{Y_\mu}$ so that

$$X_n^1(\mu) = \left\{ q \in X_n : q = \mathbf{q}^{\mathbf{T}}\mathbf{\Phi}, \ \mathbf{q}^{\mathbf{T}}\mathbf{R}_{\hat{\mathbf{X}}_\mu}\mathbf{q} = \mathbf{1} \right\}.$$

It follows that the optimization problem (4.11) is equivalent to

$$(\bar{\mu}, \bar{\mathbf{q}}) = \operatorname*{argmax}_{\mu \in \mathcal{P}; \mathbf{q} \in \mathbb{R}^{\mathbf{n}}} \frac{\mathbf{q}^{\mathbf{T}}\big(\mathbf{R}_{\hat{\mathbf{X}}_\mu} - B_\mu^{\mathbf{T}}\mathbf{R}_{\mathbf{Y}_\mu}^{-1} B_\mu\big)\mathbf{q}}{\mathbf{q}^{\mathbf{T}}\mathbf{R}_{\hat{\mathbf{X}}_\mu}\mathbf{q}}, \tag{4.14}$$

where $\bar{\mathbf{q}}$ is the coefficient vector of $\bar{q}$. This problem can be solved analogously to the corresponding optimization problem (4.2), (4.5) of the inf-sup condition so that we obtain the alternative algorithm UPDATE-δ for updating Y_n.

Algorithm 3. Update to achieve δ-proximality.

1: **function** UPDATE-$\delta(Y_n, X_n)$
2:    Choose $0 < \delta < 1$.
3:    Select a sufficiently large sample $\mathcal{S} \subset \mathcal{P}$.
4:    **repeat**
5:      Assemble the Gramians and cross-Gramians $\mathbf{R}_{\mathbf{Y}_\mu}$, $\mathbf{R}_{\hat{\mathbf{X}}_\mu}$, B_μ (see (4.4)).
6:      Compute

$$\delta_{max} = \max_{\mu \in \mathcal{S}; \mathbf{q} \in \mathbb{R}^{\mathbf{n}}} \frac{\mathbf{q}^{\mathbf{T}}\big(\mathbf{R}_{\hat{\mathbf{X}}_\mu} - B_\mu^{\mathbf{T}}\mathbf{R}_{\mathbf{Y}_\mu}^{-1} B_\mu\big)\mathbf{q}}{\mathbf{q}^{\mathbf{T}}\mathbf{R}_{\hat{\mathbf{X}}_\mu}\mathbf{q}},$$

$$(\bar{\mu}, \bar{\mathbf{q}}) = \operatorname*{argmax}_{\mu \in \mathcal{S}; \mathbf{q} \in \mathbb{R}^{\mathbf{n}}} \frac{\mathbf{q}^{\mathbf{T}}\big(\mathbf{R}_{\hat{\mathbf{X}}_\mu} - B_\mu^{\mathbf{T}}\mathbf{R}_{\mathbf{Y}_\mu}^{-1} B_\mu\big)\mathbf{q}}{\mathbf{q}^{\mathbf{T}}\mathbf{R}_{\hat{\mathbf{X}}_\mu}\mathbf{q}}.$$

7:      Update $Y_n \leftarrow \operatorname{span}\{Y_n, R_{Y_{\bar{\mu}}}^{-1} B_{\bar{\mu}} \bar{q}_{\bar{\mu}}\}$ with $\bar{q}_{\bar{\mu}} = \sum_{i=1}^n (\bar{\boldsymbol{q}}_{\bar{\mu}})_i \phi_i$.
8:    **until** $\delta_{max} \leq \delta$
9:    **return** Y_n
10: **end function**

The efficient practical execution of Algorithm UPDATE-δ requires assembling the matrices $\mathbf{R}_{\hat{\mathbf{X}}_\mu}$ in the typical offline/online fashion. This is possible when instead of Assumption 4.1 the following holds.

Assumption 4.6. *The Riesz maps R_{Y_μ}, $R_{\hat{X}_\mu}$ and hence their inner products $(\cdot,\cdot)_{Y_\mu}$, $(\cdot,\cdot)_{\hat{X}_\mu}$ depend affinely on the parameter $\mu \in \mathcal{P}$.*

By (3.6), Assumption 4.6 is valid if the Y_μ-norm can be chosen independent of μ, *i.e.*, when (3.35) holds. Moreover Assumption 4.6 can also be satisfied for parameter dependent Y_μ-norms as *e.g.*, in view of (3.8), for the transport equation.

Finally, it is important to note that the number of operations used by both algorithms UPDATE-INF-SUP and UPDATE-δ (under Assumption 4.6) only depends on the size of the sample set $\mathcal{S}$ and the dimensions n and $m(n)$ of the reduced bases. Especially, it is independent of the dimension of the truth spaces which renders these algorithms feasible.

Assumption 4.6 is clearly more restrictive, *i.e.* the use of UPDATE-δ is more constrained than UPDATE-INF-SUP which applies under the standard assumptions of affine dependence and for any norm on X_μ.

4.3. Interrelation between both stabilization schemes

We discuss next the interrelation between the schemes UPDATE-INF-SUP and UPDATE-δ.

Proposition 4.7. *Assume that we use the $\| \cdot \|_{\hat{X}_\mu}$-norm for X_μ and the spectral decomposition*

$$\mathbf{R}_{\hat{\mathbf{X}}_\mu} = \boldsymbol{L}_{\mathbf{X}_\mu}^{\mathbf{T}} \boldsymbol{L}_{\mathbf{X}_\mu} = \mathbf{Q}_\mu^{\mathbf{T}} \boldsymbol{\Lambda}_\mu^{1/2} \boldsymbol{\Lambda}_\mu^{1/2} \mathbf{Q}_\mu.$$

in (4.6) *for the scheme* UPDATE-INF-SUP, *where Λ_μ is the diagonal matrix with the eigenvalues and $\mathbf{Q}_\mu$ the matrix of corresponding eigenvectors. Then the outputs of* UPDATE-INF-SUP *and* UPDATE-δ *coincide.*

Proof. Let $\mathbf{M}_\mu := \boldsymbol{B}_\mu^{\mathbf{T}} \mathbf{R}_{\mathbf{Y}_\mu}^{-1} \boldsymbol{B}_\mu$. Clearly, since $\mathbf{Q}_\mu$ is orthogonal,

$$\lambda_{\max}(\mu) := \max_{\mathbf{q} \in \mathbb{R}^{\mathbf{n}}} \frac{\mathbf{q}^{\mathbf{T}}(\mathbf{R}_{\hat{\mathbf{X}}_\mu} - \boldsymbol{B}_\mu^{\mathbf{T}} \mathbf{R}_{\mathbf{Y}_\mu}^{-1} \boldsymbol{B}_\mu)\mathbf{q}}{\mathbf{q}^{\mathbf{T}} \mathbf{R}_{\hat{\mathbf{X}}_\mu} \mathbf{q}}$$

is the largest eigenvalue of the matrix $\mathbf{I} - \boldsymbol{\Lambda}_\mu^{-1/2} \mathbf{Q}_\mu \mathbf{M}_\mu \mathbf{Q}_\mu^{\mathbf{T}} \boldsymbol{\Lambda}_\mu^{-1/2}$ so that

$$\lambda_{\max}(\mu) = 1 - \lambda_{\min}(\Lambda_\mu^{-1/2} \mathbf{Q}_\mu \mathbf{M}_\mu \mathbf{Q}_\mu^{\mathbf{T}} \boldsymbol{\Lambda}_\mu^{-1/2}). \tag{4.15}$$

On the other hand, using the $\hat{X}_\mu$-norm in UPDATE-INF-SUP, *i.e.*, replacing $\mathbf{R}_{\mathbf{X}_\mu}$ by $\mathbf{R}_{\hat{\mathbf{X}}_\mu}$ in (4.6), and using the spectral decomposition $\mathbf{R}_{\hat{\mathbf{X}}_\mu} = \mathbf{Q}_\mu^{\mathbf{T}} \boldsymbol{\Lambda}_\mu^{1/2} \boldsymbol{\Lambda}_\mu^{1/2} \mathbf{Q}_\mu$ for $\boldsymbol{L}_{\mathbf{X}_\mu}^{\mathbf{T}} \boldsymbol{L}_{\mathbf{X}_\mu}$, the matrix $\mathbf{D}_\mu$ in (4.7) takes the form $\mathbf{D}_\mu = \boldsymbol{L}_{\mathbf{Y}_\mu}^{-\mathbf{T}} \boldsymbol{B}_\mu \mathbf{Q}_\mu^{\mathbf{T}} \boldsymbol{\Lambda}^{-1/2}$. Clearly, the smallest singular value of $\mathbf{D}_\mu$ is just $\lambda_{\min}(\mathbf{D}_\mu^{\mathbf{T}} \mathbf{D}_\mu)^{1/2}$ and the corresponding eigenvector agrees with the right singular vector of $\mathbf{D}_\mu$. Since

$$\mathbf{D}_\mu^{\mathbf{T}} \mathbf{D}_\mu = \boldsymbol{\Lambda}_\mu^{-1/2} \mathbf{Q}_\mu \mathbf{M}_\mu \mathbf{Q}_\mu^{\mathbf{T}} \boldsymbol{\Lambda}_\mu^{-1/2}$$

we see that in this case the enrichments produced by both schemes agree, which confirms the claim. $\square$

4.4. Termination of stabilizing greedy loops

4.4.1. The general case

Under the most general assumptions, neither insisting on (3.35) nor on (3.36) we resort to a very crude argument that ensures termination of the stabilizations loops UPDATE-INF-SUP and UPDATE-δ. Our findings can be summarized as follows.

Proposition 4.8. *Both schemes* UPDATE-INF-SUP *and* UPDATE-δ *always terminate after finitely many steps.*

Proof. We prove the assertion only for the scheme UPDATE-INF-SUP. The argument for UPDATE-δ is identical. To this end, let Y_{n+1} and X_{n+1} be the spaces obtained by applying UPDATE-INF-SUP to the input spaces Y_n and X_n. According to the update rule (4.3) used by UPDATE-INF-SUP, the enlarged space Y_{n+1} is contained in the truth space $Y_\mathcal{N}$. Thus, since $Y_\mathcal{N}$ is finite dimensional, the statement of the proposition follows if each added function is linearly independent to the previous ones. To this end, assume the algorithm has already grown Y_n to $\tilde{Y}_n$ and let $\bar{v} = R_{Y_{\bar\mu}}^{-1} B_{\bar\mu} \bar{q}$ be the next function to be added (see (3.3)). Now, assume by contradiction that it is already contained in $\tilde{Y}_n$. Since $\bar{v}$ is a supremizer this implies that

$$\sup_{v \in \tilde{Y}_n} \frac{b_{\bar\mu}(\bar{q}, v)}{\|v\|_{Y_\mu} \|\bar{q}\|_{X_\mu}} \geq \zeta \beta_\mathcal{N}.$$

Recalling that $\bar\mu$ and $\bar{q}$ are the worst possible choices according to (4.2), this violates the stopping criterion in Line 12 of UPDATE-INF-SUP. Thus, it follows that $\bar{v}$ is linearly independent from $\tilde{Y}_n$ showing finite termination of UPDATE-INF-SUP. $\square$

The fact that, by the above argument, the number of stabilization steps may depend on the dimension of the truth space is certainly very pessimistic and not satisfactory from a practical point of view. In fact, much more can be said under some additional assumptions.

4.5. Uniformly equivalent norms

Suppose now that all the spaces Y_μ, X_μ agree as sets with Y, X, respectively (see (3.33)), and that (3.35), (3.36)) hold. Then we can replace $\|\cdot\|_{Y_\mu}$ by an uniformly equivalent reference norm $\|\cdot\|_Y$. Since the Riesz map R_Y is now independent of μ, Assumption 4.6 holds and the stabilizing schemes UPDATE-INF-SUP and UPDATE-δ are equivalent, see Proposition 4.7. Moreover, recall that, by Remark 3.2, the supremizer for $q_n \in X_n$ in the inf-sup condition is given by $R_Y^{-1} B_\mu q_n$. The key observation is that because of the affine decomposition (3.43) of B_μ all these supremizers together generate a finite dimensional space.

Remark 4.9. Given $X_n \subset X$, $X_n = \mathrm{span}\,\{\phi_j : j = 1,\ldots,n\}$, let

$$\hat{Y}_n := \left\{ R_Y^{-1} B_\mu p : p \in X_n,\ \mu \in \mathcal{P} \right\} \subseteq \mathrm{span}\left\{ R_Y^{-1} B_\mu \phi_j : j = 1,\ldots,n,\ \mu \in \mathcal{P} \right\}. \tag{4.16}$$

Then $\hat{Y}_n$ is a finite dimensional space of dimension $\dim \hat{Y}_n \le m_B n$, where m_B is the number of terms in the affine expansion (3.43). Hence, one has

$$\inf_{\mu \in \mathcal{P}} \inf_{q \in X_n} \sup_{v \in \hat{Y}_n} \frac{b_\mu(q,v)}{\|q\|_{X_\mu} \|v\|_Y} = \inf_{\mu \in \mathcal{P}} \inf_{q \in X_n} \sup_{v \in Y_\mathcal{N}} \frac{b_\mu(q,v)}{\|q\|_{X_\mu} \|v\|_Y} \ge \beta_\mathcal{N}, \tag{4.17}$$

where $\beta_\mathcal{N} > 0$ is the inf-sup constant from (3.42) in Remark 3.11.

Proof. If B_k is the operator corresponding to the bilinear form $b_k(\cdot,\cdot)$ in the affine expansion (3.43) and $R_Y = R_{Y_\mu}$ is independent of $\mu \in \mathcal{P}$, we conclude that

$$\hat{Y}_n \subseteq \mathrm{span}\left\{ R_Y^{-1} B_k \phi_j : j = 1,\ldots,n,\ k = 1,\ldots,m_B \right\}$$

which proves the first part of the claim. Since all optimal test functions are contained in $\hat{Y}_n$ the discrete inf-sup condition (4.17) follows immediately from the assumed inf-sup condition (3.42) of the full problem. $\qquad\square$

The following simple observation is an immediate consequence of Remark 4.9.

Proposition 4.10. *Assume that* (3.35), *hold. Then the update algorithm* UPDATE-INF-SUP, *and hence likewise* UPDATE-δ, *increases the dimension of the test space in each step and terminates with a test space of dimension at most* nm_B.

Proof. The proof is identical to the one of Theorem 4.8 by noting that all supremizers that are added during the algorithm are not only contained in the truth space $Y_\mathcal{N}$ but in the much smaller space $\hat{Y}_n \subset Y_\mathcal{N}$ which is of dimension $m_B n$. $\qquad\square$

The above reasoning applies verbally to other saddle point problems like those appearing in parameter dependent Stokes systems or constrained optimization problems. The finite dimensionality of $\hat{Y}_n$ is also the basis of the *a priori* choice of stabilizers in [13, 14, 29] to guarantee inf-sup stability although the connection with a greedy stabilization does not seem to be made there.

The reason for nevertheless applying such a greedy stabilization is that a sufficient inf-sup stability might actually be achieved at an earlier stage so that in total fewer stabilizers suffice.

4.6. A greedy perspective

As we shall see in later applications, in the context of Section 3.1 it will be important to treat also the case where only (3.36) holds but (3.35) is *not valid*. In this case the norms $\|\cdot\|_{X_\mu}, \|\cdot\|_{\hat{X}_\mu}$ are all equivalent and can be replaced by a parameter independent reference norm $\|\cdot\|_X$. For instance, in the case (2.11) one has $\|\cdot\|_{L_2(\Omega)} = \|\cdot\|_{\hat{X}_\mu} = \|\cdot\|_X$, $\mu \in \mathcal{P}$, which will be further discussed in later numerical experiments. For the remainder of this section we assume that only (3.36) is valid.

We have already seen that $(\tilde{q}, \tilde{\mu}) := \operatorname{argmin}_{q \in X_n^1, \mu \in \mathcal{P}} \left(\sup_{v \in Y_n} b_\mu(q, v) / \|v\|_{Y_\mu} \right)$ agrees with the output of (4.11) and, on account of Remark 4.4, of a greedy approximation step to the set X_n^1. Hence, the question of termination of the stabilization loop is equivalent to finding the smallest j for which

$$\sigma_{n,j} := \max_{q \in X_n^1, \mu \in \mathcal{P}} \left(\inf_{\psi \in Y_n^j} \|q - R_X^{-1} B_\mu^* \psi\|_X \right) \leq \delta, \tag{4.18}$$

where $Y_n^0 = Y_{n-1}$ and Y_n^j is the enrichment of Y_n^0 produced by the jth stabilization step. Here we assume that for the preceding pair (X_{n-1}, Y_{n-1}) we have that $Y_{n-1} = Y_{n-1}^{\ell_{n-1}}$ satisfies $\sigma_{n-1,\ell_{n-1}} \leq \delta$. We wish to see now how Y_n^j evolves from Y_{n-1}. For convenience let $K_\mu := R_X^{-1} B_\mu^*$

A straightforward application of the currently available greedy concepts from [1, 11] is complicated by the fact that the sets X_n^1 become "less compact" when n grows and that the approximating subspaces $R_X^{-1} B_\mu^* Y_n$ depend on μ through the application of B_μ^*. The following discussion is merely to shed some light on the expected behavior of $\sigma_{n,j}$, in particular, to identify some driving mechanisms, while we postpone a more detailed discussion to forthcoming work.

Our first remarks concern the continuity of the mapping $\mu \mapsto K_\mu$. To this end, recall that the space $Y = \bigcap_{\mu \in \mathcal{P}} Y_\mu$ is endowed with the norm $\| \cdot \|_Y$ from (3.34) which is here allowed to be stronger than the individual norms $\| \cdot \|_{Y_\mu}$. In view of (3.36) and (3.7), we have for any $\psi \in Y$

$$\|K_\mu \psi\|_X \leq C_0 C_M \|K_\mu \psi\|_{\hat{X}_\mu} = C_0 C_M \|\psi\|_{Y_\mu} \leq C_0 C_M \|\psi\|_Y.$$

Thus, $K_\mu \in L(Y, X)$ which is equivalent to saying $B_\mu^* \in L(Y, X')$. Now let B_k^* be the component of B_μ^* corresponding to the kth bilinear form $b_k(\cdot, \cdot)$ in (3.43) which, by assumption, are smooth. Obviously, one has

$$\|(K_\mu - K_{\mu'})\psi\|_X \leq \sum_{k=1}^{m_B} |\Theta_k^b(\mu) - \Theta_k^b(\mu')| \|B_k^* \psi\|_{X'} \leq C \max_{k=1,\dots,m_B} |\Theta_k^b(\mu) - \Theta_k^b(\mu')| \|\psi\|_Y. \tag{4.19}$$

which shows that the mapping $\mathcal{P} \to L(Y, X), \mu \mapsto K_\mu$ is continuous in μ. By compactness of $\mathcal{P}$, we can find for each $\epsilon > 0$ a finite ϵ-net comprised of $N_\epsilon(\mathcal{P})$ centers $\mu_{\epsilon,j}$ such that for each $\psi \in Y$ and any $\mu \in \mathcal{P}$ there exists a $j \in \{1, \dots, N_\epsilon(\mathcal{P})\}$ such that

$$\|(K_\mu - K_{\mu_{\epsilon,j}})\psi\|_X \leq \epsilon \|\psi\|_Y. \tag{4.20}$$

In order to estimate $\|\psi\|_Y$ we introduce the constant

$$C(n, \mathcal{N}) := \max_{\mu \in \mathcal{P}; \, q \in X_n^1} \|K_\mu^{-1} q\|_Y, \tag{4.21}$$

which is finite because $K_{\bar{\mu}}^{-1} \phi_r \in Y_\mathcal{N} \subset Y$. However, a point of concern is that the quantity $C(n, \mathcal{N})$ may depend on the truth space dimension, a point that will be taken up later again. In particular, we have for any given n and any $\psi = K_{\bar{\mu}}^{-1} \phi_r$, $r \leq n$, where ϕ_r is any of the orthonormalized reduced basis functions generated by snapshots from $\mathcal{M}_X$,

$$\|(K_{\bar{\mu}} - K_{\mu_{\epsilon,j}}) K_{\bar{\mu}}^{-1} \phi_r\|_X \leq \epsilon \|K_{\bar{\mu}}^{-1} \phi_r\|_Y \leq \epsilon C(n, \mathcal{N}), \quad \bar{\mu} \in \mathcal{P}. \tag{4.22}$$

For $n = 1$ the greedy stabilization would determine a sequence $\bar{\mu}_{1,l}$, $l = 1, \dots, \ell_1$, such that

$$\sup_{\mu \in \mathcal{P}} \left\| \phi_1 - \sum_{l=1}^{\ell_1} c_l(\mu) K_\mu \left(K_{\bar{\mu}_{1,l}}^{-1} \phi_1 \right) \right\|_X \leq \delta. \tag{4.23}$$

Thus, for $\epsilon \leq \delta / 2C(n, \mathcal{N})$, this means that $\ell_1 \leq N_\epsilon(\mathcal{P})$. In fact, as long as (4.23) does not hold no two $\bar{\mu}_{1,l}$ can fall into a single ball of the ϵ-cover of $\mathcal{P}$ and as soon as every ball contains a $\bar{\mu}_{1,l}$ (4.22) says that (4.23) is valid.

It is now easy to display for any given n a space $V_n \subset Y$ which is δ-proximal for X_n. In fact, pick an ϵ_n-net for $\mathcal{P}$ where $\epsilon_n := \delta/(C(n,\mathcal{N})\sqrt{n})$ and let

$$V_n := \mathrm{span}\,\{K_{\mu_{\epsilon_n,l}}^{-1}\phi_k : l = 1,\ldots,N_{\epsilon_n}(\mathcal{P}),\, k = 1,\ldots,n\}.$$

Hence, for any $\mu \in \mathcal{P}$ there exists a center $\mu_{\epsilon_n,k}$ such that, on account of (4.22), for every $r = 1,\ldots,n$,

$$\|\phi_r - K_\mu(K_{\mu_{\epsilon_n,k}}^{-1}\phi_r)\|_X = \|(K_{\mu_{\epsilon_n,k}} - K_\mu)(K_{\mu_{\epsilon_n,k}}^{-1}\phi_r)\|_X \le \delta/n^{1/2}.$$

Since the ϕ_k are X-orthonormal we obtain for any $q \in X_n^1$

$$\inf_{\psi \in V_n} \|q - K_\mu\psi\|_X \le \sum_{r=1}^{n} |(q,\phi_r)_X|\|\phi_r - K_\mu(K_{\mu_{\epsilon_n,k}}^{-1}\phi_r)\|_X \le \delta\Big(\sum_{k=1}^{n} |(q,\phi_r)_X|^2\Big)^{1/2} \le \delta.$$

Of course, it is not clear whether the greedy procedure for building the spaces Y_n^j would actually produce a space of similar dimension $\dim V_n \le nN_{\epsilon_n}(\mathcal{P}) =: N_n$. If one did instead a separate greedy procedure for each subspace spanned by ϕ_r the argument for (4.23) would say that it terminates after at most N_n steps. Intuitively, one expects that the actual greedy algorithm terminates earlier since each individual ϕ_r has to be resolved only with accuracy δ, not with accuracy $\epsilon_n = \delta/(n^{1/2}C(n,\mathcal{N}))$ as above.

The perhaps simplest, although grossly pessimistic, way of rigorously bounding the number of greedy steps providing the spaces Y_n^j in the stabilization loop, is to use the above pigeonhole principle and consider in addition to a $\bar{\delta}$-net for $\mathcal{P}$ with centers $\mu_{\bar{\delta},l}$, $l = 1,\ldots,N_{\bar{\delta}}(\mathcal{P})$, also a $\bar{\epsilon}$-net for the compact set X_n^1 with centers q_i, $i = 1,\ldots,N_{\bar{\epsilon}}(X_n^1)$, where $\bar{\delta}, \bar{\epsilon}$ will be specified later. Now suppose that $K_{\bar{\mu}}^{-1}\bar{q}$ is the new snapshot added to Y_n^{j-1} to generate Y_n^j. There exist, by construction, indices $l \in \{1,\ldots,N_{\bar{\delta}}(\mathcal{P})\}$ and $i \in \{1,\ldots,N_{\bar{\epsilon}}(X_n^1)\}$ such that $\|\bar{q} - q_i\|_X \le \bar{\epsilon}$ and $\|(K_{\bar{\mu}} - K_{\mu_{\bar{\delta},l}})\psi\|_X \le \bar{\delta}\|\psi\|_Y$.

Then, for any $(q,\mu) \in X_n^1 \times \mathcal{P}$ with $\|q_i - q\|_X \le \bar{\epsilon}$ and $\|(K_\mu - K_{\mu_{\bar{\delta},l}})\psi\|_X \le \bar{\delta}\|\psi\|_Y$ where $q_i, \mu_{\bar{\delta},l}$ are related to the new snapshot $K_{\bar{\mu}}^{-1}\bar{q}$ as above, we obtain

$$\inf_{\psi \in Y_n^j} \|q - K_\mu\psi\|_X \le \|q - K_\mu K_{\bar{\mu}}^{-1}\bar{q}\|_X \le \|q - q_i\|_X + \|q_i - \bar{q}\|_X + \|(K_{\bar{\mu}} - K_\mu)K_{\bar{\mu}}^{-1}\bar{q}\|_X$$

$$\le 2(\bar{\epsilon} + \bar{\delta}C(n,\mathcal{N})). \tag{4.24}$$

Now choose $\bar{\epsilon}$ and $\bar{\delta}$ such that $2(\bar{\epsilon} + \bar{\delta}C(n,\mathcal{N})) = \delta$, *e.g.* by taking $\bar{\epsilon} = \delta/4$, $\bar{\delta} = \delta/(4C(n,\mathcal{N}))$. It follows from (4.24) that a new snapshot $K_{\bar{\mu}}^{-1}\bar{q}$ can only satisfy $\inf_{\psi \in Y_n^{j-1}} \|\bar{q} - K_{\bar{\mu}}\psi\|_X > \delta$ if it falls into a cover element $B_{q_i,\mu_{\bar{\delta},l}}(\bar{\epsilon},\bar{\delta}) := \{(q,\mu) : \|q_i - q\|_X \le \bar{\epsilon},\, \|(K_\mu - K_{\mu_{\bar{\delta},l}})\psi\|_X \le \bar{\delta}\|\psi\|_Y\}$ that does not contain any previous snapshot yet. This can happen at most $N_{\bar{\delta}}(\mathcal{P})N_{\bar{\epsilon}}(X_n^1)$ times which therefore bounds the number of possible greedy steps in the stabilization loop.

As mentioned before, this bound is very pessimistic. In fact, since X_n^1 is isometrically isomorphic to a unit sphere in ℓ_2^n the covering numbers $N_{\bar{\epsilon}}(X_n^1)$ increase like $(12/\bar{\epsilon})^n$, see [20], Chapter 13. The numbers $N_{\bar{\delta}}(\mathcal{P})$ instead depend only on the fixed dimension of the parameter set $\mathcal{P}$ and the smoothness of the parameter functions $\Theta_k^b(\mu)$.

One way to ameliorate the strong dependence of the $N_{\bar{\epsilon}}(X_n^1)$ on n is to relate the problem to a greedy approximation to a compact set that is *independent* of n. To this end, recall the solution set $\mathcal{M} = \mathcal{M}_X \times \{0\}$, see (3.19), which under the present assumptions is compact in $X \times \{0\}$, independent of the truth spaces. As detailed later the spaces X_n are generated by a (weak) greedy algorithm. By compactness, the (weak) greedy errors

$$\sigma_n := \sigma_n(\mathcal{M}_X, X_n) := \max \mathrm{dist}\,(\mathcal{M}_X, X_n)_X \to 0, \quad n \to \infty, \tag{4.25}$$

tend to zero at a rate that is independent of the truth dimension. A repeated greedy approximation generates an X-orthonormal system $\{\phi_j\}_{j=1}^{\infty} \subset X$. Let X° denote the closure of the span of $\{\phi_j\}_{j=1}^{\infty}$, *i.e.*

$$X^\circ := \Big\{q \in X : \sum_{j\in\mathbb{N}} (q,\phi_j)_X^2 < \infty\Big\}.$$

Let

$$\mathcal{F} := \{q \in X^\circ : |q|_* < \infty\}, \qquad |q|_* := \sup_{n \in \mathbb{N}} \sigma_n^{-1}\left(\sum_{j=n+1}^\infty (q, \phi_j)_X^2\right)^{1/2}.$$

Obviously, $\mathcal{B} := \{q \in \mathcal{F} : \max\{\|q\|_X, |q|_*\} \le 1\}$ is a compact subset of X and by construction

$$\text{dist}\,(q, X_n)_X \le \sigma_n |q|_*, \quad q \in \mathcal{F},\, n \in \mathbb{N}. \tag{4.26}$$

Moreover, the greedy errors for $\mathcal{B}$ are comparable to the greedy errors for $\mathcal{M}_X$. In particular,

$$\max \text{dist}\,(\mathcal{B}, X_n)_X \le \max \text{dist}\,(\mathcal{M}_X, X_n)_X, \quad n \in \mathbb{N}. \tag{4.27}$$

Furthermore,

$$X_n^1 := \{q \in X_n : \|q\|_{\hat{X}_\mu} \le 1,\, \mu \in \mathcal{P}\} \subseteq \{q \in X_n : \|q\|_X \le c_M^{-1}\} \subset \mathcal{F},$$

since for $q \in X_n$

$$|q|_* = \max_{j \le n} \sigma_j^{-1}\left(\sum_{k=j+1}^n (q, \phi_k)_X^2\right)^{1/2} \le \sigma_n^{-1}\|q\|_X \le c_M^{-1}\sigma_n^{-1}. \tag{4.28}$$

Therefore, recalling (4.18), we conclude that

$$\begin{aligned}
\sigma_{n,j} &\le \max_{q \in X_n^1 \cap \mathcal{B},\, \mu \in \mathcal{P}} \left(\inf_{\psi \in Y_n^j} \|q - R_{\hat{X}_\mu}^{-1} B_\mu^* \psi\|_{\hat{X}_\mu}\right)(c_M \sigma_n)^{-1} \\
&\le \max_{q \in X_n^1 \cap \mathcal{B},\, \mu \in \mathcal{P}} \left(\inf_{\psi \in Y_n^j} \|q - R_X^{-1} B_\mu^* \psi\|_X\right) C_M (c_M \sigma_n)^{-1},
\end{aligned} \tag{4.29}$$

where c_M, C_M are the constants from (3.6). Hence, termination of the stabilization loop reduces to analyzing the necessary number of steps needed to enrich $Y_{n-1}^{\ell_{n-1}} = Y_n^0$ until $\max_{q \in X_n^1 \cap \mathcal{B},\, \mu \in \mathcal{P}} \left(\inf_{\psi \in Y_n^j} \|q - R_X^{-1} B_\mu^* \psi\|_X\right) \le c_M \sigma_n / C_M$. Clearly, $N_\epsilon(X_n^1 \cap \mathcal{B}) \le N_\epsilon(\mathcal{B})$ where $\mathcal{B}$ is now a fixed compact set. We can now apply the same reasoning as above with $X_n^1, \bar{\epsilon}, \bar{\delta}$ replaced by $\mathcal{B}, \bar{\epsilon} c_M \sigma_n / C_M, \bar{\delta} c_M \sigma_n / C_M$, respectively. This leads to the alternative bound $N_{\delta \sigma_n c_M/(4C_M)}(\mathcal{B}) N_{\delta \sigma_n c_M/(C_M C(n, \mathcal{N}))}(\mathcal{P})$ for the maximal number of greedy steps. Note that in this case $C(n, \mathcal{N})$ can be replaced by

$$C(\mathcal{B}, \mathcal{N}) := \max_{\mu \in \mathcal{P}, q \in \mathcal{B}} \|K_\mu^{-1} q\|_Y.$$

Since every $q \in X_n^1$ or $q \in \mathcal{B}$ is a linear combination of snapshots $B_{\mu_l}^{-1} f = p(\mu_l)$ and since the enrichments of the test spaces Y_n^j are of the form $B_{\bar{\mu}_j}^{-*} R_X q_j$, $q_j \in X_n^1$, they are linear combinations of elements of the form $B_{\mu'}^{-*} R_X B_{\mu''}^{-1} f$. Since the operators $B_{\mu'}^{-*} R_X B_{\mu''}^{-1}$ at least preserve the regularity of f the quantities $K_\mu^{-1} q$, $q \in X_n^1$ ($q \in \mathcal{B}$), where now the inversion is understood in the infinite dimensional spaces, possess the required additional regularity in Y when f is sufficiently regular, see the discussion of the transport problem in Section 6.3.

Summary 4.11. *We can now summarize the above findings as follows:*

1. *If the constants $C(n, \mathcal{N})$ or $C(\mathcal{B}, \mathcal{N})$ are uniformly bounded independently of the choice of the truth spaces the stabilization loops terminate after a number of steps that is independent of the truth spaces. Their dependence on n can be bounded in terms of the metric entropy of X_n^1 or the metric entropies of $\mathcal{B}$ and $\mathcal{P}$, coupled in the latter case with the greedy errors σ_n for $\mathcal{M}_X$.*
2. *The constants $C(n, \mathcal{N})$, $C(\mathcal{B}, \mathcal{N})$ remain independent of the truth spaces when f is sufficiently regular.*

5. A DOUBLE GREEDY SCHEME

We shall discuss now a greedy strategy for constructing reduced spaces X_n, Y_n for the saddle point problem (3.17) which is a weak formulation of (3.1).

The basic outline of such a strategy looks as follows:

- *Stabilization:* Given a pair Y_n, X_n, enrich Y_n until $\beta_{Y_n,X_n}(\mu) \geq \zeta\beta_{\mathcal{N}}$, $\mu \in \mathcal{P}$, where $\beta_{\mathcal{N}}$ is given by (3.42) and $\beta_{Y_n,X_n}(\mu)$, $\mu \in \mathcal{P}$, is the inf-sup constant (3.32) for the reduced spaces X_n and Y_n.
- *Approximation update:* In view of the best approximation property (3.24), (3.25), we then improve the accuracy of the reduced spaces with the aid of a greedy step.

That last greedy step, in turn, requires a tight residual based surrogate as detailed next.

5.1. Tight surrogates

Suppose now that the pair of spaces $X_n \subset X$, $Y_n \subset Y$ satisfy the δ-proximity condition (3.23) for some $\delta \in (0,1)$ and abbreviate the corresponding solutions of (3.22) as $p_n(\mu) := p_{X_n,Y_n}(\mu)$, $u_n(\mu) := u_{Y_n,X_n}(\mu)$. By Propositions 3.7, 3.8, the definition (3.6) of the $\hat{X}_\mu$-norm says then that

$$\|p(\mu) - p_n(\mu)\|_{\hat{X}_\mu} = \|f - B_\mu p_n(\mu)\|_{Y'_\mu}, \quad \mu \in \mathcal{P},$$

i.e., the residual based surrogate

$$R(\mu, X_n \times Y_n) := \|f - B_\mu p_n(\mu)\|_{Y'_\mu} \tag{5.1}$$

is in this case almost ideal. In fact, combining (5.1) with (3.24) yields

$$\inf_{q \in W} \|p(\mu) - q\|_{\hat{X}_\mu} \leq R(\mu, X_n \times Y_n) \leq \frac{1}{1-\delta} \inf_{q \in X_n} \|p(\mu) - q\|_{\hat{X}_\mu}. \tag{5.2}$$

Hence, (1.8) holds with $c_R = 1 - \delta, C_R = 1$.

5.1.1. Reduction to Truth-Riesz maps

Of course, the dual norm $\|\cdot\|_{Y'_\mu}$ and hence $R(\mu, X_n \times Y_n)$ cannot be computed exactly. Instead, defining

$$\|\cdot\|_{Y'_{\mathcal{N}}} := \|P_{Y_\mu, Y_{\mathcal{N}}} R_{Y_\mu}^{-1} \cdot \|_{Y_\mu} = \sup_{v \in Y'_{\mathcal{N}}} \frac{\langle \cdot, v \rangle}{\|v\|_{Y_\mu}}, \tag{5.3}$$

we consider the following candidate

$$R_n(\mu) := \|f - B_\mu p_n(\mu)\|_{Y'_{\mathcal{N}}}, \tag{5.4}$$

where we continue to assume that $Y_{\mathcal{N}} \in \mathcal{V}(X_{\mathcal{N}}, \delta_{\mathcal{N}})$, *i.e.*, the truth spaces $X_{\mathcal{N}}, Y_{\mathcal{N}}$ comply with Remark 3.10. Then, by (3.40) and (3.41), we conclude that

$$\|p(\mu) - p_n(\mu)\|_{\hat{X}_\mu} \leq (1 - \delta_{\mathcal{N}}^2)^{-1/2}\|P_{Y_\mu,Y_{\mathcal{N}}}(R_{Y_\mu}^{-1}(f - B_\mu p_n(\mu))\|_{Y_\mu} = (1 - \delta_{\mathcal{N}}^2)^{-1/2}R_n(\mu)$$

$$\leq (1 - \delta_{\mathcal{N}}^2)^{-1/2}\|f - B_\mu p_n(\mu)\|_{Y'_\mu} \leq (1 - \delta_{\mathcal{N}}^2)^{-1/2}(1-\delta)^{-1}\|p(\mu) - P_{\hat{X}_\mu, X_n}p(\mu)\|_{\hat{X}_\mu}. \tag{5.5}$$

This immediately implies the following fact.

Proposition 5.1. *Under the above assumptions on the truth spaces the surrogate $R_n(\mu)$ given by (5.4) is tight with condition*

$$\kappa(R_n) \leq \frac{1}{(1 - \delta_{\mathcal{N}}^2)^{1/2}(1-\delta)}, \tag{5.6}$$

which, in principle, can be driven as close to one as one wishes, at a computational expense caused by a correspondingly large truth space $Y_{\mathcal{N}}$ and a possibly larger number of stabilization steps.

The equivalence

$$\|f - B_\mu q\|_{Y_{\mathcal{N}}'} \sim \|f - B_\mu q\|_{Y_\mu'}, \quad q \in X_{\mathcal{N}}, \tag{5.7}$$

which is nothing but a reformulation of (3.38) for $W = X_{\mathcal{N}}, Y_{\mathcal{N}} \in \mathcal{V}(X_{\mathcal{N}}, \delta_{\mathcal{N}})$, says that the $\|\cdot\|_{Y_{\mathcal{N}}'}$-norm yields still a meaningful error estimate even in case the truth spaces are not rich enough to resolve all features of the infinite dimensional exact solution which will be seen below in the experiments.

The above findings can be summarized as follows.

Proposition 5.2. *If* (3.35) *holds, then* $R_n(\mu)$, *defined by* (5.4), *is feasible.*

Proof. Under the given assumptions the norms $\|\cdot\|_{Y_\mu}$ can be replaced by a uniformly equivalent reference norm $\|\cdot\|_Y$ so that the Riesz map R_Y is independent of $\mu \in \mathcal{P}$. Hence, $R_n(\mu)$ can, in view of (3.40), be efficiently evaluated by a standard offline/online decomposition, see *e.g.* [28]. □

5.1.2. Iterative tightening

Recall that in the pure transport problem (3.35) does not hold, see Remark 2.1. Hence, the surrogate $R_n(\mu)$ from (5.4) is no longer feasible in the strict sense. Instead a feasible variant would be

$$R_n'(\mu) := \|f - B_\mu p_n(\mu)\|_{Y_n'} = \|P_{Y_\mu, Y_n} R_{Y_\mu}^{-1}(f - B_\mu p_n(\mu))\|_{Y_\mu}, \tag{5.8}$$

where the dual norm is now induced by the *reduced* space Y_n instead of the truth space $Y_{\mathcal{N}}$. While the δ-proximality of Y_n for X_n (see (3.23)) does ensure the equivalence $\|B_\mu q\|_{Y_\mu'} \sim \|B_\mu q\|_{Y_n'}$, $q \in X_n$, (with constants close to one, depending on δ) the analog is not clear for $\|f - B_\mu q\|_{Y_n'}$ since generally $f \notin B_\mu(X_n)$.

However, Remark 3.9 immediately tells us at least a criterion for the validity of the desired residual equivalence, namely with the aid of a somewhat strengthened δ-proximality.

Remark 5.3. Assume that for some $\bar{\delta} \in (0, 1)$ one chooses $Y_n \in \mathcal{V}(X_n, \bar{\delta})$ so that

$$\inf_{v \in Y_n} \|p - R_{\hat{X}_\mu}^{-1} B_\mu^* v\|_{X_\mu} \le \bar{\delta}\|p\|_{X_\mu}, \quad \forall\, p \in \mathcal{M}_X + X_n. \tag{5.9}$$

Then

$$(1 - \bar{\delta}^2)^{1/2}\|f - B_\mu q\|_{Y_\mu'} \le \|f - B_\mu q\|_{Y_n'} \le \|f - B_\mu q\|_{Y_\mu'}, \quad q \in X_n, \mu \in \mathcal{P}, \tag{5.10}$$

and we have $\kappa(R_n') \le (1 - \bar{\delta}^2)^{-1/2}(1 - \delta)^{-1}$.

Note that we could replace $\mathcal{M}_X$ in (5.9) by its truth approximation $\mathcal{M}_{X, \mathcal{N}}$ since, in view of (5.7), it suffices to establish $\|f - B_\mu q\|_{Y_n'} \sim \|f - B_\mu q\|_{Y_{\mathcal{N}}'}$. But the main practical issue remains how to find Y_n satisfying (5.9) at affordable cost.

To this end, we propose a systematic way of successively substantiating tightness of error estimators at the expense of an additional computational effort in the offline phase. We refer to this process as *iterative tightening*. The idea is that once a reduced space provides sufficiently accurate approximations to $\mathcal{M}_X$, condition (5.9) becomes easier to fulfill. To make use of this observation, assume we have a second pair of reduced spaces $\bar{X} \subset X_{\mathcal{N}}$ and $\bar{Y} \subset Y_{\mathcal{N}}$. We now describe how such spaces can give rise to tight surrogates and later discuss their construction.

Lemma 5.4. *Assume that the pair* $X_n + \bar{X}$ *and* $\bar{Y}$ *satisfies the (standard)* δ-proximality condition (3.23) *and that the approximation of* $p(\mu)$ *from* $X_n + \bar{X}$ *is superior to the approximation from* X_n *alone, i.e., one has for some* $0 \le \xi < 1$

$$\|p(\mu) - \bar{p}^n(\mu)\|_{\hat{X}_\mu} \le \xi\|p(\mu) - p^n(\mu)\|_{\hat{X}_\mu}, \quad \mu \in \mathcal{P}, \tag{5.11}$$

where $p^n(\mu)$ *and* $\bar{p}^n(\mu)$ *are the respective best approximations to* $p(\mu)$ *from* X_n *and* $X_n + \bar{X}$. *Then*

$$\inf_{\bar{v} \in \bar{Y}} \|p - R_{\hat{X}_\mu}^{-1} B_\mu^* \bar{v}\|_{\hat{X}_\mu} \le \bar{\delta}\|p\|_{\hat{X}_\mu}, \quad \forall\, p \in \mathcal{M}_X + X_n, \mu \in \mathcal{P}, \tag{5.12}$$

where $\bar{\delta} := (1 + \delta)\xi + \delta$. *Hence, for* ξ, δ *sufficiently small, the surrogate* $R_n'(\mu)$ *from* (5.8) *is tight with a condition given by* (5.6) *with* $\delta_{\mathcal{N}}$ *replaced by* $\bar{\delta}$.

Note that we use the space $X_n + \bar{X}$ as opposed to $\bar{X}$ alone because for the latter space the condition (5.11) would imply that $X_n \subset \bar{X}$ if both spaces are constructed from snapshots, which would be too restrictive for the application below.

Proof. For each deviation $p(\mu) - p$, $p \in X_n$, we obtain

$$\inf_{\bar{v} \in \bar{Y}} \|p(\mu) - p - R_{X_\mu}^{-1} B_\mu^* \bar{v}\|_{\hat{X}_\mu} \leq \|p(\mu) - \bar{p}^n(\mu)\|_{\hat{X}_\mu} + \inf_{\bar{v} \in \bar{Y}} \|\bar{p}^n(\mu) - p - R_{\hat{X}_\mu}^{-1} B_\mu^* \bar{v}\|_{\hat{X}_\mu}$$

$$\leq \|p(\mu) - \bar{p}^n(\mu)\|_{\hat{X}_\mu} + \delta \|\bar{p}^n(\mu) - p\|_{\hat{X}_\mu} \leq \big((1+\delta)\xi + \delta\big) \|p(\mu) - p\|_{\hat{X}_\mu}. \quad (5.13)$$

Thus for $(1+\delta)\xi + \delta$ sufficiently small the extended δ-proximality condition (5.9) is satisfied for the trial space X_n and test space $\bar{Y}$. Thus, Remark 5.3 applies which says that the surrogate $R'_n(\mu)$ from (5.8) is tight with the claimed condition. $\square$

We shall describe ways of constructing the spaces $\bar{X}$ and $\bar{Y}$ later in Section 6.

5.2. Approximation update

Either scheme UPDATE-INF-SUP or UPDATE-δ outputs a pair X_n, Y_n that is uniformly inf-sup stable, *i.e.*, the corresponding inf-sup constant is uniformly bounded away from zero $\beta_{Y_n, X_n}(\mu) \geq \zeta \beta_{\mathcal{N}}$, $\mu \in \mathcal{P}$. By Proposition 5.1, the surrogate $R_n(\mu)$, defined by (5.4), is tight with a condition controlled by the δ-proximality parameters. The feasibility of this surrogate depends on the way how the spaces Y_μ depend on the parameter $\mu \in \mathcal{P}$, see Proposition 5.2. In applications, an infeasible surrogate is replaced by $R'_n(\mu)$ from (5.8) combined with iterative tightening.

This suggest the following *outer* greedy step UPDATE-APPROXIMATION, defined in Algorithm 5.15, which aims at improving on the accuracy of the reduced model.

Algorithm 4. Update to improve the approximation quality.

Require: Finite dimensional spaces $Y_n \subset Y$, $X_n \subset X$ that satisfy the inf-sup condition

$$\inf_{\mu \in \mathcal{P}} \inf_{q \in X_n} \sup_{v \in Y_n} \frac{b_\mu(q, v)}{\|q\|_{X_\mu}} \geq \zeta \beta_{\mathcal{N}} \qquad (5.14)$$

for some $0 < \zeta \leq 1$.
1: **function** UPDATE-APPROXIMATION(Y_n, X_n)
2:     Select a sufficiently large sample $\mathcal{S} \subset \mathcal{P}$.
3:     compute

$$\hat{\mu} := \operatorname*{argmax}_{\mu \in \mathcal{P}} R_n(\mu)$$

4:     Compute the solution $[\hat{u}, \hat{p}] \in Y_{\hat{\mu}} \times X_{\hat{\mu}}$ of

$$\begin{aligned} (u, v)_{Y_{\hat{\mu}}} + b_{\hat{\mu}}(p, v) &= \langle f, v \rangle, & v \in Y_{\hat{\mu}}, \\ b_{\hat{\mu}}(q, u) &= \langle g, q \rangle, & q \in X_{\hat{\mu}}. \end{aligned} \qquad (5.15)$$

5:     Set

$$\operatorname{span} \{X_n, \hat{p}\} \to X_n$$

6:     If (3.36) holds orthonormalize $\{\phi_1, \ldots, \phi_{n-1}, \hat{p}_n\}$ in X.
7:     **return** X_n
8: **end function**

5.3. Putting things together

The overall double-greedy method (see Algorithm DoGre below) for computing reduced spaces for (7.1) consists now in combining the inner greedy stabilization loop with the outer greedy approximation step for the saddle point formulation (3.17).

Algorithm 5. Double greedy scheme

1: **function** DoGre
2:     Initialize $Y_1 = \{0\}$ and $X_1 = \text{span}\,\{u(\mu_1)\}$ for an arbitrary $\mu_1 \in \mathcal{P}$.
3:     $Y_1 \leftarrow$ Update-inf-sup(Y_1, X_1).
4:     **while** $\max_{\mu \in \mathcal{P}} R_n(\mu) > \tau$ **do**
5:         $X_n \leftarrow$ Update-Approximation(Y_n, X_n).
6:         $Y_n \leftarrow$ Update-inf-sup(Y_n, X_n).
7:     **end while**
8:     **return** Y_n, X_n
9: **end function**

To analyze of algorithm DoGre recall the solution manifold $\mathcal{M} = \mathcal{M}_X \times \{0\}$ from (3.19). Since the inner stabilization loops ensure, by Proposition 3.7 and (5.2), tightness of the surrogates, we can invoke Theorem 1.3. The above findings can now be summarized as follows.

Theorem 5.5. *Assume that* (3.36) *holds. Let* $p_n(\mu) := p_{X_n, Y_n}(\mu)$, $u_n(\mu) := u_{Y_n, X_n}(\mu)$ *denote the solution components of* (3.22) *for* $W = X_n$, $V = Y_n$, *were* $[Y_n, X_n]$ *are the reduced spaces produced by algorithm* DoGre *using the surrogate* (5.4). *Let*

$$\sigma_n(\mathcal{M}_X) := \sup_{\mu \in \mathcal{P}} \|p(\mu) - p_n(\mu)\|_{\hat{X}_\mu}, \quad d_n(\mathcal{M}_X) := \inf_{\dim(Z_n)=n} \Big(\max \text{dist}(\mathcal{M}_X, Z_n)_X \Big).$$

(a) Then, if $d_n(\mathcal{M}_X) = O(n^{-\alpha})$, *for some* $\alpha > 0$ *or if* $d_n(\mathcal{M}_X) = O(e^{-cn^\alpha})$, *for some* $c, \alpha > 0$, *one has*

$$\sigma_n(\mathcal{M}_X) = O(n^{-\alpha}), \quad \sigma_n(\mathcal{M}) = O(e^{-\tilde{c}n^\alpha}), \quad n \to \infty, \tag{5.16}$$

respectively, with constants depending on the parameters $\delta_\mathcal{N}, \delta, \zeta$ *in* Update-δ *or* Update-Inf-Sup, *and on the constants in* (3.36), (3.7). *Moreover,* (5.16) *remains valid for* $\sigma_n(\mathcal{M}_X)$ *replaced by*

$$\hat{\sigma}_n(\mathcal{M}_X) := \sup_{\mu \in \mathcal{P}} \big\{ \|p(\mu) - p_n(\mu)\|_X + \|u(\mu) - u_n(\mu)\|_{Y_\mu} \big\}. \tag{5.17}$$

(b) Assume that both (3.35) *and* (3.36) *hold. Then, the assertion (a) holds where in addition* $\dim(Y_n \times X_n) \leq (1 + m_B)n$, $n \in \mathbb{N}$. *All bounds remain valid up to the tolerance* tol* *when all computations are carried out within this accuracy. Moreover, the surrogate* (5.4) *in algorithm* DoGre, *Steps 4 and 5 is feasible.*

Proof. The output $[Y_n, X_n]$ of Steps 5 and 6 in DoGre is uniformly inf-sup stable so that the surrogate (5.4) used in Step 5 is uniformly tight, with a condition depending on the stabilization thresholds $\delta, \delta_\mathcal{N}$. Concerning $\hat{\sigma}_n(\mathcal{M}_X)$ we use (3.25). By (3.36) the surrogates remain uniformly tight for the reference norm $\| \cdot \|_X$. Hence, Theorem 1.3 applies. The rest of the assertion follows from Propositions 4.10 and 5.2. □

In general, under the assumption (a), the well conditioned surrogate (5.4) is not feasible. Employing the feasible surrogate (5.8) instead, requires, in order to guarantee rate optimality, an additional iterative tightening as described in Section 5.1.2 and later in connection with numerical experiments. Note also that under the assumption (a) $\dim Y_n$ could be significantly larger than n, see the discussion in Section 4.6. In the case of uniformly equivalent norms, *i.e.*, when both conditions (3.35) and (3.36) hold, the dimension of the stabilizing spaces Y_n remains proportional to the dimension of the reduced primal space.

We conclude this section with a remark on the online evaluation of $p_n(\mu)$. Recall that the corresponding component $u_n(\mu) \in Y_n$ is only an auxiliary variable tending to zero.

Remark 5.6. Assume that (3.35) and (3.36) hold, *i.e.* the spaces X_μ and Y_μ can be choose parameter independent. Instead of solving for a given μ the saddle point problem (3.22) for $W = X_n$, $V = Y_n$, whose dimension is $n + m(n)$, one can compute in the offline phase the *test basis functions* $\psi_{k,j}$, $j = 1, \ldots n$, $k = 1, \ldots, m_B$

$$(\psi_{k,j}, v)_Y = b_k(\phi_j, v), \quad v \in Y_n, \ j = 1, \ldots, n, \tag{5.18}$$

where b_k are the components of the affine decomposition (3.43). Then, defining

$$\psi_j^n(\mu) := \sum_{k=1}^{m_B} \Theta_k^b(\mu) \psi_{k,j}, \tag{5.19}$$

on account of Proposition 3.6, for each $\mu \in \mathcal{P}$, the solution $p_n(\mu) = p_{X_n, Y_n}(\mu)$ of the saddle point problem (3.22), also solves the Petrov−Galerkin problem

$$b_\mu(p_n(\mu), \psi_j^n(\mu)) = \langle f_\mathcal{N}, \psi_j^n(\mu) \rangle, \quad j = 1, \ldots, n. \tag{5.20}$$

Hence the online complexity is indeed determined by the size n of the trial basis.

6. APPLICATION TO THE MODEL PROBLEMS

6.1. Singularly perturbed convection-diffusion problems

We refer to the setting in Section 2.3.1 and consider the convection-diffusion problem (2.5) for large Peclet numbers.

To this end, we shall briefly discuss two scenarios concerning the truth spaces, namely (a) boundary layers are to be resolved completely by the truth spaces, and (b) due to a possibly very small diffusion, even the truth spaces cannot resolve the boundary layers.

In case (a) solutions in the truth spaces could be obtained by simple standard Galerkin discretizations and a modified variational formulation according to (3.6) is only needed for the computation of reduced basis functions which then also resolve boundary layers well.

In this example we prescribe the space Y_μ and adjust X_μ according to (3.6). We first decompose B_μ into its symmetric and skew-symmetric parts:

$$s_\mu(u,v) := \frac{1}{2} \Big(\langle B_\mu u, v \rangle + \langle B_\mu v, u \rangle \Big), \quad k_\mu(u,v) := \frac{1}{2} \Big(\langle B_\mu u, v \rangle - \langle B_\mu v, u \rangle \Big),$$

and define

$$\|v\|_{Y_\mu}^2 := s_\mu(v,v) = \epsilon |v|_{H^1(\Omega)}^2 + \left\| \left(c - \frac{1}{2} \operatorname{div} b(\mu) \right)^{1/2} v \right\|_{L_2(\Omega)}^2, \tag{6.1}$$

see [5,30,33] for details. $\| \cdot \|_{Y_\mu}$ is then equivalent to the standard $H^1(\Omega)$-norm with constants depending on the diffusion ϵ. This works perfectly when the discretization (adaptive or not) resolves the boundary layers. However, when layers are not resolved, although stable, the scheme (3.22) would give rise to unpleasant numerical artifacts, due to the nature of the involved norms, see the detailed discussion in [5,34].

Therefore, we briefly recall next an alternative variational formulation of (2.6) avoiding the numerical artifacts, regardless of choosing sufficiently large truth spaces that fully resolve boundary layers or not. In essence, in case the finite element truth space does not resolve boundary layers this scheme behaves like a solver of the corresponding transport problem for $\epsilon = 0$ which is, however, ill-posed when insisting on zero boundary conditions on all of $\partial\Omega$. We resort to a remedy proposed in [5,34]. We retain the construction of the norms

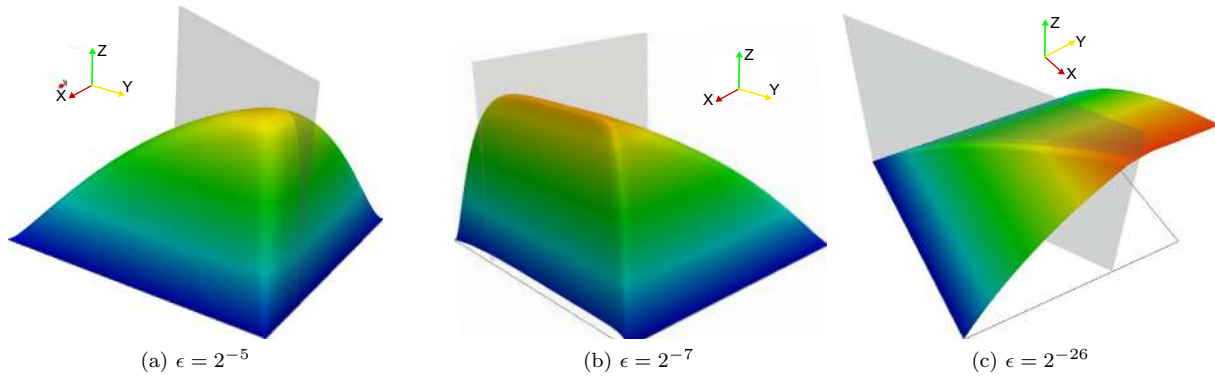(a) $\epsilon = 2^{-5}$ (b) $\epsilon = 2^{-7}$ (c) $\epsilon = 2^{-26}$

FIGURE 1. Solutions of the convection-diffusion problem (6.8). (a) RB dimension $n = 7$, $m(n) = 16$, angle $\mu = 0.885115$, (b) RB dimension $n = 8$, $m(n) = 23$, angle $\mu = 0.259830$, (c) RB dimension $n = 20$, $m(n) = 57$, angle $\mu = 0.587137$.

but modify the outflow boundary condition. Instead of building them into the trial space, we impose them only weakly. To this end, let

$$\Gamma_+(\mu) := \{x \in \partial\Omega : n(x) \cdot b(\mu, x) > 0\}$$

be the outflow boundary where $n(x)$ is the outward unit normal at x. Now, we take

$$\bar{X}_\mu := \{q \in H^1(\Omega) : q|_{\Gamma_-(\mu)} = 0\} \tag{6.2}$$

as a set with norm defined below. Here and in the following, restrictions to the boundary are implicitly considered in a trace sense. Thus, zero boundary conditions are only built into $\bar{X}_\mu$ on part of the boundary. To find a weak form of the boundary conditions at the outflow boundary recall from (3.12) the connection of (3.6) with the optimization problem

$$\|f - B_\mu \bar{p}\|_{Y'_\mu}^2 \to \min \tag{6.3}$$

where $\bar{p}$ belongs now to the larger space $\bar{X}_\mu$. So far we have not changed Y_μ which is still $H_0^1(\Omega)$ endowed with the norm (6.1). Due to the missing outflow boundary conditions, B_μ has a nontrivial kernel so that the optimization problem is not uniquely solvable. One simple remedy is to add the outflow boundary condition as a penalty term:

$$\|f - B_\mu \bar{p}\|_{Y'_\mu}^2 + \omega \|\bar{p}\|_{H_b(\mu)}^2 \to \min, \tag{6.4}$$

where $\|\cdot\|_{H_b(\mu)}$ is a norm for $H_{00}^{1/2}(\Gamma_+(\mu))$ and $\omega > 0$, see [5]. Practically, this weak enforcement of the outflow boundary condition applied to a subspace $W \subset \bar{X}_\mu$ has the following effect: typically boundary layers are found at the outflow boundary which are too narrow to be resolved at affordable cost. If ω is chosen small, then the enforcement of the outflow boundary condition has little weight so that it is almost ignored which, in turn, removes layer artifacts. If, however, W is sufficiently rich so as to resolve layers, $\inf_{q \in W} \|f - B_\mu q\|_{Y'_\mu}$ becomes so small that the boundary penalty becomes important and the boundary conditions are approximately satisfied, see Figure 1a. The rationale is that as long as the layer is not resolved the error with respect to conventional norms (including the SUPG-norm) is mostly concentrated in the layer region, which therefore stays, roughly speaking, as large as not realizing the boundary conditions at the outflow boundary at all. Putting a small weight on this error contribution actually increases accuracy away from the outflow boundary, see [5]. Putting it in a slightly different way, by allowing more freedom in the outflow boundary layer, the X_μ-norm is changed in such a manner that the error in the boundary layer has very small weight. This in turn allows one to better control the error away from the layer.

To apply the theory of Section 3.1, we define the test space

$$\bar{Y}_\mu := Y_\mu \times H_b(\mu)' = H_0^1(\Omega) \times H_{00}^{1/2}(\Gamma_+(\mu))',$$

$$\|[v,g]\|_{\bar{Y}_\mu}^2 := \|v\|_{Y_\mu}^2 + \omega \|g\|_{H_b(\mu)'}^2, \tag{6.5}$$

and the operator

$$\bar{B}_\mu p := [B_\mu p, p|_{\Gamma_+(\mu)}].$$

According to the definition of the graph-norm (3.6), this yields the norm

$$\|p\|_{\bar{X}_\mu}^2 := \|\bar{B}_\mu p\|_{\bar{Y}_\mu'}^2 = \|\bar{B}_\mu p\|_{Y_\mu'}^2 + \|p\|_{H_b(\mu)}^2 \tag{6.6}$$

for the trial space $\bar{X}_\mu$. Note that, in view of (6.1), the first part of the $\bar{X}_\mu$-norm depends on the diffusion as well as on the convection parameter. The convection term is measured in a dual norm which can be viewed as the infinite dimensional counterpart to the mesh dependent weighted L_2-norm of the convection term in the classical SUPG norm, see *e.g.* [18, 26, 30, 33]. For a more detailed discussion of various versions of norms used for convection-diffusion equations and related effects we refer to [5, 12, 26].

Note that in the optimization problem (6.4) the norm of the boundary penalty is not a dual norm. This allows us to replace the system (3.22), which in or case is a 3×3 block system, by the simpler system

$$\begin{array}{rl} \langle R_{Y_\mu} u_V, v \rangle + \langle B_\mu p_W, v \rangle & = \langle f, v \rangle, \; v \in V, \\ \langle B_\mu^* u_V, q \rangle - \mu \langle p_W, q \rangle_{H_b(\mu)} = & 0, \quad q \in W, \end{array} \tag{6.7}$$

in all practical computations. This system is derived by the same reasoning as in Section 3.1 applied to the first term $\|f - B_\mu p_W\|_{Y_\mu'}$ of the optimization problem (6.4) only. In [5, 34], it is shown how to transform this system to the equivalent saddle point problem (3.22), so that the theory of the present paper still applies to (6.7).

In summary, we have found a stable variational formulation of the convection-diffusion problem (2.6) that fits into the general framework of Section 3.1. However, note that the spaces $\bar{Y}_\mu$ and $\bar{X}_\mu$ may differ even as sets for different $\mu \in \mathcal{P}$, see (6.5), (6.2). Specifically, the dependence of $\bar{Y}_\mu$ on μ lies only in the boundary conditions. However, for a polyhedral domain Ω one can find a finite cover $\{\mathcal{P}_l : l = 1, \ldots, P\}$ of $\mathcal{P}$ so that the outflow boundary portions $\Gamma_+(\mu) = \Gamma_{+,l}$ stay the same for $\mu \in \mathcal{P}_l$. Hence, the spaces $\bar{Y}_\mu, \bar{X}_\mu$ all agree as sets for $\mu \in \mathcal{P}_l$. Clearly, the solution manifold $\mathcal{M}$ (see (3.19)) is a finite union of solution manifolds $\mathcal{M}(l)$ corresponding to the subsets $\mathcal{P}_l$. Since each $\mathcal{M}(l)$ is compact so is the finite union $\mathcal{M}$. Note that for $\mu \in \mathcal{P}_l$ the Riesz map $R_{\bar{Y}_\mu}$ is independent of μ. Therefore, we can apply Theorem 5.5 to each component $\mathcal{P}_l$ leading to the following result.

Corollary 6.1. *The scheme* DoGRE *based on* (6.7) *is rate-optimal.*

In this case the online evaluations can be based on Remark 5.6. Some first numerical experiments are presented in the following section.

6.2. Numerical experiments for convection-diffusion problems

We consider the convection-diffusion problem

$$-\epsilon \Delta p + \begin{pmatrix} \cos \mu \\ \sin \mu \end{pmatrix} \cdot \nabla p + p = 1, \text{ in } \Omega = (0,1)^2, \quad p = 0, \text{ on } \partial\Omega. \tag{6.8}$$

In all test cases we use the variational formulation based on (6.4) regardless of the choice of the truth spaces. First, we treat scenario (a), *i.e.*, with $\epsilon = 2^{-5}$ which is already convection dominated. However, we use a truth space that completely resolves the layers. Specifically, for $X_\mathcal{N}$ and $Y_\mathcal{N}$ we choose bilinear finite elements which are continuous on a rectangular uniform grid of mesh size 2^{-10} and 2^{-11}, respectively. For all computations we used an equidistant sample set $\mathcal{S} \subset \mathcal{P} = [0.2, \pi - 0.2]$ of cardinality 500. Using finite element *a posteriori* error
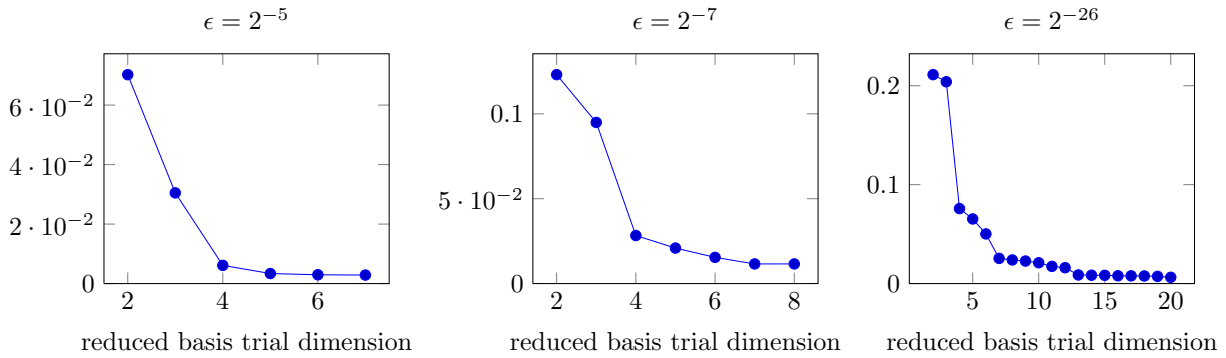
FIGURE 2. Surrogates of the reduced basis approximation for the convection-diffusion problem (6.8).

estimators from [5], based on (5.1), the respective truth-accuracy is bounded by 0.00286384. We note that these *a posteriori* bounds represent the truth residual and hence the energy error only within some fixed constants. This is in contrast to the surrogate bounds for the reduced spaces which are much tighter. The number of adaptively generated basis functions for the reduced test space together with the corresponding constant of the δ-proximality, as well as the maximal surrogate are given in Table 1a. Figure 1a shows a reduced basis solution for the angle $\mu = 0.885115$. The parameter dependent direction of the first order term is visualized by a plane.

The intermediate case $\epsilon = 2^{-7}$ shown in Figure 1b demonstrates how the formulation handles a not fully resolved boundary layer which is not far-off being resolved either.

The other example, scenario (b), refers to the same problem (6.8) again, however, with a very small viscosity $\epsilon = 2^{-26}$ and we choose truth spaces with a mesh size of 2^{-9} and 2^{-10}, respectively. Hence, this case is even more strongly convection dominated and poses difficulties for resolving the boundary layers even for the truth space itself. Since the boundary layers are not resolved (6.4) does not strictly enforce strong boundary conditions at the outflow boundary even in the truth space. Accordingly, the approximate solutions from the reduced space do not satisfy the boundary conditions in a strict sense either. In fact, we choose the same truth spaces as in the preceding experiment. Thus we have to employ the norms (6.5) based on the variational formulation (6.7). The numerical results are summarized in Table 1c and a corresponding reduced basis solution is displayed in Figure 1c.

Figure 2 displays a surrogate plot for the values 2^{-5}, 2^{-7} and 2^{-26} of ϵ. One observes that the error of the reduced basis approximation decays rapidly already for small reduced bases. In fact, since $\inf_{p \in X_\mathcal{N}} \|f - B_\mu p\|_{Y'_\mu} \leq \inf_{p \in X_n} \|f - B_\mu p\|_{Y'_\mu}$, the error of the truth approximation is always a lower bound for the error of the reduced basis approximation. This contrasts standard reduced basis methods where one, in our terminology, chooses $Y_\mathcal{N} = X_\mathcal{N}$. Instead, we assume a larger space $Y_\mathcal{N} \in \mathcal{V}(X_\mathcal{N}, \delta_\mathcal{N})$ which, according to Remark 3.9, implies that the surrogate (5.4) is equivalent to the true error with respect to the infinite dimensional solution, regardless of whether the truth space resolves all solution features like boundary layers or not. Comparing the surrogate plots with the Tables 1, one sees that the reduced basis errors stagnate roughly at the error level of the truth solution. Due to the very small δ-proximality thresholds, the surrogates reflect the true reduced errors very accurately, see (5.5), Proposition 5.1.

Note that in all cases the inner stabilization loop produces at most $m_B = 3$ additional test basis functions for the test space, see Proposition 4.10.

6.3. Transport problems – the worst scenario

We address now the transport equation (2.8) in Section 2.3.2. Aside from its essential appearance in more general kinetic models and Boltzmann type equations, it can be viewed as a "limit" of convection-diffusion

TABLE 1. Numerical results for the convection-diffusion problem (6.8).

(a) $\epsilon = 2^{-5}$, maximal *a posteriori* error 0.00286384.

dimension trial	test	δ	maximal surrogate	surr / *a post*
2	3	2.51e-01	7.02e-02	2.45e+01
3	6	3.75e-01	3.05e-02	1.06e+01
4	7	3.75e-01	6.12e-03	2.14e+00
5	10	3.51e-01	3.38e-03	1.18e+00
6	13	1.87e-01	2.96e-03	1.03e+00
7	16	1.64e-01	2.86e-03	1.00e+00

(b) $\epsilon = 2^{-7}$, maximal *a posteriori* error 0.0115811.

dimension trial	test	δ	maximal surrogate	surr / *a post*
2	5	6.94e-03	1.23e-01	1.06e+01
3	8	1.49e-01	9.49e-02	8.20e+00
4	11	1.22e-02	2.83e-02	2.44e+00
5	14	1.46e-02	2.09e-02	1.81e+00
6	17	5.10e-03	1.54e-02	1.33e+00
7	20	4.29e-03	1.16e-02	1.00e+00
8	23	7.40e-03	1.16e-02	1.00e+00

(c) $\epsilon = 2^{-26}$, maximal *a posteriori* error 0.001055.

dimension trial	test	δ	maximal surrogate	surr / *a post*	dimension trial	test	δ	maximal surrogate	surr / *a post*
2	5	1.35e-03	2.11e-01	2.00e+02	12	33	3.47e-04	1.60e-02	1.52e+01
4	9	1.09e-02	7.58e-02	7.19e+01	14	39	1.10e-04	8.46e-03	8.02e+00
6	15	1.61e-03	5.02e-02	4.76e+01	16	45	9.39e-05	7.87e-03	7.46e+00
8	21	7.99e-04	2.39e-02	2.26e+01	18	51	6.11e-05	7.69e-03	7.29e+00
10	27	3.55e-04	2.10e-02	2.00e+01	20	57	5.28e-05	6.35e-03	6.02e+00

problems. The particular interest lies in the complete lack of viscosity as a "classical" stabilizing ingredient, see *e.g.* [24]. Moreover, as we shall see, the conditions (3.36) and (3.35) do *not* hold simultaneously, not even for suitable subsets of $\mathcal{P}$. Moreover, the parameter dependence will be seen to be significantly less smooth.

We have already proposed a variational formulation (2.9) along with the spaces Y_μ, X_μ in (2.10) endowed with the norms (2.11). With these definitions, the operator $B_\mu : X_\mu \to Y_\mu'$ is an isomorphism with condition number 1, *i.e.* $\|\cdot\|_{X_\mu} = \|\cdot\|_{\hat{X}_\mu} = \|B_\mu \cdot\|_{Y_\mu'}$, see [7]. Notice that in this case the Riesz map R_{Y_μ} are given by

$$R_{Y_\mu} = B_\mu B_\mu^*, \quad i.e., \ (v,w)_{Y_\mu} = \langle B_\mu^* v, B_\mu^* w\rangle, \quad \|\cdot\|_{\hat{X}_\mu} = \|\cdot\|_{L_2(\Omega)},$$

so that (3.43) and (3.36) are valid. Finally, the Riesz maps R_{Y_μ} and R_{X_μ} depend affinely on the parameter so that the double greedy scheme can be applied.

However, since the Y_μ-norm is *not* independent of μ, we cannot evaluate the surrogate given by (5.4) in the usual way. As a remedy, we use the surrogate $R_n'(\mu)$ from (5.8), *i.e.*, we approximate this inverse Riesz map by projecting on the reduced basis space Y_n instead of the truth space $Y_\mathcal{N}$. To ensure that this surrogate is also tight we take up the criterion in Remark 5.3. Specifically, we wish to apply Lemma 5.4 and try to construct suitable pairs $\bar{X}, \bar{Y}$ as follows.

We run the double greedy scheme (possibly) several times which yields the sequences of reduced spaces $X_1^i, X_2^i, \dots$ and $Y_1^i, Y_2^i, \dots$, $i = 0, 1, 2, \dots$ in the ith run of the full double-greedy algorithm. Now, say we stop the first run at index N and define $\bar{X} := X_N^0$. For the second run, we use the same initial spaces as for the first run, however, the calls of UPDATE-INF-SUP(Y_n^1, X_n^1) are replaced by UPDATE-INF-SUP$(Y_n^1, \bar{X} + X_n^1)$, so that δ-proximality is guaranteed for the larger space $\bar{X} + X_n^1$. Then, with the n-dependent choice $\bar{Y} = Y_n^1$, the estimate (5.13) implies that for the second run the surrogates are tight as long as the condition (5.11) is satisfied.

Of course, neither can this latter condition be rigorously checked since we cannot rely on the surrogates, nor have we specified the terminating index $N = N_0$. We briefly sketch now several options of iteratively tightening the surrogates $R_n'(\mu)$. One could stop the first run $i = 0$ at the smallest N_0 for which $R_{N_0}'(\mu)/\tau_\mathcal{N} \le \alpha$ for some $\alpha \ll 1$, where $\tau_\mathcal{N}$ is the truth error tolerance. The second run $i = 1$ with $\bar{X} = \bar{X}^1 = X_{N_0}^0$ will stop at step N_1. In general, the ith run with $\bar{X}^i = \bar{X}^{i-1} + X_{N_{i-1}}^{i-1}$ stops at N_i. One expects that $N_{i+1} \ge N_i$ since the surrogates,

being lower bounds for the true residuals, become tighter as long as $\bar{X}^i$ grows. A practical stopping criterion would be, for instance, that $N_{i+1} \leq N_i$, or $R'_{N_i}(\mu)/\|f - B_\mu p_{N_i}(\mu)\|_{Y'_n} \sim R'_{N_{i+1}}(\mu)/\|f - B_\mu p_{N_{i+1}}(\mu)\|_{Y'_n}$.

An alternative strategy is to apply the double greedy scheme to the defect problem

$$B_\mu \bar{p}(\mu) = f - B_\mu p_{N_0}(\mu), \quad \mu \in \mathcal{P},$$

and form $\bar{X}$ as the sum of $X^0_{N_0}$ and the largest reduced space for the defect problem. Since the relative accuracy to be achieved for the defect problem only needs to meet the constant ξ in (5.11) one expects that a few steps suffice. Since $\bar{X}$ now contains "complementary" information $X^0_{N_0}$ is enlarged more effectively than in the first method.

The upshot of these comments is that investing additional computational offline effort is guaranteed to tighten the surrogates and thereby improves the choice of the reduced spaces. This is in contrast to greedy strategies based on surrogates that are not based on well-conditioned variational formulations and therefore most likely fail to detect the most effective snapshots. These issues will be addressed in forthcoming work.

Since the basis function $\phi_j \in X_n$ can now be orthonormalized in $L_2(\Omega)$ and $\|\cdot\|_{\hat{X}_\mu} = \|\cdot\|_{L_2(\Omega)}$, Theorem 5.5 applies and yields the following result.

Corollary 6.2. *If $R'_n(\mu)$ from (5.8) is based on iterative tightening with $\bar{X}_{i(n)}$ satisfying (5.11) for sufficiently small ξ, then the scheme* DoGre *using* Update-inf-sup *is rate-optimal for $\mathcal{M}_X$.*

We could also reverse the roles of the spaces Y_μ, X_u, choosing $L_2(\Omega)$ as the test space, see [12, 34]. In this case the trial spaces would essentially depend on the parameter μ so that the understanding of the solution set $\mathcal{M}$ is less clear. On the other hand, this choice would correspond to the limit of the formulation (6.7) for vanishing viscosity.

Since (3.35) does not hold we cannot apply Proposition 4.10 to predict a strict *a priori* bound on the number of stabilization steps in Update-δ or Update-Inf-Sup. Adhering to the notation in Section 4.6, we have here $K_\mu = B^*_\mu$, see (2.11). Since in the present case R_X is the identity, as pointed out there, the enrichments of the test spaces Y_n are linear combinations of elements of the form $B^{-*}_{\mu'} B^{-1}_{\mu''} f$ where μ', μ'' are different most of the time, due to the greedy selection. As a consequence, when $f \in L_2(\Omega)$, this means that indeed $B^{-*}_{\mu'} B^{-1}_{\mu''} f \in H^1(\Omega) = Y$. Of course, the H^1-norm may deteriorate when μ', μ'' get closer, which however may be offset to some extent by the expectation that these snapshots are most relevant for the stabilization of solutions with nearby parameters. It is also clear that higher regularity of f would indeed ensure sufficient regularity of the $q \in X^1_n$ (or $q \in \mathcal{B}$), independently of μ', μ'' and hence allows one to control the constants $C(n, \mathcal{N})$ (or $C(\mathcal{B}, \mathcal{N})$). This effect is reflected to some extent by the experiments below.

6.4. Numerical experiments for transport problems

We consider the analog of the convection-diffusion problem (6.8) with zero diffusion $\epsilon = 0$ and corresponding boundary conditions, *i.e.*

$$\begin{pmatrix} \cos \mu \\ \sin \mu \end{pmatrix} \cdot \nabla p + p = 1, \text{ in } \Omega = (0,1)^2, \quad p = 0, \text{ on } \Gamma_-. \tag{6.9}$$

We employ a truth trial space with mesh size 2^{-8}, using discontinuous piecewise bilinear finite elements with proper boundary conditions. To ensure stable truth discretizations, the test truth space is comprised of globally continuous piecewise bilinear finite elements, therefore being contained in $Y = \bigcap_{\mu \in \mathcal{P}} Y_\mu$, on a finer mesh with mesh size 2^{-9} to ensure δ-proximality. Recall that the spaces Y_μ now differ even as sets. The results are shown in Table 2 and a reduced basis solution for the angle $\mu = 0.244579$ is given in Figure 3. Specifically, in addition to the dimensions of the trial and test spaces in columns (1) "trial", (2) "test", it records the values of the surrogates in column (4) "surr", the error between the reduced basis solution and the best L_2-approximation of the exact solution in the truth space in column (6) "rb L_2", the error between the reduced basis solution

TABLE 2. Numerical results for the transport problem 6.9, maximal error truth L_2 0.000109832.

dimension		δ	maximal	maximal error between		surr /
trial	test		surr	rb truth	rb L_2	err
4	11	3.95e-01	8.44e-03	2.45e-02	2.45e-02	3.45e-01
6	17	4.49e-01	7.06e-03	1.40e-02	1.40e-02	5.04e-01
8	25	4.87e-01	4.16e-03	9.05e-03	9.05e-03	4.60e-01
10	33	4.32e-01	3.37e-03	5.74e-03	5.74e-03	5.87e-01
12	40	4.83e-01	2.65e-03	4.65e-03	4.65e-03	5.71e-01
14	48	4.23e-01	1.64e-03	3.39e-03	3.39e-03	4.83e-01
16	57	4.32e-01	1.50e-03	2.56e-03	2.56e-03	5.84e-01
18	65	4.66e-01	1.17e-03	2.33e-03	2.33e-03	5.03e-01
20	74	4.16e-01	1.21e-03	2.10e-03	2.10e-03	5.77e-01
22	83	3.83e-01	1.02e-03	1.93e-03	1.93e-03	5.29e-01
24	91	4.05e-01	7.27e-04	1.58e-03	1.58e-03	4.61e-01

TABLE 3. Numerical results for the transport problem 6.9 after a single cycle of iterative tightening. Maximal error truth L_2 0.0154814.

dimension		δ	maximal	maximal error between		surr /
trial	test		surr	rb truth	rb L_2	err
First reduced basis creation						
20	81	3.73e-01	2.71e-02	5.46e-02	5.62e-02	4.82e-01
Second reduced basis creation						
10	87	3.51e-01	6.45e-02	7.40e-02	7.53e-02	8.57e-01

and the truth solution in column (5) "rb truth", and finally in column (7) "surr/err" the ratio between the computed surrogate and the error in "rb L_2". All values reflect the worst case over the parameter range.

As pointed out above, unlike the convection-diffusion problem, the surrogate (5.8) for the transport problem is not necessarily well-conditioned from the start. Therefore, Table 2 contains one column which shows the ratio of the surrogate compared to the true error of the reduced basis approximation, maximized over a sample of the angles with the largest values of the surrogate. Although this is at this point not founded rigorously, we see that this ratio stays uniformly bounded with respect to the size of the reduced basis. Hence it already does reflect the accuracy of the reduced model. However, the ratio is not close to one yet, as it would be for a well-conditioned surrogate given by the truth-exact evaluation of the residual corresponding to a well-conditioned variational formulation. To further improve this ratio by approximating the residual more accurately, we resort to iterative tightening as described above.

The results for a single iteration are recorded in Table 3. It is seen that already after a single run the ratio of the surrogate and the true error between the reduced basis approximation and true solution has become much closer to one.

In agreement with the discussion in Section 4.6 the experiments show that a slightly larger number of test basis functions than for the convection-diffusion problem is needed here. In particular, unlike in the elliptic case non-smooth data (right hand side, boundary shape, and boundary conditions) affect the smoothness of the dependence of the solutions on the parameter. In our examples at most a low order polynomial decay of the n-widths can be expected. According to Summary 4.11 in Section 4.6, since the right hand side is actually smooth, we expect that $\sigma_{n,j}$, defined in (4.18), that controls the termination of the inner stabilization loop drops below the desired $\delta < 1$ after an acceptable bounded number of steps independent of the dimension of the truth space. In fact, one observes that the growth of the test basis stays surprisingly moderate.
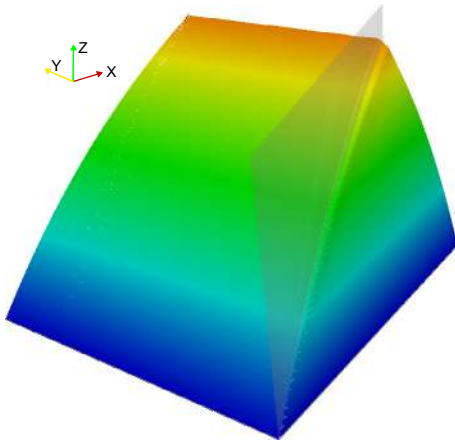
FIGURE 3. Solution of the transport problem (6.9), with reduced basis of dimension $n = 24$, $m(n) = 91$ and angle $\mu = 0.244579$.
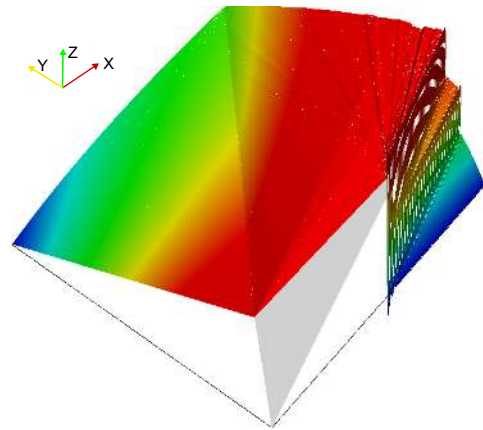

FIGURE 4. Solution of the transport problem (6.10), with reduced basis of dimension $n = 24$, $m(n) = 96$ and angle $\mu = 0.256311$.

Finally, Table 4 and Figure 4 show the results for the problem

$$\begin{pmatrix} \cos \mu \\ \sin \mu \end{pmatrix} \cdot \nabla p + p = \begin{cases} 0.5 \ x < y \\ 1 \ x \geq y \end{cases}, \quad \text{in } \Omega = (0,1)^2, \quad p = \begin{cases} 1 - y \ x \leq 0.5 \\ 0 \quad x > 0.5 \end{cases}, \quad \text{on } \Gamma_-. \tag{6.10}$$

Now the right hand side as well as the boundary conditions exhibit jump discontinuities where the latter is transported trough the domain. This causes a further significant reduction of the smoothness of the dependence on the solutions on the parameter. Problem (6.10) therefore represents an extreme example involving interacting jump discontinuities caused by the right hand side and by the boundary conditions. The small ripples observed in the solution plot Figure 4 originate from the superposition of the jumps of the various snapshots involved in the solution. Since they do neither grow nor expand one can conclude that the scheme is in fact stable.

As indicated before, varying the transport direction for such data shows that the dependence of the solution on the parameter is even less smooth than in the previous example so that the Kolmogorov widths of the solution manifold are expected to decay more slowly. Hence the greedy errors cannot decay too rapidly either. Again, by Summary 4.11, the quantities $\sigma_{n,j}$ in (4.18), estimating the number of stabilization steps for X_n, are expected to decay even more slowly than in the case of zero boundary conditions. Table 4 confirms this in that slightly more test basis functions are generated than in example 6.9. Nevertheless, one observes that in the initial phase already a few reduced basis functions decrease the error very effectively so that a reduced space with trial dimension as low as ten realizes an accuracy that would require a conventional finite element space of much larger dimension. Overall, the performance, at least in the given range of truth accuracy, is only slightly weaker than for the milder case of zero inflow boundary conditions. The precise implications on the approximation of functionals of the solution and possible strategies for alternative ways of enriching the trial dictionary will be explored in forthcoming work.

7. GENERAL SADDLE POINT PROBLEMS

The crucial role of saddle point problems for the generation of well-conditioned variational formulations is apparent from the preceding discussion. On the other hand, the concepts developed in this context have an immediate bearing on more general saddle point problems of "classical type". By this we mean (parameter dependent)

TABLE 4. Numerical results for the transport problem 6.10, maximal error truth L_2 0.0154814.

| dimension		δ	maximal	maximal error between		surr /
trial	test		surr	rb truth	rb L_2	err
4	14	4.97e-01	5.91e-02	1.29e-01	1.30e-01	4.54e-01
6	23	4.92e-01	4.29e-02	1.00e-01	1.02e-01	4.22e-01
8	31	4.29e-01	4.34e-02	7.78e-02	7.95e-02	5.46e-01
10	40	4.15e-01	3.84e-02	7.78e-02	7.95e-02	4.83e-01
12	49	3.71e-01	3.48e-02	7.40e-02	7.53e-02	4.63e-01
14	57	3.76e-01	3.12e-02	6.20e-02	6.41e-02	4.87e-01
16	64	3.74e-01	2.99e-02	6.20e-02	6.41e-02	4.67e-01
18	73	4.63e-01	2.86e-02	6.20e-02	6.41e-02	4.47e-01
20	81	3.73e-01	2.71e-02	5.46e-02	5.62e-02	4.82e-01
22	87	4.09e-01	2.42e-02	5.46e-02	5.62e-02	4.32e-01
24	96	3.91e-01	2.51e-02	4.51e-02	4.79e-02	5.25e-01
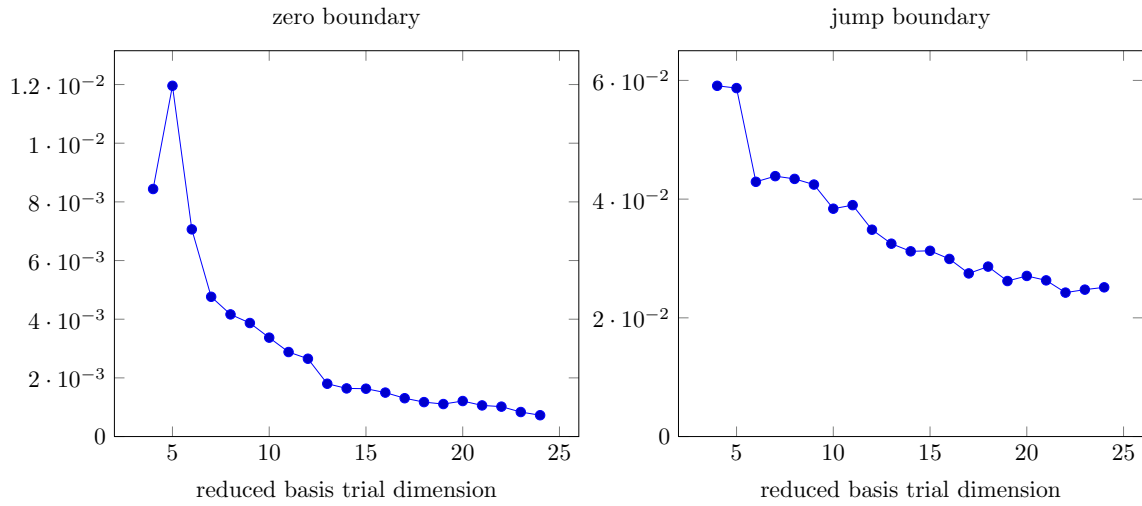
FIGURE 5. Surrogates of the reduced basis approximation for the transport problems (6.9) and (6.10).

variational formulations *e.g.* of the Stokes system or those arising in mixed formulations and constrained optimization problems. To see this, it is useful to point out the main distinctions between the two settings when considering the following general formulation for parameter dependent bilinear forms $a_\mu(\cdot,\cdot) : Y_\mu \times Y_\mu \to \mathbb{R}$, $b_\mu(\cdot,\cdot) : X_\mu \times Y_\mu \to \mathbb{R}$ such that for $\mu \in \mathcal{P}$

$$\begin{aligned} a_\mu(u(\mu),v) + b(p(\mu),v) &= \langle f,v\rangle, \quad v \in Y_\mu, \\ b_\mu(q,u(\mu)) &= \langle g,q\rangle, \quad q \in X_\mu. \end{aligned} \tag{7.1}$$

For classical problems the following conditions

$$|a_\mu(v,w)| \le C_a(\mu)\|v\|_{Y_\mu}\|w\|_{Y_\mu}, \quad |b_\mu(q,v)| \le C_b(\mu)\|v\|_{Y_\mu}\|q\|_{X_\mu}, \quad v,w \in Y_\mu \quad q \in X_\mu, \tag{7.2}$$

as well as

$$\inf_{q\in X_\mu}\sup_{v\in Y_\mu}\frac{b_\mu(q,v)}{\|q\|_{X_\mu}\|v\|_{Y_\mu}} \ge \beta(\mu), \qquad a_\mu(v,v) \ge c_a(\mu)\|v\|_{Y_\mu}^2, \ v \in V(\mu), \tag{7.3}$$

where

$$V(\mu) := \{v \in Y_\mu : b_\mu(q, v) = 0, \, \forall \, q \in M\} = \ker B_\mu^*,$$

are usually met.

In comparison, the specific structure of the "stabilizing" saddle point problem (3.17) is the following.

Remark 7.1. B_μ is an isomorphism and for $a_\mu(v, w) := \langle R_{Y_\mu} v, w \rangle$ one has $V(\mu) = \{0\}$ and condition (7.3) holds with $c_a(\mu) = C_a(\mu) = 1$ even on Y_μ. Moreover, when using the norm $\| \cdot \|_{\hat{X}_\mu}$ on X_μ, (7.2) holds with $C_b(\mu) = \beta(\mu) = 1$. However, on the downside, one may encounter failure of either (3.35) or (3.36).

For the classical problems considered in [13, 14, 29] one can state the following.

Remark 7.2. The conditions (3.35), (3.36) are both satisfied so that a single reference norm $\| \cdot \|_{Y \times X}$ can be used. Hence, that renormation (3.6) is not necessary for achieving tightness of residual based surrogates which now involve both component spaces $Y' \times X'$ which can be evaluated by the standard offline-online decomposition, see *e.g.* [13, 14].

It is well-known (see *e.g.* [2]) that, given (7.2) and (7.3), the validity of the mapping property **MP** and the best approximation property **BAP** hinges again on the inf-sup condition

$$\inf_{q \in W} \sup_{v \in V} \frac{b_\mu(q, v)}{\|v\|_{Y_\mu} \|q\|_{X_\mu}} \geq \beta_{V,W}(\mu). \tag{7.4}$$

For classical problems it is known for $V = Y, W = X$, has to be ensured for the truth spaces $V = Y_\mathcal{N}, W = X_\mathcal{N}$ through suitably chosen finite element spaces, say, and again need to be ensured by stabilizing strategies for the reduced spaces $V = Y_n, W = X_n$.

In view of Remark 7.2, both schemes UPDATE-δ and UPDATE-INF-SUP can be applied. Since the spaces Y_n no longer just serve as stabilizers but need to contribute to the target approximation accuracy of the full solution manifold

$$\mathcal{M} := \{[u(\mu), p(\mu)] : \text{ solves } (7.1), \, \mu \in \mathcal{P}\} =: \mathcal{M}_Y \times \mathcal{M}_X, \tag{7.5}$$

the only changes that need to be incorporated in a slightly modified version DoGre2 of DoGre are:

- In **Algorithm** (4) replace Step 5 by:
  Set
  $$\text{span}\,\{X_n, \hat{p}\} \to X_n, \qquad \text{span}\,\{Y_n, \hat{u}\} \to Y_n,$$
  i.e., both component spaces are updated in the outer greedy step.
- In **Algorithm** (5) Step 5 is replaced by
  $Y_n, X_n \leftarrow \text{UPDATE-APPROXIMATION}(Y_n, X_n)$
- Replace the surrogate by
  $$R^*(\mu, V \times W) := \|f - A_\mu u_{W,V}(\mu) - B_\mu p_{W,V}(\mu)\|_{Y'_\mathcal{N}} + \|g - B^* u_{W,V}(\mu)\|_{X'_\mathcal{N}},$$

  see [13, 14].

Clearly, under the given assumptions (3.35), (3.36), $\mathcal{M}$ is compact. Denoting again by $p_n(\mu), u_n(\mu)$ the solution components produced by the scheme DoGre2 and comparing the greedy errors (5.17)

$$\sigma_n(\mathcal{M}) := \sup_{\mu \in \mathcal{P}} \big\{\|p(\mu) - p_n(\mu)\|_X + \|u(\mu) - u_n(\mu)\|_Y\big\},$$

with the n-widths $d_n(\mathcal{M})_{X \times Y}$ and keeping Proposition 4.10 in mind, we extend the results in [13, 14, 29] as follows.

Corollary 7.3. *The scheme* DoGre2 *applied to* (7.1) *is under the above assumptions rate-optimal.*

8. Concluding remarks

The generation of well-conditioned variational formulations for non-coercive or indefinite problems has been proposed as the central ingredient of a general strategy for constructing tight surrogates for RBMs also for such problem classes. In contrast to previous work, well-conditioned tight surrogates are obtained in a feasible way in all settings warranting a near-optimal performance of the corresponding RBM, which does not seem to be achievable with the aid of previously known concepts. We emphasize that these concepts apply as well to space-time discretizations of unsteady problems (see [7]) offering interesting perspectives with regard to robustly capturing long-term dynamics. The presented application to two simple model problems is to be viewed as a first proof of concept. The two examples are to bring out some essential obstructions and raise issues that have so far not been addressed in this context. In particular, they hint at the principal limitations of RBMs in their standard formulations, especially regarding the smoothness of the parameter dependence.

References

[1] 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.

[2] F. Brezzi and M. Fortin, Mixed and Hybrid Finite Element Methods, in vol. 15 of *Springer Ser. Comput. Math.* Springer-Verlag (1991).

[3] 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: M2AN* **46** (2012) 595–603.

[4] J.M. Cascon, C. Kreuzer, R.H. Nochetto and K.G. Siebert, Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.* **46** (2008) 2524–2550.

[5] A. Cohen, W. Dahmen and G. Welper, Adaptivity and Variational Stabilization for Convection-Diffusion Equations. *ESAIM: M2AN* **46** (2012) 1247–1273.

[6] W. Dahmen, Parameter dependent transport equations, in Workshop J.L.L.-SMP: Reduced Basis Methods in High Dimensions. Available at http://www.ljll.math.upmc.fr/fr/archives/actualites/2011/workshop_ljll_smp_rbihd.html

[7] W. Dahmen, C. Huang, C. Schwab and G. Welper, Adaptive Petrov−Galerkin methods for first order transport equations. *SIAM J. Numer. Anal.* **50** (2012) 2420–2445.

[8] L.F. Demkowicz and J. Gopalakrishnan, A class of discontinuous Petrov−Galerkin Methods I: The transport equation. *Comput. Methods Appl. Mech. Engrg.* **199** (2010) 1558–1572.

[9] L. Demkowicz and J. Gopalakrishnan, A class of discontinuous Petrov−Galerkin methods. Part II: Optimal test functions. *Numer. Methods for Partial Differ. Equ.* **27** (2011) 70–105.

[10] S. Deparis, Reduced basis error bound computation of parameter-dependent Navier-Stokes equations by the natural norm approach. *SIAM J. Numer. Anal.* **46** (2008) 2039–2067.

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

[12] A. Ern and J.-L. Guermond, Theory and practice of finite elements. Springer (2004).

[13] A. Gerner and K. Veroy-Grepl, Certified reduced basis methods for parametrized saddle point problems, preprint (2012). To appear in *SIAM J. Sci. Comput.*

[14] A.-L. Gerner and K. Veroy, Reduced basis *a posteriori* error bounds for the Stokes equations in parameterized domains: A penalty approach. *M3AS: Math. Models Methods Appl. Sci.* **21** (2011) 2103–2134.

[15] M.A. Grepl, Certified Reduced Basis Methods for Nonaffine Linear Time-Varying and Nonlinear Parabolic Partial Differential Equations. *M3AS: Math. Models Methods Appl. Sci.* **22** (2012) 40.

[16] M.A. Grepl and A.T. Patera, *A Posteriori* Error Bounds for Reduced-Basis Approximations of Parametrized Parabolic Partial Differential Equations. *ESAIM: M2AN* **39** (2005) 157–181.

[17] B. Haasdonk, Convergence rates for the POD-greedy method. *ESAIM: M2AN* **47** (2013) 859–873.

[18] T. Hughes and G. Sangalli, Variational Multiscale Analysis: the Fine-scale Green's Function, Projection, Optimization, Localization, and Stabilized Methods. *SIAM J. Numer. Anal.* **45** (2007) 539–557.

[19] G. Kanschat, E. Meinköhn, R. Rannacher and R. Wehrse, Numerical methods in multidimensional radiative transfer, Springer (2009).

[20] G.G. Lorentz, M. von Golitschek and Yu. Makovoz, Constructive approximation: *Advanced problems*, vol. 304. Springer Grundlehren, Berlin (1996).

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

[22] 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 J. Numer. Anal.* **43** (2005) 2098–2120.

[23] N.-C. Nguyen, G. Rozza and A.T. Patera, Reduced basis approximation for the time-dependent viscous Burgers' equation. *Calcolo* **46** (2009) 157–185.

[24] T. Patera and K. Urban, An improved error bound for reduced basis approximation of linear parabolic problems, submitted to *Mathematics of Computation* (in press 2013).

[25] A.T. Patera and 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*.

[26] H.-J. Roos, M. Stynes and L. Tobiska, Robust Numerical Methods for Singularly Perturbed Differential Equations, in vol. 24 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2nd Edition (2008).

[27] G. Rozza and D.B.P. Huynh and A. Manzoni, Reduced basis approximation and *a posteriori* error estiamtion for Stokes flows in parametrized geometries: roles of the inf-sup stability constants, *Numer. Math.* DOI: 10.1007/s00211-013-0534-8.

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

[29] G. Rozza and K. Veroy, On the stability of reduced basis techniques for Stokes equations in parametrized domains. *Comput. Methods Appl. Mechanics Engrg.* **196** (2007) 1244–1260.

[30] G. Sangalli, A uniform analysis of non-symmetric and coercive linear operators. *SIAM J. Math. Anal.* **36** (2005) 2033–2048.

[31] M. Schlottbom, On Forward and Inverse Models in Optical Tomography, Ph.D. Thesis. RWTH Aachen (2011).

[32] S. Sen, K. Veroy, D.B.P. Huynh, S. Deparis, N.C. Nguyn and A.T. Patera, Natural norm *a posteriori* error estimators for reduced basis approximations. *J. Comput. Phys.* **217** (2006) 37–62.

[33] R. Verfürth, Robust *a posteriori* error estimates for stationary convection-diffusion equations. *SIAM J. Numer. Anal.* **43** (2005) 1766–1782.

[34] G. Welper, Infinite dimensional stabilization of convection-dominated problems, Ph.D. Thesis. RWTH Aachen (2012).

[35] J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo and V. Calo, A class of discontinuous Petrov−Galerkin methods. Part IV: Wave propagation. *J. Comput. Phys.* **230** (2011) 2406–2432.