Mathematical Modelling and Numerical Analysis The original publication is available at www.esaim-m2an.org
Modélisation Mathématique et Analyse Numérique

AN ADAPTIVE PARAMETRIZED-BACKGROUND DATA-WEAK APPROACH TO VARIATIONAL DATA ASSIMILATION

Tommaso Taddei [1]

Abstract. We present an Adaptive Parametrized-Background Data-Weak (APBDW) approach to the steady-state variational data assimilation (state estimation) problem for systems modeled by partial differential equations. The variational formulation is based on the Tikhonov regularization of the PBDW formulation [Y Maday, AT Patera, JD Penn, M Yano, Int J Numer Meth Eng, 102(5), 933-965] for pointwise noisy measurements. We propose an adaptive procedure based on *a posteriori* estimates of the L^2 state-estimation error to improve performance. We also present *a priori* estimates for the L^2 state-estimation error that motivate the approach and guide the adaptive procedure. We provide numerical experiments for a synthetic acoustic problem to illustrate the different elements of the methodology, and we consider an experimental thermal patch configuration to demonstrate the applicability of our approach to real physical systems.

Résumé. Nous présentons une méthode variationelle d'assimilation de données pour des systèmes modélisés par des équations aux dérivées partielles, nommément "Adaptive-Parameterized-Background Data-Weak" (APBDW). Cette approche est fondée sur la régularisation Tychonoff de la formulation PBDW [Y Maday, AT Patera, JD Penn, M Yano, Int J Numer Meth Eng, 102(5), 933-965], et consiste en une procédure adaptative pour considérer le bruit expérimental. Des estimations a priori et a posteriori de l'état L^2 (estimation d'erreur) motivent l'approche et servent de guide la procédure adaptative. Nous présentons des résultats numériques pour deux problémes de modéle synthétique pour illustrer les éléments de la méthodologie. Nous considérons aussi une configuration expérimentale de patch thermique pour montrer que notre approche est applicable dans le cadre de systèmes physiques.

1991 Mathematics Subject Classification. 62-07,93E24.

February 26, 2018.

1. Introduction

Data assimilation refers to the estimation of the state u^{true} of a physical system over the domain of interest $\Omega \subset \mathbb{R}^d$ by combining experimental data with a mathematical model of the dynamics of the system. For real-time and in situ applications, data assimilation techniques should provide an estimate of the state rapidly with little or no communication with extensive offline resources. Furthermore, for safety reasons, it is key to certify the reliability of our estimate using either probabilistic (i.e., confidence intervals) or deterministic (i.e., error bounds) approaches.

Keywords and phrases: variational data assimilation; parametrized partial differential equations; model order reduction; kernel methods

[1] Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA, ttaddei@mit.edu

The goal of this work is to develop a variational data assimilation procedure that combines a parameterized best-knowledge (bk) mathematical model and experimental data to rapidly obtain a reliable estimate of the state $u^{\text{true}} \in \mathcal{U}$ over Ω. In this work, we shall focus on steady-state data assimilation. We denote by $\{y_m\}_{m=1}^M$ the set of experimental measurements, and we denote by $u^{\text{bk}}(\mu) \in \mathcal{U}$ the solution to our parameterized bk mathematical model for the parameter value $\mu \in \mathcal{P}^{\text{bk}}$, $G^{\text{bk},\mu}(u^{\text{bk}}(\mu)) = 0$. Here, the space $\mathcal{U} = \mathcal{U}(\Omega)$ is a suitable Hilbert space defined over Ω, $G^{\text{bk},\mu}(\cdot)$ denotes the parameterized bk mathematical model, and $\mathcal{P}^{\text{bk}} \subset \mathbb{R}^P$ reflects the uncertainty in the value of the parameters associated with the model. Since experimental apparatuses are typically affected by errors, measurements are in general of the form $y_m = \ell_m^o(u^{\text{true}}) + \epsilon_m$, where $\ell_m^o : \mathcal{U} \to \mathbb{R}$ is a known functional and ϵ_m reflects the observational noise. On the other hand, the uncertainty in the parameters of the model leads to the definition of the bk manifold $\mathcal{M}^{\text{bk}} := \{u^{\text{bk}}(\mu) : \mu \in \mathcal{P}^{\text{bk}}\} \subset \mathcal{U}$, which collects the solution to the parameterized bk model for all values of the parameter in $\mathcal{P}^{\text{bk}}$.

In [36,37], Maday et al. introduced the so-called Parameterized-Background Data-Weak (PBDW) approach. The key idea of the PBDW formulation is to seek an approximation $u^\star = z^\star + \eta^\star$ to the true field u^{true} employing projection-by-data. The first contribution to $u^\star$, $z^\star \in \mathcal{Z}_N$, is the "deduced background estimate." The linear N-dimensional space $\mathcal{Z}_N \subset \mathcal{U}$ is informed by the bk manifold $\mathcal{M}^{\text{bk}}$, which we hope is close to the true field. The second contribution to $u^\star$, $\eta^\star \in \mathcal{U}_M$, is the "update estimate". The linear M-dimensional space $\mathcal{U}_M$ is the span of the Riesz representations of the M observation functionals $\{\ell_m^o\}_{m=1}^M$. The pair $(z^\star, \eta^\star) \in \mathcal{Z}_N \times \mathcal{U}_M$ is then computed by searching for $\eta^\star$ of minimum norm subject to the observation constraints $\ell_m^o(z^\star + \eta^\star) = y_m$ for $m = 1, \ldots, M$. While the background estimate incorporates our *a priori* knowledge of the state, the update addresses the deficiencies of the bk model by improving the approximation properties of the search space.

As discussed in the original papers, PBDW provides some new contributions. First, the variational formulation facilitates the construction of *a priori* error estimates, which might guide the optimal choice of the experimental observations. Second, the background space $\mathcal{Z}_N$ accommodates anticipated uncertainty associated with the parameters of the model; the construction of $\mathcal{Z}_N$ based on $\mathcal{M}^{\text{bk}}$ relies on the application of parametric Model Order Reduction (pMOR) techniques. Third, unlike standard least-squares methods, PBDW provides a mechanism — the update $\eta^\star$ — to correct the deficiencies of the bk model. Finally, projection-by-data, as opposed to projection-by-model, implies that the parameterized model is not directly used during the data assimilation procedure. This feature significantly simplifies the computational procedure and guarantees real-time responses.

We observe that several of these ingredients have appeared in different contexts. The variational formulation is built upon least-squares ([25, 30]); the correction of unmodeled physics through Riesz representation of observation functionals is first introduced in the work by Bennett ([2,3]). On the other hand, the bk background space $\mathcal{Z}_N$, as opposed to a background singleton element in the original 3D-VAR ([32, 33]), is found in the context of gappy Proper Orthogonal Decomposition ([17, 59]), Generalized Empirical Interpolation Method (GEIM, [34,35]), and nearfield acoustical holography (NAH, [10,60]). Finally, the use of model order reduction techniques within the context of data assimilation is found in the already-mentioned gappy-POD and GEIM, and in a number of recent papers in the context of 4D-VAR ([9, 14, 49, 50, 54, 62]).

In this paper, we present an adaptive Parameterized-Background Data-Weak (APBDW) approach that extends the original PBDW formulation to the case of pointwise noisy measurements; $\ell_m^o := \delta_{x_m}$ for some $x_m \in \Omega$, $m = 1, \ldots, M$. Our approach is based on the Tikhonov regularization of the PBDW formulation; to explain the connection with other data assimilation formulations, we derive APBDW from the 3D-VAR minimization statement, and the partial-spline model ([56, Chapter 9]). The extension to pointwise measurements is based on the theory of Reproducing Kernel Hilbert Spaces (RKHS, [1]) and exploits the connection with kernel methods for regression ([42, 53]).

We rely on an adaptive procedure for the selection of the hyper-parameters associated with the Tikhonov penalization term, and with the characteristic length-scale of the kernel. The adaptive procedure chooses the hyper-parameters that minimize an estimate of the L^2 state-estimation error on a validation dataset. The adaptive procedure employed in this work is based on holdout validation (see, e.g., [21, Chapter 7] and [27]),

and relies on the Monte Carlo *a posteriori* error estimation procedure discussed in [51]. Adaptation allows us to properly weight the trust in the best-knowledge model.

We further present a complete *a priori* analysis for the $L^2(\Omega)$ state-estimation error in presence of noise. We both consider the case of homoscedastic random noise and the case of systematic noise. This analysis motivates the approach from a theoretical standpoint, and guides the adaptive procedure. We anticipate that our result for systematic noise can be seen as a generalization of the *a priori* result proved in [28].

The theory of RKHS allows us to consider spaces $\mathcal{U}$ for which the Riesz representers $\{K_{x_m}\}_m$ associated with the observation functionals $\{\delta_{x_m}\}_m$ are explicitly known. We demonstrate that explicit expressions for the representers greatly improve the flexibility of the approach; in addition, we find much faster convergence with respect to the number of measurements M than in the approach presented in [36, 37]. We here remark that pointwise measurements represent a convenient idealization. We have indeed that physical transducers estimate averages over a finite spatial region. However, if the radius of the region is sufficiently small compared to the characteristic length-scale of the field of interest, we can replace local averages with pointwise observations.

The outline of the paper is as follows. In section 2, we present the formulation and the well-posedness analysis. We further relate our formulation to a number of other methods proposed in the data assimilation and statistical learning literature. In section 3, we present *a priori* and *a posteriori* estimates for the $L^2(\Omega)$ state-estimation error. In section 4, we exploit the error analysis to design an adaptive procedure for the selection of the parameters. In section 5, we present numerical results for a synthetic acoustic problem. Finally, in section 6, we present the results for a physical thermal patch configuration.

2. Formulation

2.1. Preliminaries

By way of preliminaries, we introduce notation used throughout the paper. Given the Lipschitz domain $\Omega \subset \mathbb{R}^d$, we introduce the space of continuous functions over Ω, $C(\Omega)$. Then, we introduce the Hilbert space $\mathcal{U}$ such that $\mathcal{U} \subset C(\Omega)$; we endow $\mathcal{U}$ with the inner product $(\cdot,\cdot)$ and the induced norm $\|\cdot\| = \sqrt{(\cdot,\cdot)}$. For any closed linear subspace $\mathcal{Q} \subset \mathcal{U}$, we denote by $\Pi_{\mathcal{Q}} : \mathcal{U} \to \mathcal{Q}$ the orthogonal projection operator onto $\mathcal{Q}$, and we denote by $\mathcal{Q}^\perp$ its orthogonal complement. Given $x \in \Omega$, we denote by $K_x \in \mathcal{U}$ the Riesz element associated to the corresponding point evaluation functional, $(K_x, f) = f(x)$ for all $f \in \mathcal{U}$, and we introduce the function $K : \Omega \times \Omega \to \mathbb{R}$ such that $K(x,y) = (K_x, K_y)$ for all $x, y \in \Omega$.

Given a random variable X, we denote by $\mathbb{E}[X]$ and by $\mathbb{V}[X]$ the mean and the variance, where $\mathbb{E}$ denotes expectation. We denote by $X \sim \mathcal{N}(m, \sigma^2)$ a Gaussian random variable with mean m and variance σ^2. Similarly, we denote by $X \sim \text{Uniform}(\Omega)$ an uniform random variable over Ω. Furthermore, we refer to an arbitrary random variable ε such that $\mathbb{E}[\varepsilon] = 0$ and $\mathbb{V}[\varepsilon] = \sigma^2$ using the notation $\varepsilon \sim (0, \sigma^2)$.

2.2. Problem statement

We aim to estimate the deterministic state $u^{\text{true}} \in \mathcal{U}$ over the domain of interest $\Omega \subset \mathbb{R}^d$. We shall afford ourselves two sources of information: a best-knowledge (bk) mathematical model

$$G^{\text{bk},\mu}(u^{\text{bk}}(\mu)) = 0, \tag{1}$$

where $\mu \in \mathcal{P}^{\text{bk}} \subset \mathbb{R}^P$ indicates a set of uncertain parameters associated with the model; and M experimental observations $\mathcal{Y}_M = \{y_1, \ldots, y_M\}$ associated to the M distinct observation centers $\mathcal{X}_M = \{x_1, \ldots, x_M\} \subset \Omega$

$$y_m = u^{\text{true}}(x_m) + \epsilon_m, \quad m = 1, \ldots, M. \tag{2}$$

Here, $\mathcal{P}^{\text{bk}} \subset \mathbb{R}^P$ is a confidence region for the true values of the parameters of the model, while $\{\epsilon_m\}_{m=1}^M$ are unknown disturbances caused by either systematic error in the data acquisition system or experimental

random noise. For purposes of exposition, we introduce the dataset $\mathcal{D}_M = \{(x_m, y_m)\}_{m=1}^M$ associated with the experimental observations (2), and the bk manifold $\mathcal{M}^{\rm bk} = \{u^{\rm bk}(\mu) : \mu \in \mathcal{P}^{\rm bk}\}$ associated with (1).

If $\mathcal{P}^{\rm bk} = \{\bar\mu\}$, we propose to estimate the state $u^{\rm true}$ as follows:

$$u^\star_\xi := \arg\min_{u\in\mathcal{U}} \xi\|u - u^{\rm bk}(\bar\mu)\|^2 \,+\, V_M(u), \tag{3a}$$

where

$$V_M(u) = \frac{1}{M}\sum_{m=1}^M \left(u(x_m) - y_m\right)^2. \tag{3b}$$

The first term penalizes the distance of the state estimate from the background, while the second term penalizes data misfit; here, the parameter $\xi > 0$ regulates the relative importance of the background $u^{\rm bk}(\bar\mu)$ compared to the data. We observe that if $\epsilon_1, \ldots, \epsilon_M$ are independent identically distributed random disturbances such that $\mathbb{E}[\epsilon_m] = 0$, $\mathbb{E}[\epsilon_m\epsilon_{m'}] = \sigma^2\delta_{m,m'}$ then (3) corresponds to the 3D-VAR statement ([4, Chapter 2], [33]).

If we consider $\mathcal{P}^{\rm bk} \neq \{\bar\mu\}$, we can generalize (3) as follows:

$$(\mu^\star_\xi, u^\star_\xi) := \arg\min_{(\mu,u)\in\mathcal{P}^{\rm bk}\times\mathcal{U}} \xi\|u - u^{\rm bk}(\mu)\|^2 \,+\, V_M(u) \tag{4}$$

Formulation (4) is known as partial spline model ([56, Chapter 9]), and can also be restated in terms of the update $\eta^\star_\xi = u^\star_\xi - u^{\rm bk}(\mu^\star_\xi)$:

$$(\mu^\star_\xi, \eta^\star_\xi) := \arg\min_{(\mu,\eta)\in\mathcal{P}^{\rm bk}\times\mathcal{U}} \xi\|\eta\|^2 \,+ V_M(u^{\rm bk}(\mu) + \eta). \tag{5}$$

We observe that (4) (and equivalently (5)) is non-convex in μ; furthermore, evaluations of the map $\mu \mapsto u^{\rm bk}(\mu)$ involve the solution to the bk model. Therefore, it is not suitable for real-time computations. We remark that statement (4) is not the only possible variational formulation that incorporates parametric uncertainty in the model. Here, the bk model enters in the objective function, in other approaches the model is employed as constraint for the minimization of V_M; in the latter class of methods, the parameter ξ can be interpreted as a Lagrangian multiplier, and enters directly into the minimization statement. We refer to [40] for a representative example of the second class of approaches for cardiovascular applications.

If we introduce the rank-N approximation ([13]) of the bk field $u^{\rm bk}(\mu)$, $u^{\rm bk}_N(x,\mu) = \sum_{n=1}^N \,\phi_n(\mu)\,\zeta_n(x)$, for $x \in \Omega$ and $\mu \in \mathcal{P}^{\rm bk}$, we can approximate statement (5) as

$$(\mu^\star_\xi, \eta^\star_\xi) := \arg\min_{(\mu,\eta)\in\mathcal{P}^{\rm bk}\times\mathcal{U}} \xi\|\eta\|^2 \,+ V_M\left(\sum_{n=1}^N \phi_n(\mu)\zeta_n + \eta\right) \tag{6}$$

Then, we can relax (6) as $(\boldsymbol{\phi}^\star_\xi, \eta^\star_\xi) = \arg\min_{(\boldsymbol{\phi},\eta)\in\mathbb{R}^N\times\mathcal{U}} \xi\|\eta\|^2 \,+\, V_M\left(\sum_{n=1}^N \phi_n\,\zeta_n + \eta\right)$, which can also be rewritten as

$$(z^\star_\xi, \eta^\star_\xi) := \arg\inf_{(z,\eta)\in\mathcal{Z}_N\times\mathcal{U}} J_\xi(z,\eta) := \xi\|\eta\|^2 \,+\, V_M(z+\eta), \tag{7}$$

where $\mathcal{Z}_N = \text{span}\{\zeta_n\}_{n=1}^N \subset \mathcal{U}$ is the N-dimensional linear space induced by $\{\zeta_n\}_{n=1}^N$. We further denote by $u^\star_\xi = z^\star_\xi + \eta^\star_\xi$ the corresponding state estimate.

Statement (7) is the Adaptive Parametrized-Background Data-Weak (APBDW) formulation, and $u^\star_\xi = z^\star_\xi + \eta^\star_\xi$ is the APBDW state estimate. We observe that APBDW is a (convex) relaxation of the partial spline model for a parametric affine background: instead of penalizing the distance between the state estimate and the manifold $\mathcal{M}^{\rm bk} = \{u^{\rm bk}(\mu) : \mu \in \mathcal{P}^{\rm bk}\}$, we penalize the distance from the linear space $\mathcal{Z}_N$. Our derivation allows us to interpret $z^\star_\xi$ as the *deduced background*: $z^\star_\xi$ is the component of the state informed by the prior knowledge of the system, and represents anticipated uncertainty in the mathematical model. Similarly, we can interpret

$\eta_\xi^\star$ as the *update*, the component of the state that accommodates unanticipated or non-parametric uncertainty. Consistently, we refer to $\mathcal{Z}_N$ as the *background space*. Finally, we observe that the parameter ξ should be chosen based on the accuracy of the background space — hence on the accuracy of the bk mathematical model — and on the magnitude of the disturbances $\epsilon_1, \ldots, \epsilon_M$. In section 4, we propose an adaptive strategy to select ξ.

2.3. **Well-posedness analysis**

Given the background space $\mathcal{Z}_N \subset \mathcal{U}$, and the observation centers $\mathcal{X}_M = \{x_m\}_{m=1}^M$, we introduce the update space $\mathcal{U}_M$ as

$$\mathcal{U}_M = \text{span}\{K_{x_m}\}_{m=1}^M, \tag{8}$$

and the stability constant $\beta_{N,M}$ as

$$\beta_{N,M} := \inf_{z \in \mathcal{Z}_N} \sup_{q \in \mathcal{U}_M} \frac{(z,q)}{\|z\|\|q\|}. \tag{9}$$

We observe that for perfect measurements (i.e., $y_m = u^{\text{true}}(x_m)$) the inner product (u^{true}, q) is a weighted sum of experimental observations

$$\left(u^{\text{true}}, q = \sum_{m=1}^M \alpha_m K_{x_m}\right) = \sum_{m=1}^M \alpha_m (u^{\text{true}}, K_{x_m}) = \sum_{m=1}^M \alpha_m y_m, \tag{10}$$

for any $q \in \mathcal{U}_M$. For this reason, we say that $\mathcal{U}_M$ is *experimentally observable*. We further observe that the stability constant $\beta_{N,M}$ is a non-increasing function of background span (N), and a non-decreasing function of observable span (M). Furthermore, $\beta_{N,M} = 0$ for $M < N$.

In the remainder of the paper, we assume that the observation centers $x_1, \ldots, x_M \in \Omega$ are distinct. This assumption has two important consequences: first, the update space $\mathcal{U}_M$ is M-dimensional; second, the matrix $\mathbb{K} \in \mathbb{R}^{M,M}$, $\mathbb{K}_{m,m'} = K(x_m, x_{m'})$, is symmetric positive definite. The proof of the first statement is straightforward, while the proof of the second statement is provided in [58, Theorem 10.4]. We extensively use these results in the proofs of this section.

Next Proposition contains the main result of this section.

Proposition 2.1. *Suppose that $\mathcal{Z}_N \subset \mathcal{U}$, and let $\beta_{N,M}$ be defined in (9). Let us further suppose that the observation sites $x_1, \ldots, x_M \in \Omega$ are distinct. Let $\xi > 0$. Then, the following hold.*

(i) *Any solution $(z_\xi^\star, \eta_\xi^\star)$ to (7) belongs to $\mathcal{Z}_N \times \mathcal{Z}_N^\perp \cap \mathcal{U}_M$.*

(ii) *The pair $(z_\xi^\star, \eta_\xi^\star)$ is a solution to (7) if and only if $u_\xi^\star = z_\xi^\star + \eta_\xi^\star$ is a solution to the problem*

$$u_\xi^\star := \arg\inf_{u \in \mathcal{U}} J_\xi^{(1)}(u) := \xi \|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + V_M(u). \tag{11}$$

(iii) *If $\beta_{N,M} > 0$, there exists a unique solution $(z_\xi^\star, \eta_\xi^\star)$ to (7). Furthermore, $(z_\xi^\star, \eta_\xi^\star)$ solves the following problem:*

$$\begin{cases} 2\xi(\eta_\xi^\star, q) + \frac{2}{M} \sum_{m=1}^M \left(z_\xi^\star(x_m) + \eta_\xi^\star(x_m) - y_m\right) q(x_m) & = 0 \quad \forall\, q \in \mathcal{U}_M; \\ (\eta_\xi^\star, p) & = 0 \quad \forall\, p \in \mathcal{Z}_N. \end{cases} \tag{12}$$

In view of the proof of Proposition 2.1, we first present two lemmas. The first lemma is proven in [35, Proposition Appendix A.1].

Lemma 2.2. *Let $\mathcal{U}_M := \text{span}\{K_{x_m}\}_{m=1}^M$ and let $\beta_{N,M}$ be defined as in (9). Then, we have that*

$$\beta_{N,M} = \inf_{\eta \in \mathcal{U}_M^\perp} \frac{\|\Pi_{\mathcal{Z}_N^\perp} \eta\|}{\|\eta\|}. \tag{13}$$

Lemma 2.3. *Let $\mathcal{U}_{M'} := \text{span}\{K_{x_m}\}_{m=1}^{M'}$, $M' \le M$. Let us introduce $\beta_{N,M'} = \inf_{z\in\mathcal{Z}_N} \sup_{v\in\mathcal{U}_{M'}} \frac{(z,v)}{\|z\|\|v\|}$, and the matrix $\mathbb{K}^{(M')} \in \mathbb{R}^{M',M'}$, $\mathbb{K}^{(M')}_{m,m'} = K(x_m, x_{m'})$. Let us further define*

$$c_{N,M} := \max_{M'=1,\ldots,M} \hat{c}_{N,M'}, \qquad \hat{c}_{N,M'} = \min\left(\frac{1}{2}\lambda_{\min}(\mathbb{K}^{(M')}),\ \frac{\lambda_{\min}(\mathbb{K}^{(M')})}{2+\lambda_{\min}(\mathbb{K}^{(M')})}\,\beta^2_{N,M'}\right), \tag{14a}$$

where $\lambda_{\min}(\mathbb{K}^{(M')})$ denotes the minimum eigenvalue of the matrix $\mathbb{K}^{(M')}$.

Then, the following bound holds:

$$\widetilde{J}(u) \;=\; \|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \sum_{m=1}^{M} \left(u(x_m)\right)^2 \ge c_{N,M}\|u\|^2, \qquad \forall\, u \in \mathcal{U}. \tag{14b}$$

Proof. We first claim that for any M' such that $\beta_{N,M'} > 0$ we have

$$\widetilde{J}_{M'}(u) \;=\; \|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \sum_{m=1}^{M'} \left(u(x_m)\right)^2 \ge \hat{c}_{N,M'}\|u\|^2, \qquad \forall\, u \in \mathcal{U}. \tag{15}$$

Given (15), we find that

$$\widetilde{J}(u) \ge \widetilde{J}_{M'}(u) \ge \hat{c}_{N,M'}\|u\|^2 \;\; \forall\, M' \le M \;\Rightarrow\; \widetilde{J}(u) \ge \left(\max_{M'} \hat{c}_{N,M'}\right)\|u\|^2,$$

which is the thesis.

We now show (15). Given $u \in \mathcal{U}$, we introduce $u_1 = \Pi_{\mathcal{U}_{M'}^\perp} u$, $u_2 = \Pi_{\mathcal{U}_{M'}} u = \sum_{m=1}^{M'} (\mathbf{u}_2)_m K_{x_m}$. Then, we observe that

$$u_1(x_m) = (\underbrace{K_{x_m}}_{\in \mathcal{U}_{M'}},\, u_1) = 0, \qquad m = 1, \ldots, M'. \tag{16}$$

We further observe that

$$\sum_{m=1}^{M'} \left(u_2(x_m)\right)^2 = \|\mathbb{K}^{(M')}\mathbf{u}_2\|_2^2, \qquad \|u_2\|^2 = \mathbf{u}_2^T\, \mathbb{K}^{(M')}\, \mathbf{u}_2,$$

which implies that

$$\min_{u_2\in\mathcal{U}_{M'}} \frac{\sum_{m=1}^{M'} \left(u_2(x_m)\right)^2}{\|u_2\|^2} \;=\; \min_{\mathbf{u}_2\in\mathbb{R}^{M'}} \frac{\|\mathbb{K}^{(M')}\mathbf{u}_2\|_2^2}{\mathbf{u}_2^T\, \mathbb{K}^{(M')}\, \mathbf{u}_2} = \lambda_{\min}(\mathbb{K}^{(M')}). \tag{17}$$

Combining (16) and (17), we obtain

$$\sum_{m=1}^{M'} \left(u(x_m)\right)^2 = \sum_{m=1}^{M'} \left(u_2(x_m)\right)^2 \ge\; \lambda_{\min}(\mathbb{K}^{(M')})\, \|u_2\|^2.$$

Now, recalling the identity $2ab \ge -\frac{1}{\epsilon}a^2 - \epsilon b^2$ valid for any $\epsilon > 0$, and Lemma 2.2, we obtain:

$$\begin{aligned}
\widetilde{J}_{M'}(u) = \widetilde{J}_{M'}(u_1+u_2) \ge\;& \|\Pi_{\mathcal{Z}_N^\perp} u_1\|^2 + \|\Pi_{\mathcal{Z}_N^\perp} u_2\|^2 + 2\,(\Pi_{\mathcal{Z}_N^\perp} u_1, \Pi_{\mathcal{Z}_N^\perp} u_2) \;+\; \lambda_{\min}(\mathbb{K}^{(M')})\, \|u_2\|^2 \\
\ge\;& (1-\epsilon)\,\beta^2_{N,M'}\|u_1\|^2 \;+\; \left(1-\tfrac{1}{\epsilon}\right)\|\Pi_{\mathcal{Z}_N^\perp} u_2\|^2 \;+ \lambda_{\min}(\mathbb{K}^{(M')})\|u_2\|^2
\end{aligned}$$

Let us consider $\epsilon \in \left(\frac{1}{1+\lambda_{\min}(\mathbb{K}^{(M')})}, 1\right)$. Recalling that $\|\Pi_{\mathcal{Z}_N^\perp} u_2\| \leq \|u_2\|$, we obtain

$$\begin{aligned}\widetilde{J}_{M'}(u) \geq & \quad (1-\epsilon)\beta_{N,M'}^2 \|u_1\|^2 + \left(\lambda_{\min}(\mathbb{K}^{(M')}) + 1 - \tfrac{1}{\epsilon}\right) \|u_2\|^2 \\ \geq & \quad \min\left(\lambda_{\min}(\mathbb{K}^{(M')}) + 1 - \tfrac{1}{\epsilon},\, (1-\epsilon)\beta_{N,M'}^2\right) \underbrace{\left(\|u_1\|^2 + \|u_2\|^2\right)}_{=\|u\|^2}.\end{aligned}$$

Estimate (15) follows by considering $\epsilon = \frac{2}{2+\lambda_{\min}(\mathbb{K}^{(M')})}$. □

We observe that $c_{N,M}$ is monotonic increasing with M; therefore, it is asymptotically bounded from below in the limit $M \to \infty$. We also recall that $\lambda_{\min}(\mathbb{K}^{(M')}) > 0$ since the sites $x_1, \ldots, x_{M'}$ are distinct. Therefore, $c_{N,M}$ is strictly positive if $\beta_{N,M}$ is strictly positive.

Proof. (Proposition 2.1). We first prove that $\eta_\xi^\star \in \mathcal{U}_M \cap \mathcal{Z}_N^\perp$ (Statement (i)). Thesis follows by observing that $J_\xi(z,\eta) = J_\xi(z, \Pi_{\mathcal{U}_M}\eta) + \xi\|\Pi_{\mathcal{U}_M^\perp}\eta\|^2$, and $J_\xi(z,\eta) = J_\xi(z + \Pi_{\mathcal{Z}_N}\eta, \Pi_{\mathcal{Z}_N^\perp}\eta) + \xi\|\Pi_{\mathcal{Z}_N}\eta\|^2$. We omit the details.

We now show that $(z_\xi^\star, \eta_\xi^\star)$ solves (7) if and only if $u_\xi^\star = z_\xi^\star + \eta_\xi^\star$ solves (11) (Statement (ii)). Exploiting Statement (i), we have

$$\min_{(z,\eta)\in\mathcal{Z}_N\times\mathcal{U}} J_\xi(z,\eta) = \min_{(z,\eta)\in\mathcal{Z}_N\times\mathcal{Z}_N^\perp} J_\xi(z,\eta).$$

Thesis follows by observing that $J_\xi^{(1)}(u) = J_\xi(\Pi_{\mathcal{Z}_N} u, \Pi_{\mathcal{Z}_N^\perp} u)$, and recalling that $\mathcal{U} = \mathcal{Z}_N \oplus \mathcal{Z}_N^\perp$.

We now prove (iii). Applying Lemma 2.3, we find that the objective function $J_\xi^{(1)} : \mathcal{U} \mapsto \mathbb{R}$ is strictly convex if $\beta_{N,M} > 0$. Existence and uniqueness of the solution to (11) then follow from [16, Theorem 3, Chapter 8.2]. Exploiting Statement (ii), we find that the solution $(z_\xi^\star, \eta_\xi^\star)$ to (7) exists and is unique. Furthermore, recalling that the solution $u_\xi^\star$ must be a zero of the first variation of $J_\xi^{(1)}$, we obtain

$$\delta J(u_\xi^\star, v) = 2\xi(\Pi_{\mathcal{Z}_N^\perp} u_\xi^\star, v) + \frac{2}{M}\sum_{m=1}^{M} \left(u_\xi^\star(x_m) - y_m\right) v(x_m) = 0. \qquad \forall\, v \in \mathcal{U},$$

which implies (12). Thesis follows. □

Before concluding, we present a number of observations. First, in Proposition 2.1, we rely on the assumption that $\mathcal{Z}_N \subset \mathcal{U}$. This is required to define $\beta_{N,M}$ in (9), and also the single-field formulation (11). In section 2.4, we derive sufficient conditions for the well-posedness of (7) that do not rely on the hypothesis $\mathcal{Z}_N \subset \mathcal{U}$. Second, statement (i) of Proposition 2.1 is extremely important from a practical standpoint since it provides an *a priori* finite-dimensional representation formula for the solution to (7): we rely on this finite-dimensional representation to derive an efficient algebraic counterpart of the variational statement.

2.4. **Algebraic formulation**

In this section, we present the PBDW algebraic formulation, and we study the stability properties of the linear system. Then, as anticipated in section 2.3, we present a well-posedness result that does not rely on the assumption that $\mathcal{Z}_N \subset \mathcal{U}$.

2.4.1. *PBDW algebraic statement*

In view of the algebraic formulation, we first introduce the matrices $\mathbb{K} \in \mathbb{R}^{M,M}$, $\mathbb{Z} \in \mathbb{R}^{N,N}$, $\mathbb{L} \in \mathbb{R}^{M,N}$ such that

$$\mathbb{K}_{m,m'} = K(x_m, x_{m'}), \quad \mathbb{Z}_{n,n'} = (\zeta_n, \zeta_{n'}), \quad \mathbb{L}_{m,n} = \zeta_n(x_m), \tag{18}$$

for $m, m' = 1, \ldots, M$, and $n, n' = 1, \ldots, N$.

Next Proposition shows the algebraic counterpart of the APBDW statement (7). We note that the computation of the solution to (7) does not require the computation of the matrix $\mathbb{Z}$.

Proposition 2.4. *Let $\beta_{N,M} > 0$, and let $\xi > 0$. Then, the solution to (7) $u_\xi^\star = z_\xi^\star + \eta_\xi^\star$ is given by*

$$u_\xi^\star(\cdot) = \sum_{n=1}^{N} z_{\xi,n}^\star \zeta_n(\cdot) + \sum_{m=1}^{M} \eta_{\xi,m}^\star K_{x_m}(\cdot), \tag{19a}$$

where the pair $(\boldsymbol{z}_\xi^\star, \boldsymbol{\eta}_\xi^\star) \in \mathbb{R}^N \times \mathbb{R}^M$ solves

$$\begin{bmatrix} \xi M \mathbb{I} + \mathbb{K} & \mathbb{L} \\ \mathbb{L}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_\xi^\star \\ \boldsymbol{z}_\xi^\star \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_M \\ \mathbf{0} \end{bmatrix}, \quad \boldsymbol{y}_M = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} \tag{19b}$$

Proof. Recalling Proposition 2.1 (Statement (i)), we have that $u_\xi^\star$ is of the form (19a). Then, substituting (19a) in (12) and choosing $q = K_{x_m}$, $p = \zeta_n$, we find

$$\begin{bmatrix} 2\xi\mathbb{K} + \frac{2}{M}\mathbb{K}^2 & \frac{2}{M}\mathbb{K}\mathbb{L} \\ \mathbb{L}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_\xi^\star \\ \boldsymbol{z}_\xi^\star \end{bmatrix} = \begin{bmatrix} \frac{2}{M}\mathbb{K}\boldsymbol{y}_M \\ \mathbf{0} \end{bmatrix}.$$

Since $\mathbb{K}$ is invertible, thesis follows by multiplying the first equation by $\frac{M}{2}\mathbb{K}^{-1}$. □

We now wish to investigate the spectral properties of the linear system (19b). With this in mind, we first present a standard result (see, e.g., [35, Lemma 3.3]).

Lemma 2.5. *The inf-sup constant $\beta_{N,M}$ is the square root of the minimum eigenvalue of the following generalized eigenproblem:*

$$\mathbb{L}^T \mathbb{K}^{-1} \mathbb{L}\, \mathbf{z}_n = \nu_n \mathbb{Z}\, \mathbf{z}_n, \qquad n = 1, \dots, N. \tag{20}$$

Next Proposition provides a bound for the minimum eigenvalue of the saddle point system (19b).

Proposition 2.6. *Suppose that $\beta_{N,M} > 0$, and let $\zeta_1, \dots, \zeta_N$ be an orthonormal basis of $\mathcal{Z}_N$. Let $\lambda_\xi^{\min}$ be the minimum (in absolute value) eigenvalue of the saddle point system (19b). Then, the following bound holds:*

$$|\lambda_\xi^{\min}| \geq \min\left(\lambda_{\min}(\mathbb{K}) + \xi M,\ \ \beta_{N,M}^2 - \xi M \frac{\lambda_{\max}(\mathbb{L}^T\mathbb{L})}{\lambda_{\min}(\mathbb{K})(\xi M + \lambda_{\min}(\mathbb{K}))}\right), \tag{21}$$

and the bound holds with equality for $\xi = 0$.

Proof. Following the argument in [5, Section 3.4], we observe that the saddle-point system (19b) is congruent to the block-diagonal matrix $\begin{bmatrix} \mathbb{K} + \xi M \mathbb{I} & 0 \\ 0 & -\mathbb{L}^T(\mathbb{K} + \xi M \mathbb{I})^{-1}\mathbb{L} \end{bmatrix}$. Therefore, we find:

$$|\lambda_\xi^{\min}| = \min\left(\lambda_{\min}(\mathbb{K}) + \xi M,\ \ \lambda_{\min}\left(\mathbb{L}^T(\mathbb{K} + \xi M \mathbb{I})^{-1}\mathbb{L}\right)\right).$$

We now estimate $\lambda_{\min}\left(\mathbb{L}^T(\mathbb{K} + \xi M \mathbb{I})^{-1}\mathbb{L}\right)$. Towards this end, we first observe that

$$(\mathbb{K} + \xi M \mathbb{I})^{-1} = \mathbb{K}^{-1} - \xi M \mathbb{X}_\xi, \quad \mathbb{X}_\xi = (\mathbb{K} + \xi M \mathbb{I})^{-1}\mathbb{K}^{-1}.$$

Therefore, recalling standard linear algebra results and Lemma 2.5, we find

$$\begin{aligned}\lambda_{\min}\left(\mathbb{L}^T(\mathbb{K}+\xi M\mathbb{I})^{-1}\mathbb{L}\right) = &\min_{\mathbf{v}} \frac{\mathbf{v}^T\mathbb{L}^T(\mathbb{K}+\xi M\mathbb{I})^{-1}\mathbb{L}\mathbf{v}}{\|\mathbf{v}\|_2^2}\\ \geq &\min_{\mathbf{v}} \frac{\mathbf{v}^T\mathbb{L}^T\mathbb{K}^{-1}\mathbb{L}\mathbf{v}}{\|\mathbf{v}\|_2^2} - \xi M \max_{\mathbf{v}} \frac{\mathbf{v}^T\mathbb{L}^T\mathbb{X}_\xi\mathbb{L}\mathbf{v}}{\|\mathbf{v}\|_2^2}\\ \geq &\ \beta_{N,M}^2 - \xi M\lambda_{\max}(\mathbb{X}_\xi)\lambda_{\max}(\mathbb{L}^T\mathbb{L}).\end{aligned}$$

Thesis follows by observing that

$$\lambda_{\max}(\mathbb{X}_\xi) = \max_{\mathbf{v}} \frac{\mathbf{v}^T(\mathbb{K}+\xi M\mathbb{I})^{-1}\,\mathbb{K}^{-1}\mathbf{v}}{\|\mathbf{v}\|_2^2} \leq (\lambda_{\min}(\mathbb{K})+\xi M)^{-1}\ \frac{1}{\lambda_{\min}(\mathbb{K})}.$$

□

2.4.2. *An improved well-posedness result*

Proposition 2.7 shows a well-posedness result that does not rely on the assumption $\mathcal{Z}_N \subset \mathcal{U}$.

Proposition 2.7. *Let $\mathcal{Z}_N \subset C(\Omega)$. Then, the solution $(z_\xi^\star, \eta_\xi^\star) \in \mathcal{Z}_N \times \mathcal{U}$ to (7) exists and is unique if and only if the matrix $\mathbb{L}$ has rank N. Furthermore, the state estimate $u_\xi^\star = z_\xi^\star + \eta_\xi^\star$ satisfies the representation formula (19).*

Proof. We observe that $J_\xi(z,\eta) = J_\xi(z, \Pi_{\mathcal{U}_M}\eta) + \xi\|\Pi_{\mathcal{U}_M^\perp}\eta\|^2$. Therefore, any solution $\eta_\xi^\star$ to (7) belongs to $\mathcal{U}_M$. This implies that any solution to (7) is of the form (19a).

Substituting (19a) in the minimization statement, we find

$$\min_{(\mathbf{z}^\star,\boldsymbol{\eta}^\star)\in\mathbb{R}^N\times\mathbb{R}^M} \xi\,\boldsymbol{\eta}^T\,\mathbb{K}\,\boldsymbol{\eta} + \frac{1}{M}\,\|\mathbb{K}\boldsymbol{\eta} + \mathbb{L}\boldsymbol{z} - \boldsymbol{\ell}_M^{\text{obs}}\|_2^2.$$

By deriving the stationary conditions, we find

$$\begin{cases} \left(\xi\mathbb{K} + \frac{1}{M}\mathbb{K}^2\right)\,\boldsymbol{\eta}^\star \,+\, \frac{1}{M}\mathbb{K}\mathbb{L}\mathbf{z}^\star &= \frac{1}{M}\mathbb{K}\mathbf{y} \\ \mathbb{L}^T\,\mathbb{K}\,\boldsymbol{\eta}^\star + \mathbb{L}^T\mathbb{L}\mathbf{z}^\star &= \mathbb{L}^T\mathbf{y} \end{cases} \tag{22}$$

By premultiplying $(22)_1$ by $M\mathbb{K}^{-1}$, we find

$$(\xi M\mathbb{I} + \mathbb{K})\,\boldsymbol{\eta}^\star + \mathbb{L}\mathbf{z}^\star = \mathbf{y}. \tag{23a}$$

If we now premultiply the latter equation by $\mathbb{L}^T$ and we subtract $(22)_2$, we obtain

$$\mathbb{L}^T\boldsymbol{\eta} = \mathbf{0}. \tag{23b}$$

Saddle system (23a) - (23b) is well-posed since $\mathbb{K}$ is invertible and $\mathbb{L}$ is full-rank by hypothesis. Thesis follows. □

We observe that if $\mathcal{Z}_N \subset \mathcal{U}$, the condition $\text{rank}(\mathbb{L}) = N$ is equivalent to $\beta_{N,M} > 0$ (cf. Lemma 2.5). Therefore, Proposition 2.7 is equivalent to Proposition 2.1 if $\mathcal{Z}_N \subset \mathcal{U}$.

2.5. **Construction of the spaces**

2.5.1. *Construction of the background space $\mathcal{Z}_N$*

If we denote by $\epsilon_{\text{mod}}^{\text{bk}} = \inf_{\mu\in\mathcal{P}^{\text{bk}}} \|u^{\text{true}} - u^{\text{bk}}(\mu)\|$ the *modeling error*, and by $\epsilon_N^{\text{bk}} = \inf_{z\in\mathcal{Z}_N} \|u^{\text{true}} - z\|$ the *best-fit error* associated with $\mathcal{Z}_N$, we aim to choose $\mathcal{Z}_N$ such that $\epsilon_{\text{mod}}^{\text{bk}} \approx \epsilon_N^{\text{bk}}$. Since the stability of the APBDW

formulation strongly depends on the value of N, we further wish to keep N small compared to the number of observations M. We observe that we may bound the best-fit error as follows:

$$\epsilon_N^{\text{bk}} = \inf_{z\in\mathcal{Z}_N} \|u^{\text{true}} - z\| \leq \sup_{\mu\in\mathcal{P}^{\text{bk}}} \inf_{z\in\mathcal{Z}_N} \|u^{\text{bk}}(\mu) - z\| + \inf_{\mu\in\mathcal{P}^{\text{bk}}} \|u^{\text{bk}}(\mu) - u^{\text{true}}\| = \epsilon_{\text{disc},N}^{\text{bk}} + \epsilon_{\text{mod}}^{\text{bk}}, \tag{24}$$

where $\epsilon_{\text{disc},N}^{\text{bk}} = \sup_{\mu\in\mathcal{P}^{\text{bk}}} \inf_{z\in\mathcal{Z}_N} \|u^{\text{bk}}(\mu) - z\|$ is the *discretization error*. Therefore, if $\epsilon_{\text{mod}}^{\text{bk}}$ is small, we can practically construct $\mathcal{Z}_N$ to minimise $\epsilon_{\text{disc},N}^{\text{bk}}$.

If the bk model is defined over the domain of interest, we observe that the task of constructing the space $\mathcal{Z}_N$ is equivalent in objective to the task of constructing the reduced trial space in parametric Model Order Reduction (pMOR). Therefore, we can resort to state-of-the-art techniques proposed in the pMOR literature to generate $\mathcal{Z}_N$, such as Proper Orthogonal Decomposition (POD, [6, 24, 29]), Proper Generalized Decomposition (PGD, [11, 12]), Taylor expansions ([18]), and Greedy algorithms.

In this work, we rely on the Weak-Greedy algorithm for the construction of the space $\mathcal{Z}_N$. The algorithm was first proposed in [43, 44] in the context of Reduced Basis method ([22, 45, 47]), and has been applied to elliptic and parabolic, linear and nonlinear, differential equations. The convergence with respect to N of the reduced space obtained using this Greedy procedure has been extensively studied in [7, 8, 15] and linked to the so-called Kolmogorov N-width [41]. We refer to [45, Chapter 7] for a thorough overview of the computational procedure; we further refer to [13, Section 8] for a review of the theoretical results.

We briefly address the more general case in which the bk model is defined in a super-domain Ω^{bk} that contains Ω. In this case, we might first appeal to one of the techniques presented above to build a space $\widehat{\mathcal{Z}}_N$ for the manifold $\widehat{\mathcal{M}}^{\text{bk}} = \{u^{\text{bk}}(\mu) : \Omega^{\text{bk}} \to \mathbb{R} : \mu \in \mathcal{P}^{\text{bk}}\}$. Then, we might define $\mathcal{Z}_N := \{z|_\Omega : z \in \widehat{\mathcal{Z}}_N\}$. If the manifold $\widehat{\mathcal{M}}^{\text{bk}}$ is low-dimensional and reducible[1], this approach should guarantee accurate reduced spaces for the bk manifold $\mathcal{M}^{\text{bk}}$. However, if $\widehat{\mathcal{M}}^{\text{bk}}$ is not reducible, and Ω is strictly contained in Ω^{bk}, this approach might either be unfeasible or lead to poor approximation spaces. This issue is subject of ongoing research.

2.5.2. *Construction of the update space: Reproducing Kernel Hilbert Spaces*

In [36, 37], authors first choose an inner product $(\cdot,\cdot)$ for the space $\mathcal{U}$, and then appeal to a Finite Element (FE) discretization to compute the Riesz representations of the observation functionals, $K_{x_1}, \ldots, K_{x_M}$. This requires the solution to M FE problems. For pointwise measurements, we can exploit the theory of Reproducing Kernel Hilbert Spaces (RKHS, [1]) to avoid the implicit construction of the update space.

An Hilbert space $(\mathcal{U}, \|\cdot\|)$ is a RKHS if the point evaluation functionals are continuous, i.e. $\delta_x \in \mathcal{U}'$ for all $x \in \Omega$. This is equivalent (cf. [58, Theorem 10.2]) to assume that there exists a function $K : \Omega \times \Omega \to \mathbb{R}$ such that (i) $K(\cdot, x) \in \mathcal{U}$ for all $x \in \Omega$, and (ii) $(K(\cdot, x), f) = f(x)$ for all $x \in \Omega$ and $f \in \mathcal{U}$. The function K is called Reproducing Kernel. For any $x \in \Omega$, we observe that K_x is the Riesz element associated with the point evaluation functional δ_x.

A function $K : \Omega \times \Omega \to \mathbb{R}$ is a symmetric positive definite (SPD) kernel if (i) $K(x, y) = K(y, x)$ for all $x, y \in \Omega$, and (ii) for any set of N distinct points in Ω, $\{x_n\}_{n=1}^N \subset \Omega$, the matrix $\mathbb{K} \in \mathbb{R}^{N,N}$ defined as $\mathbb{K}_{n,n'} = K(x_n, x_{n'})$ is positive definite. We already mentioned that if $\mathcal{U}$ is a RKHS such that point evaluation functionals are linearly independent then the corresponding reproducing kernel is SPD. The converse is also true: given the SPD kernel K there exists a RKHS for which K is the reproducing kernel, which is referred to as *native space* of K. The latter result is known as Moore-Aronszajn theorem and was first proved in [1].

The duality between RKHS and SPD kernels has important implications for our discussion. We might first propose an explicit SPD kernel — for instance, the csRBF kernel (25) — and then appeal to Moore-Aronszajn theorem to recover the variational formulation. This prevents us from having to solve M FE problems to build the update space $\mathcal{U}_M$.

An important class of kernels, which is employed in the numerical simulations, is given by the compactly supported radial basis functions of minimal degree (csRBFs), also known as Wendland functions. This class of

[1] We refer to [45, Chapter 5] for a formal discussion about the reducibility of parametric manifolds.

kernels was first proposed by Wendland in [57], and is defined as $K_\chi(x,y) = \phi_{d,k}(\gamma\|x-y\|_2)$ where $\chi = [k,\gamma]$ and

$$\phi_{d,k}(r) = \begin{cases} p_{d,k}(r) & 0 \le r \le 1; \\ 0 & r > 1. \end{cases} \tag{25a}$$

The polynomial $p_{d,k}$ has the following form for $k = 0, 1$ and for all d:

$$p_{d,k}(r) = \begin{cases} (1-r)^{\ell_{d,k}} & k = 0 \\ (1-r)^{\ell_{d,k}+1}\left((\ell_{d,k}+1)r+1\right) & k = 1 \end{cases} \tag{25b}$$

and $\ell_{d,k} = \left\lfloor \frac{d}{2} \right\rfloor + k + 1$. We observe that it is possible to generalize (25b) to the more general case $k \in \mathbb{N}$; we refer to [58, Table 9.1] for the explicit formulas.

Next result clarifies the connection between csRBF and Sobolev spaces. We refer to [58, Theorem 10.35] for the proof.

Theorem 2.8. *Let us consider the compactly supported RBF K_χ, $K_\chi(x,y) = \phi_{d,k}(\gamma\|x-y\|_2)$, introduced in (25). Let $\Omega = \mathbb{R}^d$, and let either one of these conditions hold: (i) $d \ge 3$, $k \ge 0$, or (ii) $d \ge 1$, $k > 0$. Then, the native space for K_χ is the Sobolev space $H^{(d+1)/2+k}(\mathbb{R}^d)$.*

2.6. Computational procedure

Algorithm 1 summarises the computational procedure. As in [36, 37], we exploit an Offline-Online computational decomposition. During the offline stage, which is performed before acquiring the experimental measurements, we generate the background space $\mathcal{Z}_N$, we select the observation centers $x_1, \dots, x_M$ and we assemble the matrix $\mathbb{L}$ in (18). During the online stage, we first acquire data $y_1, \dots, y_M$, we choose the parameters of the Kernel K and the constant ξ, and then we solve the linear system (19b). The offline stage is computationally intensive since it requires the solution to the bk model for multiple values of the parameter; on the other hand, the online stage is independent of the dimensionality of the FE mesh used to compute $\mathcal{Z}_N$, and is thus computationally inexpensive for moderate values of M. We further observe that we have not discussed yet how to select the observation centers during the offline stage, and how to select kernel parameters and the constant ξ: we address these points in section 4.

Algorithm 1 APBDW approach. Offline-online computational procedure

Offline stage

1: Choose a family of kernels (e.g. (25))
2: Generate the background $\mathcal{Z}_N \subset \mathcal{U}$ (cf. section 2.5.1)
3: Select the observation centers $x_1, \dots, x_M \in \Omega$ (cf. section 4)
4: Compute the matrix $\mathbb{L}$ (18)

Online stage

1: Acquire the measurements $y_1, \dots, y_M$
2: Choose the parameters of the kernel and the regularizer weight ξ (cf. section 4)
3: Assemble the matrix (19b) and solve the linear system (19b)
4: (If needed) Evaluate the state using (19a).

2.7. Connection with other formulations

2.7.1. *Connection with other data assimilation procedures*

In the statistical learning literature, APBDW is closely related to the approach presented and analyzed in [26] by Kimeldorf and Wahba. In more detail, the two approaches are equivalent if we choose $\mathcal{Z}_N$ as the set of all

polynomials of degree less or equal to κ, $\kappa > d/2 - 1$, and $\mathcal{U} = H^{\kappa+1}(\Omega)$ endowed with a proper inner product. We also observe that, exploiting the connection with [26], we can re-interpret our formulation in a Bayesian setting as a Gaussian linear system with improper prior (see [55]). In this work, we do not pursue this feature of the approach.

Furthermore, for $N = M$ APBDW formulation is equivalent to the solution to the Generalized Empirical Interpolation Method (GEIM, [34]), while for $\xi \to \infty$ and $\mathcal{Z}_N$ built using a Proper Orthogonal Decomposition (POD, [29]), APBDW is asymptotically equivalent to Gappy-POD ([17, 59]). On the other hand, in the limit $\xi \to 0^+$ our formulation (7) is asymptotically equivalent to the "noise-free" formulation presented in [36]. Next Proposition shows the two asymptotic results.

Proposition 2.9. *Let $\beta_{N,M} > 0$. Let $u^\star_\xi = \eta^\star_\xi + z^\star_\xi$ be the solution to (7). Then, we have*

$$\lim_{\xi \to 0^+} \|u^\star - u^\star_\xi\| = 0, \tag{26}$$

and

$$\lim_{\xi \to \infty} \|u^\star_\xi - z_{\rm LS}\| = 0, \tag{27}$$

where $u^\star = z^\star + \eta^\star$ satisfies

$$(\eta^\star,\, z^\star) := \arg \min_{(z,\eta) \in \mathcal{Z}_N \times \mathcal{U}} \|\eta\| \qquad \textit{subject to } \; z(x_m) + \eta(x_m) = y_m, \;\; m = 1, \dots, M; \tag{28}$$

and $z^\star_{LS} \in \mathcal{Z}_N$ satisfies

$$z^\star_{LS} := \arg \min_{z \in \mathcal{Z}_N} V_M(z). \tag{29}$$

Proof. Let us first consider the limit $\xi \to 0^+$. Recalling [36, Section 2.4], we have that $u^\star$ is of the form (19a) with coefficients $\boldsymbol{\eta}^\star_0, \boldsymbol{z}^\star_0$ satisfying

$$\begin{bmatrix} \mathbb{K} & \mathbb{L} \\ \mathbb{L}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}^\star_0 \\ \mathbf{z}^\star_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_M \\ \mathbf{0} \end{bmatrix},$$

Therefore, we just have to show that

$$\lim_{\xi \to 0^+} \left(\boldsymbol{u}^\star_\xi = \begin{bmatrix} \boldsymbol{\eta}^\star_\xi \\ \boldsymbol{z}^\star_\xi \end{bmatrix} \right) = \boldsymbol{u}^\star_{\xi=0} = \begin{bmatrix} \boldsymbol{\eta}^\star_{\xi=0} \\ \boldsymbol{z}^\star_{\xi=0} \end{bmatrix}$$

Exploiting Proposition 2.6, we can show that each eigenvalue of the saddle-point system (19b) satisfies $|\lambda_{\xi,j}| \geq \frac{1}{2} \min(\lambda_{\min}(\mathbb{K}), \beta^2_{N,M}) > 0$ for all $\xi \leq \bar{\xi}$, for some $\bar{\xi} > 0$. Therefore, we find $\|\boldsymbol{u}^\star_\xi\|_2 \leq C$ for all $\xi \leq \bar{\xi}$.

Let $\{\xi_j\}_j$ be a positive sequence such that $\xi_j \to 0^+$, and let $\boldsymbol{u}^\star_{\xi_j}$ be the solution to (19b) for $\xi = \xi_j$. Since $\{\boldsymbol{u}^\star_{\xi_j}\}_j$ is uniformly bounded, applying Bolzano-Weierstrass theorem, we obtain that, up to a subsequence, $\boldsymbol{u}^\star_{\xi_\ell} \to \widehat{\boldsymbol{u}}$. We further observe that

$$\begin{bmatrix} \boldsymbol{y}_M \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbb{K} & \mathbb{L} \\ \mathbb{L}^T & 0 \end{bmatrix} \boldsymbol{u}^\star_{\xi_\ell} + \xi_\ell M \begin{bmatrix} \mathbb{I} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{u}^\star_{\xi_\ell} \underbrace{\to}_{\ell \to \infty} \begin{bmatrix} \boldsymbol{y}_M \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbb{K} & \mathbb{L} \\ \mathbb{L}^T & 0 \end{bmatrix} \widehat{\boldsymbol{u}}.$$

Since the linear system (19b) for $\xi = 0$ admits a unique solution, we must have $\widehat{\boldsymbol{u}} = \boldsymbol{u}^\star_{\xi=0}$. Using the same reasoning, we find that $\boldsymbol{u}^\star_{\xi=0}$ is the only limit point of the uniformly bounded sequence $\{\boldsymbol{u}^\star_{\xi_j}\}_j$. Therefore, the entire sequence is convergent (see, e.g., [38, page 67]).

We now consider the case $\xi \to \infty$. As for the previous case, we must prove that for $\xi \to \infty$

$$\mathbf{z}^\star_\xi \to \mathbf{z}_{\rm LS} = (\mathbb{L}^T \mathbb{L})^{-1} \mathbb{L}^T \boldsymbol{y}_M, \quad \boldsymbol{\eta}^\star_\xi \to \mathbf{0}.$$

The proof exploits the same argument of the previous case, $\xi \to 0^+$. We omit the details. □

2.7.2. *A two-stage regression procedure: connection with Kalman filter*

We can rewrite the linear system (19b) as follows:

$$\begin{cases} \mathbb{L}^T(\xi M\mathbb{I}+\mathbb{K})^{-1}\mathbb{L}\,\boldsymbol{z}_\xi^\star = \mathbb{L}^T(\xi M\mathbb{I}+\mathbb{K})^{-1}\boldsymbol{y}_M \\ (\xi M\mathbb{I}+\mathbb{K})\boldsymbol{\eta}_\xi^\star = \boldsymbol{y}_M - \mathbb{L}\boldsymbol{z}_\xi^\star. \end{cases} \tag{30}$$

Formulation (30) is the algebraic counterpart of the following two-stage procedure:

$$\begin{cases} z_\xi^\star := \arg\min\limits_{z\in\mathcal{Z}_N} \|\mathcal{L}_M(z)-\boldsymbol{y}_M\|_{\mathbb{W}_\xi}; \\ \eta_\xi^\star := \arg\min\limits_{\eta\in\mathcal{U}} \xi\|\eta\|^2 + \dfrac{1}{M}\displaystyle\sum_{m=1}^{M}(\eta(x_m)-y_m^{\mathrm{err}})^2, \quad y_m^{\mathrm{err}} := y_m - z_\xi^\star(x_m), \end{cases} \tag{31}$$

where $\mathbb{W}_\xi = (\mathbb{K}+\xi M\mathbb{I})^{-1}$, $\mathcal{L}_M(z) = [z(x_1),\dots,z(x_M)]$, and $\|\mathbf{d}\|_{\mathbb{W}} = \sqrt{\mathbf{d}^T\mathbb{W}\mathbf{d}}$. Problem $(31)_1$ corresponds to a weighted least-square regression problem, while $(31)_2$ corresponds to a smoothing problem applied to the error field $u^{\mathrm{true}} - z_\xi^\star$.

Equation (31) can be exploited to interpret the APBDW components $z_\xi^\star$ and $\eta_\xi^\star$: the deduced-background estimate $z_\xi^\star \in \mathcal{Z}_N$ represents our *predicted state estimate* based on prior knowledge (here encoded in the background space); on the other hand, the update $\eta_\xi^\star \in \mathcal{U}_M$ represents the *innovation* induced by the measurements and only depends on the residuals $y_m^{\mathrm{err}} = y_m - z_\xi^\star(x_m)$. This interpretation of the APBDW components clarifies the connection between our approach and least-square statistical linear estimation (LSSLE, [52]), and thus with Kalman filtering[2] ([25,30]). We remark, however, that, while in LSSLE the predicted state estimate is uniquely determined by prior knowledge and possibly by observations at previous times, in our approach both $z_\xi^\star$ and $\eta_\xi^\star$ are determined through experimental data.

We further observe that — from the perspective of approximation theory — APBDW formulation introduces a hierarchy between the approximation provided by the background space $\mathcal{Z}_N$, and the approximation provided by $\mathcal{U}_M$. In more detail, the background space $\mathcal{Z}_N$ should provide primary approximation, while the update space $\mathcal{U}_M$ is designed to complete any deficiency in $\mathcal{Z}_N$. This asymmetry between $\mathcal{Z}_N$ and $\mathcal{U}_M$ is motivated by the underlying assumption that elements of $\mathcal{Z}_N$ have better *generalization properties* (see, e.g., [53]) than elements in $\mathcal{U}_M$, since $\mathcal{Z}_N$ is directly informed by the mathematical model. In this respect, by adapting the parameter ξ, we can properly tune the relative importance of primary and secondary approximation.

3. Error analysis

We present *a priori* and *a posteriori* estimates for the $L^2(\Omega)$ state-estimation error $\|u^{\mathrm{true}} - u_\xi^\star\|_{L^2(\Omega)}$. The importance of the error analysis is twofold. First, it motivates our formulation from a theoretical viewpoint. Second, it provides insights about the role of the different pieces of our formulation: the regularization parameter ξ, the background space $\mathcal{Z}_N$, the kernel K and the centers $\mathcal{X}_M$.

3.1. *A priori* error analysis

In order to derive error bounds for the $L^2(\Omega)$ state-estimation error $\|u^{\mathrm{true}} - u_\xi^\star\|_{L^2(\Omega)}$, we must first introduce assumptions on our dataset $\mathcal{D}_M$. To our knowledge, three different scenarios have been considered so far.

(1) *Random-design regression*: the pairs $\{(x_m, y_m)\}_{m=1}^M$ are drawn independently from a joint unknown distribution $\rho_{(X,Y)}$. In this case, the objective of learning is to estimate the conditional expectation $\mathbb{E}[Y|X=x]$.

[2]We recall that Kalman filter can be interpreted as a sequential form of statistical LSSLE (see, e.g. [52, section 4]).

(2) *Fixed-design regression*: the centers $\mathcal{X}_M = \{x_1, \dots, x_M\}$ are fixed (non-random) points in Ω, while the responses $\mathcal{Y}_M = \{y_m\}_{m=1}^M$ satisfy $y_m = u^{\text{true}}(x_m) + \epsilon_m$, where $u^{\text{true}} : \Omega \to \mathbb{R}$ is the deterministic field of interest and $\epsilon_1, \dots, \epsilon_M$ are independent identically distributed (i.i.d.) random variables with zero mean and variance σ^2, $\epsilon_m \sim (0, \sigma^2)$.

(3) *Scattered data approximation*: both centers $\mathcal{X}_M$ and responses $\mathcal{Y}_M$ are non-random, and we assume that there exists some unknown $\delta > 0$ such that $|y_m - u^{\text{true}}(x_m)| \le \delta$ for all $m = 1, \dots, M$.

The first scenario has been extensively studied in the statistical learning literature (see, e.g., [42, 53]). We refer to [20] for a complete review of the error bounds available. The second scenario has also been studied in statistics; we refer to the survey [19] for further details about a specific class of kernels. Finally, the third scenario has been studied in approximation theory and radial basis functions (see, e.g., [58]). From the modeling perspective, the first scenario refers to the case in which we do not have control on the observation centers, the second scenario addresses the problem of random error in the measurements, and the third scenario addresses the problem of systematic deterministic error.

In the next two sections, we present error bounds for both the second and the third scenarios. Our analysis for the third scenario exploits results first proved in a work of Krebs, Louis and Wendland ([28]). We state upfront that in the remainder of this section we assume that $\mathcal{Z}_N \subset \mathcal{U}$.

3.1.1. *An* a priori *error bound for scattered data approximation*

We state the main result of this section.

Proposition 3.1. *Let Ω be a Lipschitz domain and let $\mathcal{U}$ be the Sobolev space $H^\tau(\Omega)$ with $\tau > d/2$. Let $\beta_{N,M}$ in (9) be strictly positive. Let us further assume that measurements are of the form $y_m = u^{\text{true}}(x_m) + \epsilon_m$ with $|\epsilon_m| \le \delta$ for $m = 1, \dots, M$.*

Then, if $u^{\text{true}} \in \mathcal{U}$, the following holds:

$$\|u^{\text{true}} - u_\xi^\star\|_{L^2(\Omega)}^2 \le C_{N,\mathcal{X}_M} \left(h_{\mathcal{X}_M}^{2\tau} \left(2\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\| + \frac{\delta}{2}\frac{1}{\sqrt{\xi}} \right)^2 + h_{\mathcal{X}_M}^d M \left(\delta + \frac{\sqrt{\xi}}{2}\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\| \right)^2 \right), \tag{32a}$$

where $C_{N,\mathcal{X}_M}$ is defined as

$$C_{N,\mathcal{X}_M} := \sup_{u \in \mathcal{U}} \frac{\|u\|_{L^2(\Omega)}^2}{h_{\mathcal{X}_M}^{2\tau}\|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + h_{\mathcal{X}_M}^d \|u\|_{\ell^2(\mathcal{X}_M)}^2}, \tag{32b}$$

$\|u\|_{\ell^2(\mathcal{X}_M)} := \sqrt{\sum_{m=1}^M u(x_m)^2}$*, and the fill distance $h_{\mathcal{X}_M}$ as*

$$h_{\mathcal{X}_M} = \sup_{x \in \Omega} \min_{m=1,\dots,M} \|x - x_m\|_2. \tag{32c}$$

Proof of Proposition 3.1 is technical and for this reason we present it in Appendix A. In the remainder of this section, we state a number of remarks.

Remark 3.2. We observe that the constant $C_{N,\mathcal{X}_M}$ is associated with the maximum eigenvalue associated to a generalized eigenproblem. Provided that the inf-sup constant $\beta_{N,M} > 0$ and $h_{\mathcal{X}_M} < 1$, the constant $C_{N,\mathcal{X}_M}$ defined in (32b) can be estimated as follows:

$$C_{N,\mathcal{X}_M} \le \frac{1}{\min\{c_{N,M}, 1 - h_{\mathcal{X}_M}^{2\tau-d}\}}\, C, \tag{33}$$

where $c_{N,M}$ is defined in (14). We rigorously prove (33) in Appendix A. We remark that to practically estimate $C_{N,\mathcal{X}_M}$, we need to numerically approximate the maximum eigenvalue of the generalized eigenproblem associated with $C_{N,\mathcal{X}_M}$. Recalling the definition of $c_{N,M}$ (14), we find that $C_{N,\mathcal{X}_M}$ is asymptotically bounded as $M \to \infty$ for fixed N; on the other hand, the dependence on N heavily depends on the background $\mathcal{Z}_N$. Accurate estimates

of $C_{N,\mathcal{X}_M}$ require the estimate of the matrix $\mathbb{Z}$ in (18); the latter requires that the basis $\zeta_1, \ldots, \zeta_N$ can be represented as $\zeta_n(\cdot) = \sum_{q=1}^{Q_n} a_{n,q} K_{\tilde{x}_{n,q}}(\cdot)$ for some $\{a_{n,q}\}_q$ and $\{\tilde{x}_{n,k}\}_q \subset \Omega$, $n = 1, \ldots, N$.

Remark 3.3. For quasi-uniform grids, $h_{\mathcal{X}_M} \sim M^{-1/d}$, for $M \to \infty$, the right-hand side reduces to

$$\|u^{\text{true}} - u_\xi^\star\|^2_{L^2(\Omega)} \lesssim \mathcal{O}\left(\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|^2 \, h^{2\tau}_{\mathcal{X}_M} \left(1 + \frac{1}{\Lambda}\right)^2 + \delta^2 \, (1+\Lambda)^2\right) \tag{34a}$$

where

$$\Lambda = \frac{\sqrt{\xi}\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|}{\delta}, \tag{34b}$$

By minimizing with respect to Λ, we obtain that the asymptotically optimal choice of ξ is

$$\xi = \left(\frac{\delta}{\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|}\right)^{2/3} h^{4/3\tau}_{\mathcal{X}_M}. \tag{35a}$$

The optimal value of ξ is directly proportional to δ, inversely proportional to the background best-fit error $\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|$ and decreases as M increases. For this choice of the hyper-parameter, we obtain:

$$\|u^{\text{true}} - u_\xi^\star\|^2_{L^2(\Omega)} \lesssim \mathcal{O}\left(\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|^{2/3} \, h^{2/3\tau}_{\mathcal{X}_M} \, \delta^{4/3} + \delta^2\right) \qquad M \to \infty. \tag{35b}$$

We observe that for any finite $\delta > 0$, we do not expect convergence in a L^2 sense.

Remark 3.4. In the case of perfect measurements, estimate (32) reduces to

$$\|u^{\text{true}} - u_\xi^\star\|^2_{L^2(\Omega)} \leq \frac{1}{4} C_{N,\mathcal{X}_M} \left(16 h^{2\tau}_{\mathcal{X}_M} + h^d_{\mathcal{X}_M} M\xi\right) \|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|^2. \tag{36}$$

We can decouple the right-hand side of (36), as the product of two terms: (i) $C_{N,\mathcal{X}_M}\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|^2$, and (ii) $16 h^{2\tau}_{\mathcal{X}_M} + h^d_{\mathcal{X}_M} M\xi$. Recalling that $C_{N,\mathcal{X}_M}$ is asymptotically independent of M (cf. Remark 3.2 and (14)), we find that the first contribution is independent of the number of measurements M; on the other hand, the second contribution is independent of the background $\mathcal{Z}_N$. We thus observe a multiplicative effect between M convergence (associated with the update) and N convergence (associated with the deduced background). For $\xi = 0$, the M term is guaranteed to decrease as M increases since $h_{\mathcal{X}_M}$ is monotonic decreasing in M; on the other hand, it is not possible to guarantee that $C_{N,\mathcal{X}_M}\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|^2$ is monotonic decreasing with N. We investigate this multiplicative effect in the numerical results.

3.1.2. A priori *error bounds for fixed-design regression*

We first introduce some notation. First, we define the matrix $\mathbb{A}_\xi \in \mathbb{R}^{N+M,N+M}$,

$$\mathbb{A}_\xi := \begin{bmatrix} \xi M \mathbb{I} + \mathbb{K} & \mathbb{L} \\ \mathbb{L}^T & 0 \end{bmatrix}, \tag{37a}$$

associated with the linear system (19b). Then, we introduce $\Sigma \in \mathbb{R}^{N+M,N+M}$,

$$\Sigma := \begin{bmatrix} \mathbb{I} & 0 \\ 0 & 0 \end{bmatrix}. \tag{37b}$$

Finally, we introduce $\mathbb{M} \in \mathbb{R}^{N+M,N+M}$ such that

$$\mathbb{M}_{i,i'} := \int_\Omega \psi_i(x)\,\psi_{i'}(x)\,dx, \qquad \psi_i(x) = \begin{cases} K_{x_i} & i = 1,\dots,M \\ \zeta_{i-M} & i = M+1,\dots,M+N \end{cases} \tag{37c}$$

We can now state the error bound.

Proposition 3.5. *Let Ω be a Lipschitz domain and let $\mathcal{U}$ be the Sobolev space $H^\tau(\Omega)$ with $\tau > d/2$. Let $\beta_{N,M}$ in (9) be strictly positive.*

Then, if $u^{\rm true} \in \mathcal{U}$, the following holds:

$$\mathbb{E}\left[\|u^{\rm true} - u^\star_\xi\|^2_{L^2(\Omega)}\right] \le \frac{1}{2} C_{N,\mathcal{X}_M} \left(16 h^{2\tau}_{\mathcal{X}_M} + h^d_{\mathcal{X}_M} M\xi\right) \|\Pi_{\mathcal{Z}_N^\perp} u^{\rm true}\|^2 + 2\sigma^2\,\mathcal{T}^\sigma, \tag{38}$$

where $\mathcal{T}^\sigma = \text{trace}\left(\mathbb{A}_\xi^{-1}\,\mathbb{M}\,\mathbb{A}_\xi^{-1}\,\Sigma\right)$. Furthermore, if $u^{\rm true} \in \mathcal{Z}_N$, we have

$$\mathbb{E}\left[\|u^{\rm true} - u^\star_\xi\|^2_{L^2(\Omega)}\right] = \sigma^2\,\mathcal{T}^\sigma. \tag{39}$$

Proof of Proposition 3.5 is presented in Appendix A. We observe that (38) can be easily extended to correlated noise by appropriately modifying the matrix Σ. We further observe that, unlike in the previous case, it is not evident how to provide explicit estimates for the optimal value of ξ. However, since $\mathcal{T}^\sigma$ is computable, if $u^{\rm true} \in \mathcal{Z}_N$, we can estimate numerically the optimal value of ξ *a priori*. We investigate this aspect in the numerical results.

3.2. *A posteriori* error analysis

Next result provides the identity of interest.

Proposition 3.6. *Let $\{x_i\}_{i=1}^I$ be drawn independently from an uniform distribution over Ω. Let $y_i = u^{\rm true}(x_i) + \delta_i + \epsilon_i$, where $\epsilon_1,\dots\epsilon_I$ are i.i.d. random variables such that $\epsilon_i \sim (0,\sigma^2)$ and $\delta_1,\dots,\delta_I$ are deterministic unknown disturbances. Let us further assume that $\{x_i\}_{i=1}^I$ and $\{\epsilon_i\}_{i=1}^I$ are independent random sequences.*

Then, we have that the mean squared error

$$MSE_I := \frac{1}{I}\sum_{i=1}^I \left(y_i - u^\star_\xi(x_i)\right)^2 \tag{40}$$

satisfies

$$\mathbb{E}[MSE_I] = E^2_{\rm mean} + \sigma^2 + \frac{1}{I}\sum_{i=1}^I \delta_i^2 - \frac{2}{|\Omega| I}\sum_{i=1}^I \delta_i \left(\int_\Omega u^{\rm true}(x) - u^\star_\xi(x)\,dx\right) \tag{41}$$

where $E^2_{\rm mean}$ is defined as follows:

$$E^2_{\rm mean} := \frac{1}{|\Omega|}\int_\Omega \left(u^{\rm true}(x) - u^\star_\xi(x)\right)^2 dx. \tag{42}$$

Proof. To simplify notation, we introduce the random sequence $\{e_i = u^{\rm true}(x_i) - u^\star_\xi(x_i)\}_{i=1}^I$. We observe that $e_1,\dots,e_I$ are i.i.d. and $\mathbb{E}[e_i^2] = \frac{1}{|\Omega|}\|u^{\rm true} - u^\star_{\lambda,\xi}\|^2_{L^2(\Omega)}$. Then, exploiting linearity of the expected value operator and the fact that $\{x_i\}_{i=1}^I$ and $\{\epsilon_i\}_{i=1}^I$ are independent, we find

$$\mathbb{E}[MSE_I] = \mathbb{E}\left[e_1^2\right] + \mathbb{E}\left[\epsilon_1^2\right] + \frac{1}{I}\sum_{i=1}^I \delta_i^2 - \frac{2}{I}\sum_{i=1}^I \delta_i\,\mathbb{E}[e_i].$$

Thesis follows. □

In absence of systematic noise ($\delta_i \equiv 0$), identity (41) reduces to

$$\mathbb{E}[MSE_I] = E^2_{\text{mean}} + \sigma^2. \tag{43}$$

Estimate (43) shows that for random noise ($\delta_i \equiv 0$) the mean squared error (40) can be used to asymptotically bound the squared $L^2(\Omega)$ error. Furthermore, since σ^2 is independent of the state estimate, minimizing the mean squared error is equivalent to minimize the $L^2(\Omega)$ error. The latter observation motivates the adaptive strategy presented in section 4.

4. Model adaptation

As observed in the previous sections, our procedure depends on a fair amount of design choices, which include the choice of a number of hyper-parameters and the choice of the observation centers and background space $\mathcal{Z}_N$. In section 4.1, we discuss how to exploit the error analysis to perform some design choices *a priori*. Then, in section 4.2, we discuss the adaptive strategy used to tune the parameters of the formulation after having acquired data.

4.1. *A priori* considerations

We recall that the APBDW state estimate $u^\star_\xi$ is given by

$$u^\star_\xi := \arg\min_{u\in\mathcal{U}} \xi\|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \frac{1}{M}\sum_{m=1}^{M} \left(u(x_m) - y_m\right)^2 \tag{44}$$

where $\mathcal{Z}_N = \text{span}\{\zeta_n\}_{n=1}^N$. We observe that the formulation depends on the regularization parameter ξ, the sensor locations $\mathcal{X}_M = \{x_m\}_{m=1}^M$ and the choice of the reproducing kernel K associated with $(\mathcal{U}, \|\cdot\|)$.

The hyper-parameter $\xi > 0$ controls the amount of regularization introduced: for $\xi \to 0^+$, the solution to (44) interpolates exactly the data while for $\xi \to \infty$, the solution to (44) converges to the least-squares solution. Our error analysis shows that the choice of ξ strongly depends on the noise variance σ^2 and on the maximum systematic error δ and also on the accuracy of the model $\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|$; in some applications, noise level can be estimated from reanalysis, on the other hand, it is extremely difficult to estimate $\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\|$ *a priori*.

Since in this work we employ csRBF kernels, the choice of the kernel K reduces to the choice of the hyper-parameters k and γ in $K_\chi(x,y) = \phi_{d,k}(\gamma\|x-y\|_2)$, where $\phi_{d,k}$ is defined in (25). As stated in Proposition 2.8, the parameter k determines the Sobolev regularity of the RKHS. Recalling estimate (35) and Proposition 2.8, the optimal value of k should minimise $\|\Pi_{\mathcal{Z}_N^\perp} u^{\text{true}}\| h_{\mathcal{X}_M}^{(d+1)/2+k}$: it is thus extremely problem-dependent. The parameter γ regulates the length scale of the kernel functions. In our experience, for small values of M, the choice of γ weakly influences the results; we can thus pick γ *a priori* such that the kernel functions $\{K_{x_m}\}_m$ share the same length scale with the elements of $\mathcal{Z}_N$. On the other hand, for larger values of M, the choice of γ significantly influences the performance of the method and it must be adapted using data. We remark that by changing k and γ we effectively modify the inner product $(\cdot,\cdot)$ and thus the penalization term $\|\cdot\|$ in (44).

If we neglect the effect of the sensor locations on the stability constant $C_{N,\mathcal{X}_M}$, the error analysis suggests to choose the observation centers to minimize the fill distance $h_{\mathcal{X}_M}$ in (32c). For $N \simeq M$, sensor location might influence significantly the value of $C_{N,\mathcal{X}_M}$. As a result, it might be worth to choose the observation centers to maximize $C_{N,\mathcal{X}_M}$ for any given M. For the PBDW formulation, in [36], the authors propose a Greedy strategy for the selection of the observation centers to maximize the inf-sup constant $\beta_{N,M}$ defined in (9). In this respect, we observe that calculations of $\beta_{N,M}$ (and also $C_{N,\mathcal{X}_M}$) involve computation of the matrix $\mathbb{Z}$ in (18). As observed in Remark 3.2, computation of accurate approximations of the matrix $\mathbb{Z}$ for general background

spaces is unfeasible. Extension of the Greedy procedure proposed in [36] to our framework is the topic of ongoing research.

In our numerical simulations, we choose adaptively the regularization parameter ξ and the kernel parameter γ, while we pick k *a priori*, and we simply consider equispaced observation centers. We remark that equispaced observation centers prevent us from considering $N \simeq M$. In the next section, we present the algorithm used to perform online adaptation. We note that our adaptation procedure could be also applied to automatically select the hyper-parameter k.

4.2. **Adaptive procedure**

In the Statistical Learning literature, several approaches have been presented to tune the design parameters of regularized regression formulations; we refer to [21, Chapter 7] and to [27] for a thorough overview. The adaptive strategy depends on the size of the dataset, which in our context corresponds to the amount of available transducers. If we denote by L the number of available transducers and by $\mathcal{D}_L = \{(x_\ell, y_\ell)\}_{\ell=1}^L$ the corresponding dataset, for large values of L, the holdout method is the most widely used approach. On the other hand, for small values of L, κ-fold cross-validation is typically employed. In the remainder of this section, we briefly review these techniques and we discuss their application to our problem.

The holdout method partitions the dataset $\mathcal{D}_L$ into the two mutually exclusive subsets $\mathcal{D}_M = \{(x_m, y_m)\}_{m=1}^M$ and $\mathcal{D}_I = \{(x_i, y_i)\}_{i=1}^I$. Given the finite dimensional search space $\mathfrak{H}^{\text{hyper}}$ for (ξ, γ), we generate the state estimate $u^\star_{\xi,\gamma}$ based on the training set and then we compute the mean squared error over the validation set

$$MSE_I(\xi, \gamma) = \frac{1}{I} \sum_{i=1}^{I} \left(y_i - u^\star_{\xi,\gamma}(x_i)\right)^2, \tag{45}$$

for each (ξ, γ) in $\mathfrak{H}^{\text{hyper}}$. Finally, we choose the state estimate associated with the choice of (ξ, γ) that minimizes $MSE_I(\xi, \gamma)$ over $\mathfrak{H}^{\text{hyper}}$. Recalling Proposition 3.6, if $\{x_i\}_i$ are drawn from an uniform distribution over Ω and the disturbances are homoscedastic, this choice of the hyper-parameters asymptotically minimizes the L^2 state-estimation error. This result holds independently of the strategy employed to compute the state estimate and thus independently of the strategy employed to select the training observation centers. As discussed in [51], if $u^\star_{\xi,\gamma}$ is an accurate description of the true field u^{true}, MSE_I rapidly converges to its expected value. Therefore, the number I of measurements that should be reserved for validation is modest.

Cross-validation is based on the partition of the dataset $\mathcal{D}_L$ into κ equal-sized subsamples (folds) $\{\mathcal{D}_L^{(k)}\}_{k=1}^\kappa$. Of the κ folds, a single fold is retained for testing and the remaining $\kappa - 1$ folds are used for training. The procedure is then repeated κ times with each of the κ folds used once as the validation dataset. In the limit $L = \kappa$, the procedure is known as Leave-One-Out Cross-Validation (LOOCV). We observe that, even for moderate L, κ-fold Cross-Validation can be quite expensive if $\kappa \approx L$. For this reason, generalized cross-validation strategies, which focus on computing computationally inexpensive approximations of the error indicator, have been developed. We refer to [21, Chapter 7.10] and to the references therein for further details.

In this paper, we exclusively employ holdout validation and we refer to a future work for the application of more advanced cross-validation strategies. Motivated by the previous discussion, we here choose the validation sensors by sampling uniformly over Ω.

5. Numerical results for a synthetic problem

In this section, we illustrate the behavior of the APBDW formulation through a two-dimensional acoustic Helmholtz problem[3]. We here employ csRBF with $k = 1$; recalling Proposition 2.8, this corresponds to $\mathcal{U} = H^{2.5}(\mathbb{R}^2)$. We appeal to the weak-Greedy algorithm to generate the background spaces $\{\mathcal{Z}_N\}_N$. We rely on holdout validation for the choice of ξ and of the kernel parameter γ: we consider uniform grids of training

[3]This model problem is the same considered in [36, Section 3].

observation points $\{x_m\}_{m=1}^M$, and uniformly random generated validation points $\{x_i\}_{i=1}^I$. In all our tests, we consider $I = \frac{M}{2}$.

5.1. **Problem definition**

Given the domain $\Omega = (0,1)^2$, we define the acoustic model problem:

$$\begin{cases} -(1+i\epsilon\mu)\,\Delta u_g(\mu) \,-\, \mu^2 u_g(\mu) = \,\mu\left(2x_1^2 + e^{x_2}\right) \,+\, \mu g & \text{in } \Omega, \\ \partial_n\, u_g(\mu) = 0 & \text{on } \partial\Omega, \end{cases} \tag{46}$$

where $\mu > 0$ is the wave number, $\epsilon = 10^{-2}$ is a fixed dissipation, and $g \in L^2(\Omega)$ is a bias term that will be specified later. Here, the parameter $\mu > 0$ constitutes the anticipated, parametric uncertainty in the system, which might model our uncertainty in the speed of sound, while the function g constitutes the unanticipated and non-parametric uncertainty in the system.

To assess the performance of the APBDW formulation for various configurations, we define the true field u^{true} as the solution to (46) for some $\mu^{\text{true}} \in \mathcal{P}^{\text{bk}}$ and for the following two choices of the "bias" g

$$g := \begin{cases} 0 & \text{perfect model;} \\ \tilde{g} \equiv 0.5(e^{-x_1} \,+\, 1.3\cos(1.3\pi x_2)) & \text{imperfect model.} \end{cases} \tag{47a}$$

On the other hand, we define the bk manifold as

$$\mathcal{M}^{\text{bk}} := \,\{u_{g=0}(\mu) :\, \mu \in \mathcal{P}^{\text{bk}}\}. \tag{47b}$$

Figure 1 shows (the real part of) the true field for three choices of the wave number μ and for the two choices of the bias g. We approximate the solution using a triangular $\mathbb{P}^5$ finite element discretization ($\mathcal{N} = 3312$). The use of a high-order method is here motivated by the smoothness of the true field.

To assess the performance, we introduce the relative L^2 error averaged over $|\mathcal{P}^{\text{bk}}_{\text{train}}| = n_{\text{train}}$ fields associated with different choices of the parameter μ:

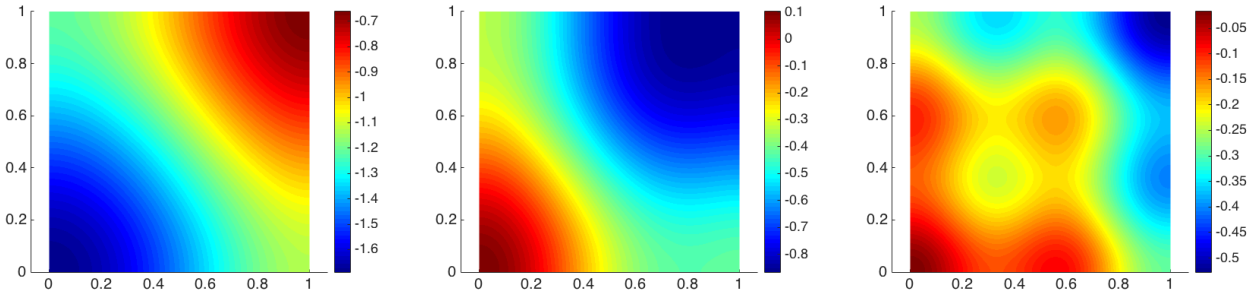
$$E^{\text{rel}}_{\text{avg}}(n_{\text{train}}) := \frac{1}{n_{\text{train}}} \sum_{\mu \in \mathcal{P}^{\text{bk}}_{\text{train}}} \frac{\|u^{\text{true}}(\mu) - u^\star_\xi(\mu)\|_{L^2(\Omega)}}{\|u^{\text{true}}(\mu)\|_{L^2(\Omega)}}. \tag{48}$$

In all our numerical tests, we consider noisy observations with additive Gaussian noise:

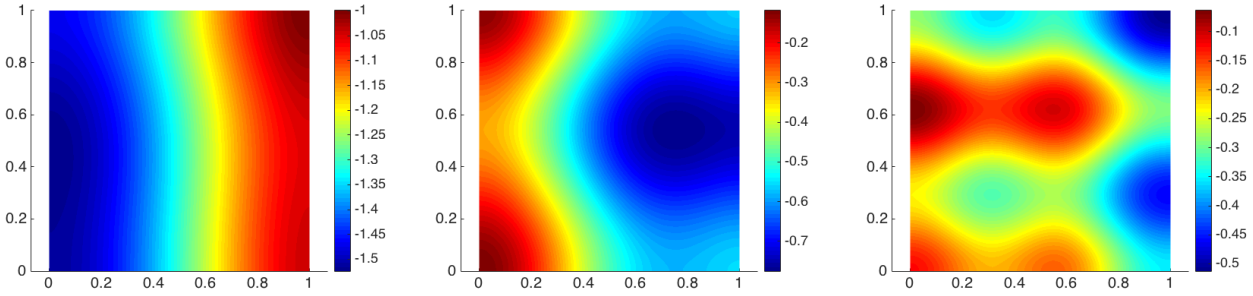
$$y_\ell = u^{\text{true}}(x_\ell) + \epsilon_\ell, \qquad \epsilon_\ell \overset{iid}{\sim} \mathcal{N}(0, \sigma^2). \tag{49}$$

5.2. **Results of the data assimilation procedure (noise-free case)**

We first visualize the APBDW state estimates for two distinct choices of u^{true}. We consider $\mu = 6.6$, and we consider $u^{\text{true}} = u_{g=0}(\mu)$ and $u^{\text{true}} = u_{g=\tilde{g}}(\mu)$; APBDW state estimates are based on the background $\mathcal{Z}_{N=5}$ and on $M = 25$ equispaced measurements. We rely on holdout validation ($I = 12$) to choose the value of the hyper-parameters ξ, γ. Figure 2 shows (the real part of) the true state, the APBDW state estimate $u^\star_\xi$, the deduced background $z^\star_\xi$ and the update $\eta^\star_\xi$. For $u^{\text{true}} = u_{g=0}(\mu)$ the update $\eta^\star_\xi$ is negligible; the reason is that the true state is well-approximated by its projection over $\mathcal{Z}_N$. On the other hand, for $u^{\text{true}} = u_{g=\tilde{g}}(\mu)$ we observe that the update is appreciable, and plays a significant role in improving the accuracy of the state estimate $u^\star_\xi$. These results strengthen the interpretation of the components of the APBDW state estimate provided in section 2: $z^\star_\xi$ addresses the parametric uncertainty in the model, while $\eta^\star$ accommodates non-parametric uncertainty.



(A) $\Re(u^{\text{true}})$, $\mu = 2$, $g = 0$ (B) $\Re(u^{\text{true}})$, $\mu = 5$, $g = 0$ (C) $\Re(u^{\text{true}})$, $\mu = 10$, $g = 0$



(D) $\Re(u^{\text{true}})$, $\mu = 2$, $g = \tilde{g}$ (E) $\Re(u^{\text{true}})$, $\mu = 5$, $g = \tilde{g}$ (F) $\Re(u^{\text{true}})$, $\mu = 10$, $g = \tilde{g}$

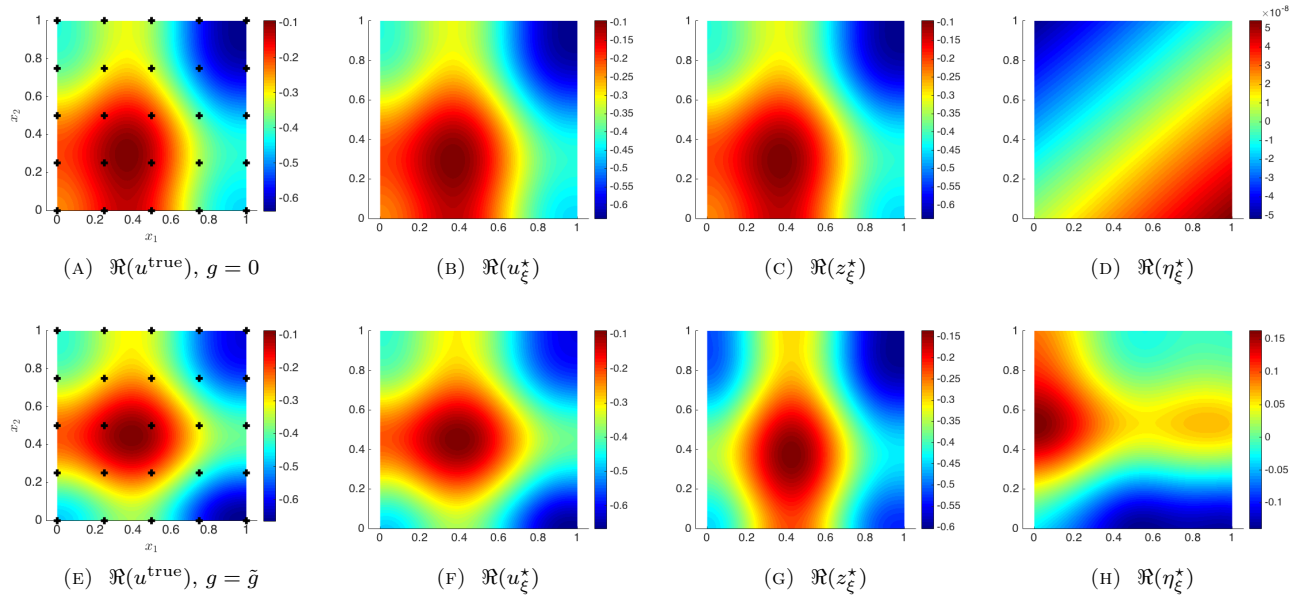
FIGURE 1. Application to a synthetic acoustic problem: visualization of the truth solutions associated with the synthetic Helmholtz problem for perfect ($g = 0$) and imperfect ($g = \tilde{g}$) models.

Figure 3 shows the convergence of $E^{\text{rel}}_{\text{avg}}$ with N for fixed M and noise-free measurements. We compute $E^{\text{rel}}_{\text{avg}}$ using (48) based on $n_{\text{train}} = 20$ fields. We observe that convergence with N is in good qualitative agreement with the behavior of the best-fit error

$$E^{\text{rel}}_N := \frac{1}{n_{\text{train}}} \sum_{\mu \in \mathcal{P}^{\text{bk}}_{\text{train}}} \frac{\|u^{\text{true}}(\mu) - \Pi_{\mathcal{Z}_N, L^2}\, u^{\text{true}}(\mu)\|_{L^2(\Omega)}}{\|u^{\text{true}}(\mu)\|_{L^2(\Omega)}},$$

where $\Pi_{\mathcal{Z}_N, L^2}$ is the projection operator with respect to the L^2 inner product. If $u^{\text{true}} \in \mathcal{M}^{\text{bk}}$, we observe fast convergence with N; on the other hand, if $u^{\text{true}} \notin \mathcal{M}^{\text{bk}}$, we reach a strictly positive plateau.

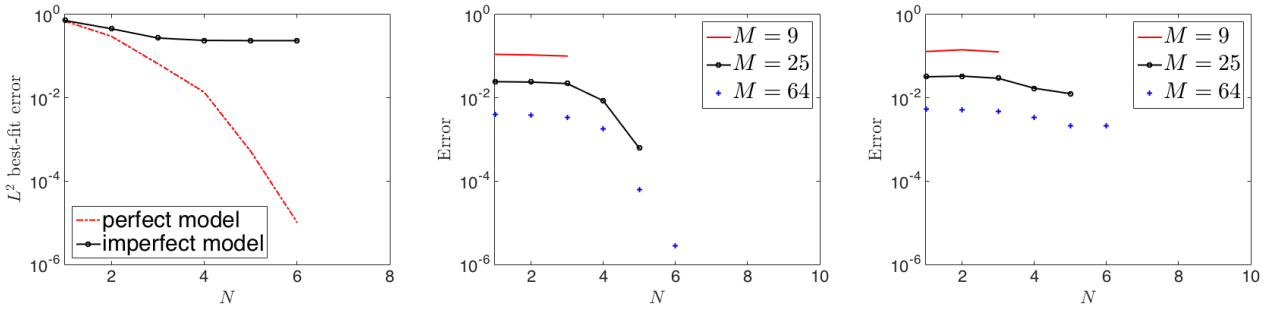
Figure 4 shows the convergence with M for fixed N and noise-free measurements. We assess performance by computing $E^{\text{rel}}_{\text{avg}}$ in (48) averaged over $n_{\text{train}} = 20$ fields. We observe that the use of csRBF kernels guarantees rapid convergence with M. We further observe that, with the exception of $N = 5$ for perfect model, the rate of convergence with M weakly depends on the value of N: in this test, we observe $E^{\text{rel}}_{\text{avg}} \simeq M^{-s}$ with $s \in [1.3, 1.5]$ for all cases considered. This empirically confirms the multiplicative effect between N convergence and M convergence observed in Remark 3.4. A possible explanation for the contrasting results for the case $(N = 5, g = 0)$ is due to discretization effects: since in this case the error $u^{\text{true}} - z^\star_\xi$ is highly oscillatory, the adaptive procedure selects large values of the parameter γ that are not well-resolved by the Finite Element mesh used to estimate the norms and to compute the true solution.



(A) $\Re(u^{\text{true}})$, $g = 0$ (B) $\Re(u^\star_\xi)$ (C) $\Re(z^\star_\xi)$ (D) $\Re(\eta^\star_\xi)$

(E) $\Re(u^{\text{true}})$, $g = \tilde{g}$ (F) $\Re(u^\star_\xi)$ (G) $\Re(z^\star_\xi)$ (H) $\Re(\eta^\star_\xi)$

FIGURE 2. Application to a two-dimensional acoustic problem: visualization of the PBDW state estimates for $N = 5$, $M = 25$ (perfect measurements). The states in Figures (a) and (e) correspond to $\mu = 6.6$.

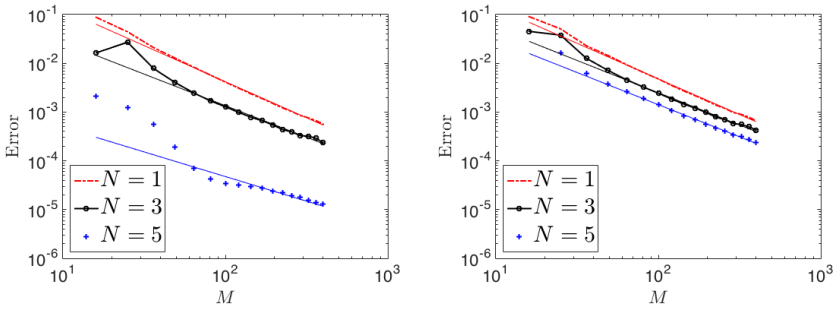


(A) L^2-best fit (B) $g = 0$ (C) $g = \tilde{g}$

FIGURE 3. Application to a two-dimensional acoustic problem: convergence of $E^{\text{rel}}_{\text{avg}}$ with N for fixed M for perfect ($g = 0$) and imperfect ($g = \tilde{g}$) model. Figure (a) shows the L^2-best-fit error.

5.3. **Interpretation of the hyper-parameters γ and ξ**

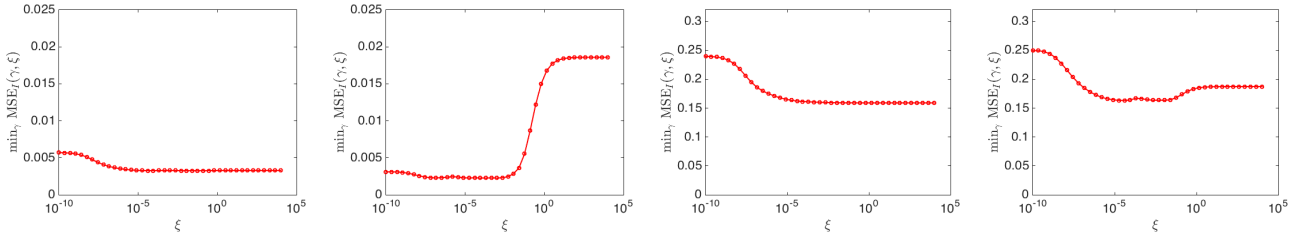
We investigate the connection between the optimal value of ξ and the signal-to-noise ratio. In Figure 5, we compute the mean squared error over the validation set for the estimation of the state associated with the parameter $\mu = 5.8$. We consider $M = 225$ and we compute the mean squared error based on $I = 110$ measurements. We both consider the case of perfect model ($g = 0$), and the case of imperfect model ($g = \tilde{g}$). For this test, we employ the background $\mathcal{Z}_{N=5}$. We observe that the optimal ξ depends on model error and on noise level. In more detail, we observe that for $g = 0$, the adaptive procedure selects large values of ξ regardless



(A) $g = 0$ (B) $g = \tilde{g}$

FIGURE 4. Application to a two-dimensional acoustic problem: convergence of $E^{\text{rel}}_{\text{avg}}$ with M for fixed N for perfect ($g = 0$) and imperfect ($g = \tilde{g}$) model. Estimated convergence rates for perfect model: -1.48 ($N = 1$), -1.30 ($N = 3$), and -1.00 ($N = 5$). Estimated convergence rates for imperfect model: -1.46 ($N = 1$), -1.32 ($N = 3$), and -1.32 ($N = 5$).

of the noise level, while for $g = \tilde{g}$ it selects $\xi \approx 10^{-7}$ for $\sigma = 0.05$ and $\xi \approx 10^{-5}$ for $\sigma = 0.4$. Therefore, the optimal ξ increases as noise increases, and decreases as best-fit error increases. The latter empirical observation is in good agreement with (35a), although the latter has been rigorously shown only for systematic noise.



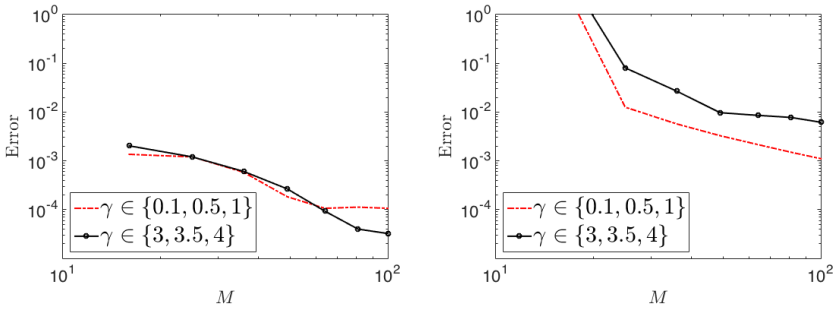
(A) $\sigma = 0.05$, $g = 0$ (B) $\sigma = 0.05$, $g = \tilde{g}$ (C) $\sigma = 0.4$, $g = 0$ (D) $\sigma = 0.4$, $g = \tilde{g}$

FIGURE 5. Application to a two-dimensional acoustic problem: interpretation of ξ. Results correspond to $u^{\text{true}} = u_g(\mu = 5.8)$. ($M = 225$, $I = 112$, $N = 5$).

In Figure 6, we investigate the influence of the kernel parameter γ. We study the behavior of $E^{\text{rel}}_{\text{avg}}$ with M associated with $n_{\text{train}} = 10$ different values of the parameter μ, for the five-dimensional background $\mathcal{Z}_{N=5}$ and for two different search spaces $\mathfrak{H}^{\text{hyper}}$: in more detail, in the first case we seek γ in $\{0.1, 0.5, 1\}$, and in the second case we choose γ in $\{3, 3.5, 4\}$. Since we consider perfect measurements, results are not sensitive to the choice of ξ. We observe that in the perfect-model case (Figure 6(a)) large values of γ significantly improve performance; on the other hand, in the imperfect model case (Figure 6(b)), the first choice of $\mathfrak{H}^{\text{hyper}}$ leads to more accurate results for all values of M considered. This can be explained by observing that γ has to match the length-scale of the field $u^{\text{true}} - z^{\star}_{\xi}$, and strongly depends on the distance between observations (and thus M). This test motivates the importance of adapting the value of γ. We remark that adaptation in γ relies on the availability of explicit expressions for the Riesz elements K_{x_m}.

5.4. **Noisy measurements**

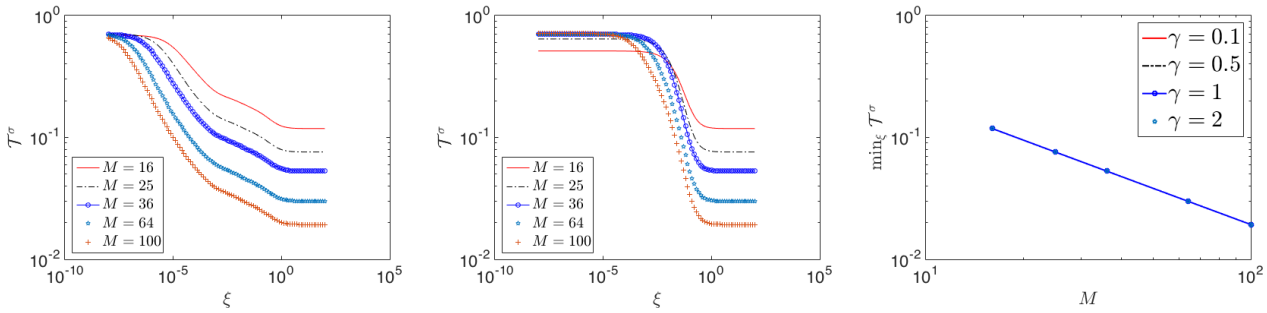
We first study the behavior of the constant $\mathcal{T}^{\sigma}$ introduced in (38). Figures 7(a) and (b) show the behavior of $\mathcal{T}^{\sigma}$ for equispaced measurements with respect to the value of ξ and for two values of γ. We observe that $\mathcal{T}^{\sigma}$



(A) $g = 0$, $N = 5$ (B) $g = \tilde{g}$, $N = 5$

FIGURE 6. Application to a two-dimensional acoustic problem: interpretation of γ ($N = 5$, $\sigma = 0$).

is monotonic decreasing in ξ and reaches a lower bound for $\xi \to \infty$. Figure 7(c) shows the behavior of $\min_\xi \mathcal{T}^\sigma$ with respect to the number of measurements: $\min_\xi \mathcal{T}^\sigma$ is independent of γ and converges to 0 with rate M^{-1}.



(A) $\gamma = 0.1$ (B) $\gamma = 2$ (C) $\gamma = 2$

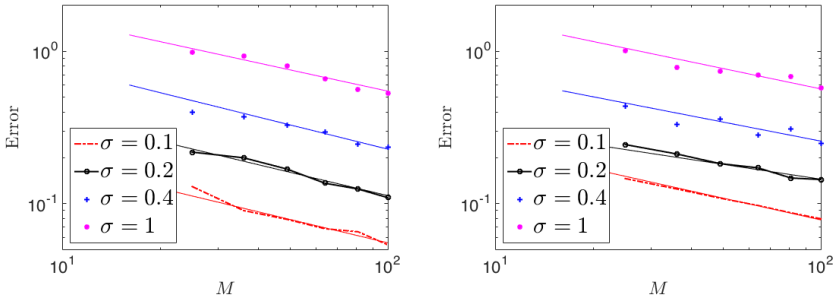
FIGURE 7. Application to a two-dimensional acoustic problem: $\mathcal{T}^\sigma$. Figures (a) and (b): behavior of $\mathcal{T}^\sigma$ with ξ for $\gamma = 0.1$ and $\gamma = 2$. Figure (c): behavior of $\min_\xi \mathcal{T}^\sigma$ with M for several values of γ.

Figure 8 shows performance in presence of noise. As in the previous tests, we assess performance by computing $E^{\text{rel}}_{\text{avg}}$ in (48) for $n_{\text{train}} = 1$ ($\mu = 6.6$); to take into account randomness in the results, we average over 24 realizations of the random noise. We consider the background $\mathcal{Z}_{N=5}$. In the case of perfect model, the estimated convergence rate in the noisy case is roughly $M^{-0.5}$ for all values of standard deviations σ considered: this is in agreement with the results shown in Figure 7(c) and with the mathematical analysis. On the other hand, in the case of imperfect model, the estimated convergence rate in the noisy case is roughly $M^{-0.4}$. Interestingly, also in this case, the convergence rate weakly depends on σ.

6. Application to a thermal patch configuration

6.1. **Experimental apparatus**

The thermal patch system consists of a 1.5[mm] thick acrylic sheet heated from behind by a resistive patch. Heat is generated through an electrical resistance with input power equal to $0.667W$. The goal of the data



(A) $g = 0$, $N = 5$ (B) $g = \tilde{g}$, $N = 5$

FIGURE 8. Application to a two-dimensional acoustic problem: convergence with M for fixed N for perfect ($g = 0$) and imperfect ($g = \tilde{g}$) model in presence of homoscedastic Gaussian noise. Estimated convergence rates for perfect model: -0.5114 ($\sigma = 0.1$), -0.5091 ($\sigma = 0.2$), -0.5297 ($\sigma = 0.4$), and -0.4641 ($\sigma = 1$). Estimated convergence rates for imperfect model: -0.4759 ($\sigma = 0.1$), -0.3235 ($\sigma = 0.2$), -0.4155 ($\sigma = 0.4$), and -0.4443 ($\sigma = 1$).

assimilation procedure is to estimate the temperature field over a portion $\Omega^{\text{obs,dim}}$ of the external surface of the plate at the steady-state limit.

We use an IR camera (`Fluke Ti 9`) to take measurements in the rectangular region $\Omega^{\text{obs,dim}} = [-23.85, 23.85] \times [-17.85, 17.85]$[mm] centered on the patch. Figure 9(a) shows the IR camera. After the patch power is turned on, we take measurements using a sampling time of 4 seconds until steady state is reached; the total duration of the experiment is roughly 5 minutes. The external temperature is about $20^o C$, roughly constant throughout the experiment. Each surface measurement taken from the IR camera corresponds to 160×120 pixel-wise measurements; the pixel size is roughly $\Delta h^{\text{device}} = 0.3$[mm], which is much smaller than the spatial length scale of the phenomenon of interest.

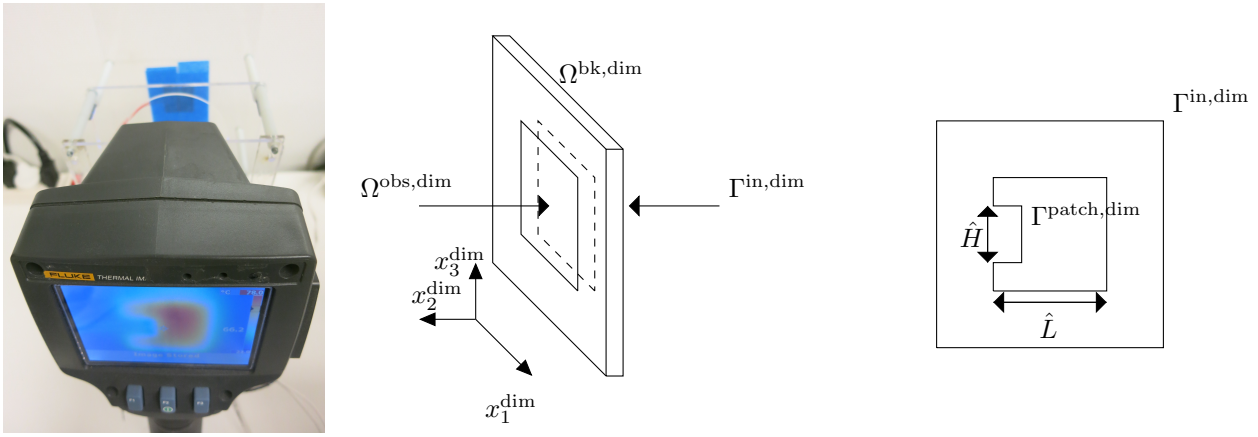
In view of the mathematical description of the problem, we present formal definitions for the geometric quantities involved. First, we introduce the domain $\Omega^{\text{bk,dim}} \subset \mathbb{R}^3$ corresponding to the three-dimensional acrylic sheet. We denote by $\Gamma^{\text{patch,dim}} \subset \mathbb{R}^2$ the surface of the sheet attached to the patch, and we denote by $\Gamma^{\text{in,dim}}$ the face of the sheet that contains $\Gamma^{\text{patch,dim}}$. We recall that $\Omega^{\text{obs,dim}} \subset \partial\Omega^{\text{bk,dim}}$ is the region in which the IR camera takes measurements. Then, we introduce the Cartesian coordinate system $x_1^{\text{dim}} x_2^{\text{dim}} x_3^{\text{dim}}$; according to our definitions, the IR camera takes measurements in the $x_1^{\text{dim}} x_3^{\text{dim}}$ plane. Figures 9(b) and (c) clarify the definitions of $\Omega^{\text{obs,dim}}$, $\Omega^{\text{bk,dim}}$, $\Gamma^{\text{patch,dim}}$ and $\Gamma^{\text{in,dim}}$ and show the characteristic dimensions of the patch.

In order to estimate the noise level in the dataset, we compute the difference $u^{\text{obs,dim}} - u^{\text{filt,dim}}$ where the field $u^{\text{obs,dim}}$ is obtained directly from the IR camera, and $u^{\text{filt,dim}}$ is obtained applying a Wiener filter (see, e.g., [31]) based on a 3 by 3 pixel averaging to the field $u^{\text{obs,dim}}$. Figure 10 shows two spatial slices of the difference $u^{\text{obs,dim}} - u^{\text{filt,dim}}$. By comparing $u^{\text{filt,dim}}$ and $u^{\text{obs,dim}}$, we deduce that the magnitude of noise in the measurements is approximately $\pm 0.5^o C$, roughly independent of the spatial position.

6.2. **Engineering motivation**

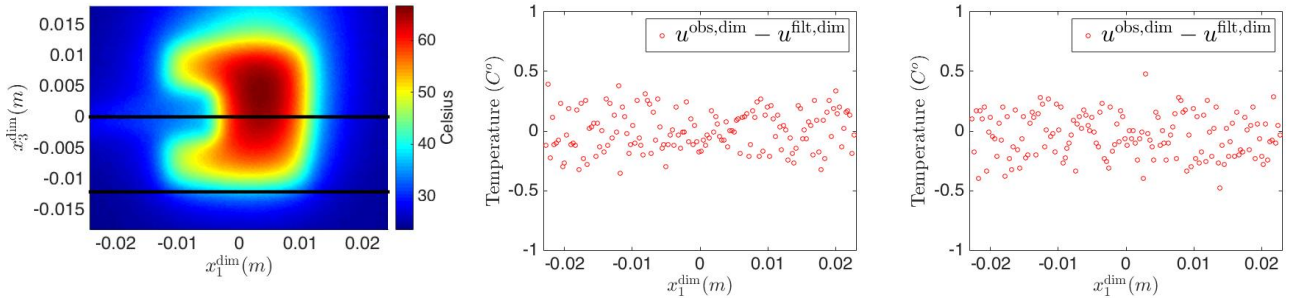
We shall now motivate this model problem from the engineering standpoint. Full-field information is typically not available; in practical applications, we envision a system with a local sensor or a small sensor array. For this reason, we want to design a data assimilation state estimation procedure that is able to reconstruct the full field based on a small amount of local measurements.

Since the IR camera provides full-field information, in this work, we synthetize local measurements – the experimental input to our methods – from the IR camera to obtain $\ell_m^{\text{obs}} = \ell(u^{\text{obs}}, x_m^{\text{obs}})$ where the observation



(A)

FIGURE 9. Thermal patch experiment. Figure (a): IR camera. Figures (b) and (c): mathematical description of the acrylic sheet. $\hat{L} = 22.606$mm, $\hat{H} = 9.271$mm.



(A) $u^{\text{obs,dim}}$ (B) $x_3^{\text{dim}} = 0.2$mm (C) $x_3^{\text{dim}} = -12.0$mm

FIGURE 10. Thermal patch experiment: comparison between filtered and unfiltered fields. Figure (a): observed thermal field $u^{\text{obs,dim}}$. Figures (b) and (c): spatial slices of the difference $u^{\text{obs,dim}} - u^{\text{filt,dim}}$.

functional $\ell(\cdot, x_m^{\text{obs}})$ is designed to represent a *fictitious* measurement in the sensor location $x_m^{\text{obs}} \in \Omega^{\text{obs}}$. We observe that the IR camera permits us to conduct convergence studies that would typically not be feasible in actual field deployment.

6.3. **Mathematical model and background space**

We resort to a steady-state heat-transfer model in which the heat-exchange between the patch and the sheet is described by means of a Robin boundary condition. In more detail, we consider the following bk model for

the thermal field $u^{\mathrm{bk,dim}} : \Omega^{\mathrm{bk,dim}} \to \mathbb{R}$:

$$\begin{cases} -\Delta u^{\mathrm{bk,dim}} = 0, & \text{in } \Omega^{\mathrm{bk,dim}}, \\ \kappa \partial_n u^{\mathrm{bk,dim}} + \psi(u^{\mathrm{bk,dim}} - \Theta^{\mathrm{room,dim}}) = g^{\mathrm{dim}} \chi_{\Gamma^{\mathrm{patch,dim}}} & \text{on } \Gamma^{\mathrm{in,dim}}, \\ \kappa \partial_n u^{\mathrm{bk,dim}} = 0 & \text{on } \partial\Omega^{\mathrm{bk,dim}} \setminus \Gamma^{\mathrm{in,dim}}, \end{cases} \tag{50}$$

where ψ is the convective heat transfer coefficient, κ is the thermal conductivity, $\Theta^{\mathrm{room,dim}} = 20^o C$ is the room temperature, and g^{dim} is the incoming flux, which models the heat exchange between the patch and the plate. Textbook values for the model parameters are $\kappa = 0.2\mathrm{W/m}$, $\psi = 10\mathrm{W/m^2}$.

Given the thermal field $u^{\mathrm{bk,dim}}$, we introduce the non-dimensional counterpart

$$u^{\mathrm{bk}}(x) = \frac{u^{\mathrm{bk,dim}}(\hat{L}x) - \Theta^{\mathrm{room,dim}}}{\Delta\Theta}, \tag{51}$$

where $\Delta\Theta = 50^o C$ is a rough approximation of the temperature difference between the far-field and the center of the patch, $\hat{L} = 22.606\mathrm{mm}$ is the length of the edge of the patch (see Figure 9). We observe that $u^{\mathrm{bk}} = u^{\mathrm{bk}}(\mu)$ satisfies

$$\begin{cases} -\Delta u^{\mathrm{bk}}(\mu) = 0, & \text{in } \Omega^{\mathrm{bk}}, \\ \partial_n u^{\mathrm{bk}}(\mu) + \mu u^{\mathrm{bk}}(\mu) = g & \text{on } \Gamma^{\mathrm{in}}, \\ \partial_n u^{\mathrm{bk}}(\mu) = 0 & \text{on } \partial\Omega^{\mathrm{bk}} \setminus \Gamma^{\mathrm{in}}, \end{cases} \tag{52a}$$

where $\mu = \hat{L}\psi/\kappa \approx 1.13$ and g is defined as follows:

$$g(x) = C\, \chi_{\Gamma^{\mathrm{patch}}}(x). \tag{52b}$$

We observe that u^{bk} depends on the parameters μ and C. Since the model is linear with respect to C and our ultimate goal is to define a linear space associated with the bk manifold, we can simply set $C = 1$. On the other hand, assuming that the estimate of κ is accurate and that $\psi \approx 10 \pm 5\mathrm{W/m^2}$, we obtain that $\mu \in \mathcal{P}^{\mathrm{bk}} = [0.5650, 1.650]$. We can thus define the bk manifold as follows:

$$\mathcal{M}^{\mathrm{bk}} = \left\{ u^{\mathrm{bk}}(\mu)|_{\Omega^{\mathrm{obs}}} : \mu \in \mathcal{P}^{\mathrm{bk}} \right\}. \tag{53}$$

We observe that our parametrized model encodes the uncertainty in the material properties ψ and κ. On the other hand, it does not take into account the nonlinear effects associated to natural convection, and to the heat-exchange between the patch and the sheet. The latter represent the non-parametric uncertainty in the model.

The background space $\mathcal{Z}_N$ associated with (52)-(53) is built using the weak-Greedy algorithm. To compute the solution to the bk model, we appeal to a $\mathbb{P}^3$ continuous Finite Element discretization based on $\mathcal{N} = 40000$ degrees of freedom.

6.4. **Numerical results**

We interpret pixel-wise measurements as pointwise evaluations associated with the center of the pixel. As in the previous test, we perform holdout validation for ξ and γ with $I = M/2$. We assess performance by computing the relative mean squared error $MSE^{\mathrm{rel}} = \|u^{\mathrm{obs}} - u^{\star}_{\xi}\|^2_{L^2(\Omega)} / \|u^{\mathrm{obs}}\|^2_{L^2(\Omega)}$ based on the full-field information.

Figure 11 shows the convergence of MSE^{rel} with M for three values of N. We observe that, while including the first snapshot leads to a substantial improvement in the performances, considering $N > 1$ does not lead to

any substantial improvement. We further observe that for $M \approx 100$ we reach the estimated noise level

$$\sigma_{\text{est}} := \frac{\|\frac{0.5^o C}{\Delta\Theta}\|_{L^2(\Omega)}}{\|u^{\text{obs}}\|_{L^2(\Omega)}}.$$

As M increases, MSE^{rel} becomes significantly lower than σ^2_{est}: this can be explained by observing that for large values of M the amount of pixels used for learning (training plus validation) is not negligible compared to the entire dataset.

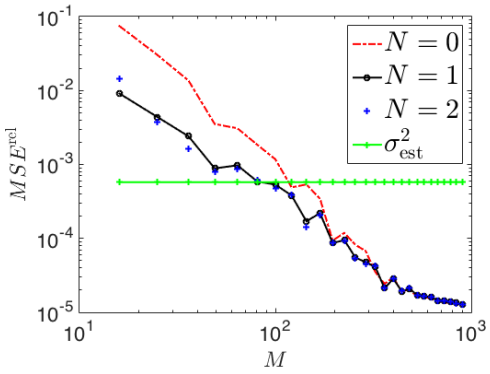


FIGURE 11. Application to the thermal patch experiment: convergence of the relative mean squared error MSE^{rel} with M for fixed N.

7. CONCLUSIONS

In this paper, we extended the PBDW formulation first proposed in [36] to pointwise noisy measurements. The extension relies on an adaptive procedure that properly takes into account the noise level, and the characteristic length-scale of the difference $u^{\text{true}} - z^\star_\xi$. The use of explicit kernels allows us to perform online adaptation to tune the characteristic length-scale of the update functions. We presented *a priori* and *a posteriori* error estimates for the L^2 state-estimation error to motivate the approach from a theoretical standpoint. We presented several numerical results to illustrate the different elements of the formulation. In more detail, numerical experiments demonstrated (i) a multiplicative effect between N convergence (associated with the primary approximation provided by the background $\mathcal{Z}_N$) and M convergence (associated with the secondary approximation provided by the update $\mathcal{U}_M$), (ii) the practical importance of adapting the shape of the Riesz representers based on data, and (iii) L^2 convergence of the APBDW estimate to the true state even for noisy measurements.

We now identify four extensions to the approach, which are subjects of future work. First, we wish to design strategies for the selection of the observation centers that address both stability and approximation. This would allow us to consider the case $N \simeq M$, and would tighten the connection between our approach and the design of the experiment. In this respect, we wish to combine the Greedy procedure presented in [36, Algorithm 2] with techniques developed in the kernel methods' literature for collocation methods for PDEs ([23, 48]) and scattered data approximation ([39, section 3.1.1], [61]). Second, we wish to consider more general observation functionals of the form $y_m = \ell^o_m(u^{\text{true}}) + \epsilon_m$. This would allow us to take into account different sources of information. In this respect, we observe that the well-posedness analysis can be trivially extended to the case in which $\ell^o_1, \ldots, \ell^o_M \in \mathcal{U}'$, while the error analysis relies on the form of the observation functionals. Third, we wish to exploit the interpretation of APBDW as convex relaxation of the partial spline model for a particular choice of the background to derive new probabilistic and deterministic error bounds and possibly improve the

performance of the state-estimation procedure. Finally, we wish to extend our technique to time-dependent problems.

Acknowledgments: the author thanks Prof. Anthony T. Patera (MIT) for fruitful discussions, Dr. James D. Penn (MIT) for the experimental results in section 6, Prof. Masayuki Yano (University of Toronto) for the development of the high-order FE code used in the numerical sections, and the unknown reviewers for their insightful comments that help improve the manuscript. This work was supported by OSD/AFOSR/MURI Grant FA9550-09-1-0613, ONR Grant N00014-11-1-0713, and the MIT-Singapore International Design Center.

Appendix A. Proofs of the error bounds

In this appendix, we provide the proofs of the results presented in section 3. The appendix is organized as follows. In section A.1, we introduce a regularized formulation and we show the connection with our formulation. In section A.2, we exploit the results for the regularized problem to prove Proposition 3.1. Finally, in section A.3, we prove the result for fixed-design regression.

A.1. **Preliminaries**

We introduce a regularized formulation of the APBDW statement proposed in this work: given $\lambda \geq 0$, $\xi > 0$, find $u^\star_{\lambda,\xi} \in \mathcal{U}$ such that

$$u^\star_{\lambda,\xi} = \arg\min_{u\in\mathcal{U}} \; J^{(1)}_{\lambda,\xi}(u) := \xi\|u\|^2_{\lambda,N} + V_M(u), \tag{54}$$

where the seminorm $\|\cdot\|_{\lambda,N}$ is defined as

$$\|w\|^2_{\lambda,N} = \lambda\|\Pi_{\mathcal{Z}_N} w\|^2 + \|\Pi_{\mathcal{Z}_N^\perp} w\|^2. \tag{55}$$

We observe that for any $\lambda > 0$, the function $\|\cdot\|_{\lambda,N}$ is a norm equivalent to $\|\cdot\|$. We also observe that for $\lambda = 0$, problem (54) corresponds to (11).

Next Proposition summarizes a number of properties of problem (54) that are crucial to prove the error bounds for $u^\star_\xi$.

Proposition A.1. *Let $\beta_{N,M} > 0$. Then, the following hold.*

1. *For any $\lambda > 0$, the solution to (54) exists and is unique. Furthermore, if we introduce $\eta^\star_{\lambda,\xi} = \Pi_{\mathcal{Z}_N^\perp} u^\star_{\lambda,\xi}$, $z^\star_{\lambda,\xi} = \Pi_{\mathcal{Z}_N} u^\star_{\lambda,\xi}$, we have that $\eta^\star_{\lambda,\xi} \in \mathrm{span}\{\Pi_{\mathcal{Z}_N^\perp} K_{x_m}\}_{m=1}^M$ and $z^\star_{\lambda,\xi} \in \mathrm{span}\{\Pi_{\mathcal{Z}_N} K_{x_m}\}_{m=1}^M$.*
2. *For any $\xi > 0$, the solution $u^\star_{\lambda,\xi}$ converges to the solution $u^\star_\xi$ to (11) when $\lambda \to 0^+$.*
3. *For any $\lambda \geq 0$, the following bounds hold:*

$$\|u^{\mathrm{true}} - u^\star_{\lambda,\xi}\|_{\lambda,N} \leq 2\|u^{\mathrm{true}}\|_{\lambda,N} + \frac{\delta}{2\sqrt{\xi}}, \tag{56a}$$

and

$$\|u^{\mathrm{true}} - u^\star_{\lambda,\xi}\|_{\ell^2(\mathcal{X}_M)} \leq \sqrt{M}\left(\delta + \frac{\sqrt{\xi}}{2}\|u^{\mathrm{true}}\|_{\lambda,N}\right). \tag{56b}$$

We prove each statement separately.

Proof. (*statement 1*) For any $\lambda > 0$, $u \mapsto \|u\|^2_{\lambda,N}$ is strictly convex, while $u \mapsto V_M(u)$ is convex. This implies that for any $\xi > 0$ the objective function $J_{\lambda,\xi}(u) = \xi\|u\|^2_{\lambda,N} + V_M(u)$ is strictly convex. Therefore, existence and uniqueness of the solution to (54) follow from [16, Theorem 3, Chapter 8.2].

We observe that $x \in \Omega \mapsto \Phi^{\lambda,N}_x = \frac{1}{\lambda}\Pi_{\mathcal{Z}_N} K_x + \Pi_{\mathcal{Z}_N^\perp} K_x$, is the feature map associated with $(\mathcal{U}, \|\cdot\|_{\lambda,N})$. We have indeed that for all $v \in \mathcal{U}$ and $x \in \Omega$

$$(\Phi^{\lambda,N}_x, v)_{\lambda,N} = \frac{\lambda}{\lambda}(\Pi_{\mathcal{Z}_N} K_x, v) + (\Pi_{\mathcal{Z}_N^\perp} K_x, v) = (K_x, \Pi_{\mathcal{Z}_N} v + \Pi_{\mathcal{Z}_N^\perp} v) = (K_x, v) = v(x).$$

Exploiting the representer theorem (see, e.g., [58, Theorem 16.1]), we have that $u^\star_{\lambda,\xi} \in \text{span}\{\Phi^{\lambda,N}_{x_m}\}_{m=1}^M$. As a result, we have that $\eta^\star_{\lambda,\xi} \in \text{span}\{\Pi_{\mathcal{Z}_N^\perp}\Phi^{\lambda,N}_{x_m}\}_{m=1}^M$, and $z^\star_{\lambda,\xi} \in \text{span}\{\Pi_{\mathcal{Z}_N}\Phi^{\lambda,N}_{x_m}\}_{m=1}^M$ for any $\lambda > 0$. □

Proof. (*statement 2*) Let $\{\lambda_j\}_j$ be a real sequence such that $\lambda_j \to 0^+$. Exploiting the first statement of Proposition A.1, we have that sequences $\{\eta^\star_{\lambda_j,\xi}\}_j$, $\{z^\star_{\lambda_j,\xi}\}_j$ belong to finite dimensional spaces that do not depend on λ. Furthermore, applying Lemma 2.3, it is possible to verify that they are uniformly bounded for all j. Applying Bolzano-Weierstrass theorem, the sequence $\{u^\star_{\lambda_j,\xi} = \eta^\star_{\lambda_j,\xi} + z^\star_{\lambda_j,\xi}\}_j$ admits a strongly convergent subsequence $\{u^\star_{\lambda_k,\xi}\}_k$ to $\widehat{u}^\star_\xi \in \mathcal{U}$.

We now show that $\widehat{u}^\star_\xi = u^\star_\xi$. We first observe that

$$J^{(1)}_{\lambda_k,\xi}(u^\star_{\lambda_k,\xi}) = \lambda_k \underbrace{\|z^\star_{\lambda_k,\xi}\|^2}_{\leq C} + \|\eta^\star_{\lambda_k,\xi}\|^2 + V_M(u^\star_{\lambda_k,\xi}) \to J^{(1)}_\xi(\widehat{u}^\star_\xi), \qquad k \to \infty.$$

We further observe that for any $\lambda_k > 0$

$$J^{(1)}_{\lambda_k,\xi}(u^\star_{\lambda_k,\xi}) \leq J^{(1)}_{\lambda_k,\xi}(u^\star_\xi), \qquad k = 1, 2, \ldots,$$

and by taking the limit on both sides, we obtain

$$J^{(1)}_\xi(\widehat{u}^\star_\xi) \leq J^{(1)}_\xi(u^\star_\xi)$$

Since $u^\star_\xi$ is the unique minimizer of (11), we must have $u^\star_\xi = \widehat{u}^\star_\xi$. Furthermore, by the same argument, $u^\star_\xi$ must be the only limit point of the sequence; therefore, the entire sequence converges to $\widehat{u}^\star_\xi$. Thesis follows. □

Proof. (*statement 3*) For $\lambda > 0$, $\|\cdot\|_{\lambda,N}$ is a norm for $\mathcal{U}$; therefore, estimates (56a) and (56b) follow directly from [28, Corollary 4.3] and [28, Lemma 4.5].

The extension to $\lambda = 0$ follows by observing that $u^\star_{\lambda,\xi}$ converges to $u^\star_\xi$ when $\lambda \to 0^+$. □

Before proving the error bounds, we prove (32b).

Lemma A.2. *Let Ω be a Lipschitz domain and let $\mathcal{U}$ be the Sobolev space $H^\tau(\Omega)$ with $\tau > d/2$. Let us assume that inf-sup constant $\beta_{N,M}$ defined in (9) is strictly positive and $h_{\mathcal{X}_M} < 1$.*

Then,

$$C_{N,\mathcal{X}_M} \leq \frac{1}{\min\{c_{N,M}, 1 - h^{2\tau-d}_{\mathcal{X}_M}\}}\, C,$$

where $c_{N,M}$ is defined in (14) and C depends on the domain Ω and on $(\cdot,\cdot)$.

Proof. Let us define the constant

$$\widehat{C}_{\mathcal{X}_M} := \sup_{u\in\mathcal{U}} \frac{\|u\|^2_{L^2(\Omega)}}{h^{2\tau}_{\mathcal{X}_M}\|u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)}}.$$

Recalling [28, Theorem 4.8], $\widehat{C}_{\mathcal{X}_M}$ is bounded from above by a constant C that does not depend on M.

Since $\beta_{N,M} > 0$, recalling Lemma 2.3, we have that

$$\|\Pi_{\mathcal{Z}_N^\perp} u\|^2 + \|u\|^2_{\ell^2(\mathcal{X}_M)} \geq c_{N,M}\|u\|^2,$$

where $c_{N,M} > 0$ is given by the expression in (14). Then, we observe that

$$\begin{aligned} h^{2\tau}_{\mathcal{X}_M}\|\Pi_{\mathcal{Z}_N^\perp}u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)} = &\ h^{2\tau}_{\mathcal{X}_M}\left(\|\Pi_{\mathcal{Z}_N^\perp}u\|^2 + \|u\|^2_{\ell^2(\mathcal{X}_M)}\right) + \left(h^d_{\mathcal{X}_M} - h^{2\tau}_{\mathcal{X}_M}\right)\|u\|^2_{\ell^2(\mathcal{X}_M)} \\ \geq &\ c_{N,M}h^{2\tau}_{\mathcal{X}_M}\|u\|^2 + \left(1 - h^{2\tau-d}_{\mathcal{X}_M}\right) h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)} \\ \geq &\ \min\{c_{N,M}, 1 - h^{2\tau-d}_{\mathcal{X}_M}\}\left(h^{2\tau}_{\mathcal{X}_M}\|u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)}\right). \end{aligned}$$

As a result,

$$C_{N,\mathcal{X}_M} = \sup_{u\in\mathcal{U}} \frac{\|u\|^2_{L^2(\Omega)}}{h^{2\tau}_{\mathcal{X}_M}\|\Pi_{\mathcal{Z}_N^\perp}u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)}} \leq \underbrace{\left(\sup_{u\in\mathcal{U}} \frac{\|u\|^2_{L^2(\Omega)}}{h^{2\tau}_{\mathcal{X}_M}\|u\|^2 + h^d_{\mathcal{X}_M}\|u\|^2_{\ell^2(\mathcal{X}_M)}}\right)}_{=C} \frac{1}{\min\{c_{N,M}, 1 - h^{2\tau-d}_{\mathcal{X}_M}\}}$$

Thesis follows. □

A.2. **Proof of the error bound for scattered data**

Proof. (*Proposition 3.1*) The proof replicates the argument of [28, Theorem 4.11]. Recalling the definition of $C_{N,\mathcal{X}_M}$, we have

$$\|u^{\text{true}} - u^\star_\xi\|^2_{L^2(\Omega)} \leq C_{N,\mathcal{X}_M}\left(h^{2\tau}_{\mathcal{X}_M}\|\Pi_{\mathcal{Z}_N^\perp}(u^{\text{true}} - u^\star_\xi)\|^2 + h^d_{\mathcal{X}_M}\|u^{\text{true}} - u^\star_\xi\|^2_{\ell^2(\mathcal{X}_M)}\right).$$

Then, using (56a) and (56b), we obtain

$$\|u^{\text{true}} - u^\star_\xi\|^2_{L^2(\Omega)} \leq C_{N,\mathcal{X}_M}\left(h^{2\tau}_{\mathcal{X}_M}\left(2\|\Pi_{\mathcal{Z}_N^\perp}u^{\text{true}}\|^2 + \frac{\delta}{2}\frac{1}{\sqrt{\xi}}\right)^2 + h^d_{\mathcal{X}_M}M\left(\delta + \frac{\sqrt{\xi}}{2}\|\Pi_{\mathcal{Z}_N^\perp}u^{\text{true}}\|\right)^2\right),$$

which is the thesis. □

A.3. **Proof of the error bound for fixed-design regression**

We first introduce some notation and preliminary definitions. We decompose the datum $\mathbf{y}_M$ as

$$\mathbf{y}_M = \mathbf{y}_M^{\text{true}} + \boldsymbol{\epsilon}, \qquad \mathbf{y}_M^{\text{true}} = [u^{\text{true}}(x_1), \dots, u^{\text{true}}(x_M)], \quad \boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_M],$$

and we define $\boldsymbol{\epsilon}_{aug} = \begin{bmatrix} \boldsymbol{\epsilon} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{M+N}$. We observe that $\mathbb{V}[\boldsymbol{\epsilon}_{aug}] = \sigma^2\Sigma$, where Σ is defined in(37). Then, we introduce the solution $u^{\star,\sigma=0}_\xi$ to (11) for $\mathbf{y}_M = \mathbf{y}_M^{\text{true}}$. We further introduce the vectors of coefficients $\mathbf{u}^\star, \mathbf{u}^{\star,\sigma=0} \in \mathbb{R}^{M+N}$,

$$\mathbf{u}^\star = \begin{bmatrix} \boldsymbol{\eta}^\star \\ \boldsymbol{z}^\star \end{bmatrix}, \quad \mathbf{u}^{\star,\sigma=0} = \begin{bmatrix} \boldsymbol{\eta}^{\star,\sigma=0} \\ \boldsymbol{z}^{\star,\sigma=0} \end{bmatrix},$$

associated with $u^\star_\xi$ and $u^{\star,\sigma=0}_\xi$.

We have now the elements to prove Proposition 3.5.

Proof. (*Proposition 3.5*) We observe that

$$\|u^\star_\xi - u^{\star,\sigma=0}_\xi\|^2_{L^2(\Omega)} = (\mathbf{u}^\star - \mathbf{u}^{\star,\sigma=0})^T\,\mathbb{M}\,(\mathbf{u}^\star - \mathbf{u}^{\star,\sigma=0}) = \boldsymbol{\epsilon}^T_{aug}\left(\mathbb{A}_\xi^{-1}\,\mathbb{M}\,\mathbb{A}_\xi^{-1}\right)\boldsymbol{\epsilon}_{aug}.$$

Then, applying [46, Theorem C, Chapter 14.4], we find

$$\mathbb{E}\left[\|u_\xi^\star - u_\xi^{\star,\sigma=0}\|^2_{L^2(\Omega)}\right] \;=\; \sigma^2\,\mathrm{trace}\left(\mathbb{A}_\xi^{-1}\,\mathbb{M}\,\mathbb{A}_\xi^{-1}\,\Sigma\right). \tag{57}$$

We now distinguish two cases. If $u^{\mathrm{true}} \in \mathcal{Z}_N$, then $u_\xi^{\star,\sigma=0} = u^{\mathrm{true}}$ and (57) implies (39). On the other hand, if $u^{\mathrm{true}} \notin \mathcal{Z}_N$ thesis follows by observing that

$$\mathbb{E}\left[\|u_\xi^\star - u^{\mathrm{true}}\|^2_{L^2(\Omega)}\right] \le\; 2\,\|u^{\mathrm{true}} - u_\xi^{\star,\sigma=0}\|^2_{L^2(\Omega)} \;+\; 2\,\mathbb{E}\left[\|u_\xi^\star - u_\xi^{\star,\sigma=0}\|^2_{L^2(\Omega)}\right]$$

and then combining estimates (36) and (57). □

References

[1] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, pages 337–404, 1950.

[2] A. Bennett. Array design by inverse methods. *Progress in oceanography*, 15(2):129–156, 1985.

[3] A. Bennett and P. McIntosh. Open ocean modeling as an inverse problem: tidal theory. *Journal of Physical Oceanography*, 12(10):1004–1018, 1982.

[4] A. F. Bennett. *Inverse modeling of the ocean and atmosphere*. Cambridge University Press, 2002.

[5] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta numerica*, 14(1):1–137, 2005.

[6] G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual review of fluid mechanics*, 25(1):539–575, 1993.

[7] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM Journal on Mathematical Analysis*, 43(3):1457–1472, 2011.

[8] A. Buffa, Y. Maday, A. T. Patera, C. Prud homme, and G. Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(03):595–603, 2012.

[9] Y. Cao, J. Zhu, I. M. Navon, and Z. Luo. A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 53(10):1571–1583, 2007.

[10] G. Chardon, A. Cohen, and L. Daudet. Sampling and reconstruction of solutions to the helmholtz equation. *arXiv preprint arXiv:1301.0237*, 2013.

[11] F. Chinesta and E. Cueto. *PGD-based modeling of materials, structures and processes*. Springer, 2014.

[12] F. Chinesta, P. Ladeveze, and E. Cueto. A short review on model order reduction based on proper generalized decomposition. *Archives of Computational Methods in Engineering*, 18(4):395–404, 2011.

[13] A. Cohen and R. DeVore. Approximation of high-dimensional parametric pdes. *Acta Numerica*, 24:1–159, 5 2015.

[14] P. Courtier, J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994.

[15] R. DeVore, G. Petrova, and P. Wojtaszczyk. Greedy algorithms for reduced bases in banach spaces. *Constructive Approximation*, 37(3):455–466, 2013.

[16] L. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[17] R. Everson and L. Sirovich. Karhunen–loeve procedure for gappy data. *JOSA A*, 12(8):1657–1664, 1995.

[18] J. Fink and W. Rheinboldt. On the error behavior of the reduced basis technique for nonlinear finite element approximations. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 63(1):21–28, 1983.

[19] T. Gasser and H. G. Müller. Kernel estimation of regression functions. In *Smoothing techniques for curve estimation*, pages 23–68. Springer, 1979.

[20] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Science & Business Media, 2006.

[21] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning*, volume 2. Springer, 2009.

[22] J. S. Hesthaven, G. Rozza, and B. Stamm. Certified reduced basis methods for parametrized partial differential equations. *SpringerBriefs in Mathematics*, 2015.

[23] Y. Hon, R. Schaback, and X. Zhou. An adaptive greedy algorithm for solving large rbf collocation problems. *Numerical Algorithms*, 32(1):13–25, 2003.

[24] M. Kahlbacher and S. Volkwein. Galerkin proper orthogonal decomposition methods for parameter dependent elliptic systems. *Discussiones Mathematicae, Differential Inclusions, Control and Optimization*, 27(1):95–117, 2007.

[25] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

[26] G. Kimeldorf and G. Wahba. Some results on tchebycheffian spline functions. *Journal of mathematical analysis and applications*, 33(1):82–95, 1971.

[27] R. Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145, 1995.
[28] J. Krebs, A. Louis, and H. Wendland. Sobolev error estimates and a priori parameter selection for semi-discrete tikhonov regularization. *Journal of Inverse and Ill-Posed Problems*, 17(9):845–869, 2009.
[29] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal on Numerical analysis*, 40(2):492–515, 2002.
[30] Z. Li and I. Navon. Optimality of variational data assimilation and its relationship with the kalman filter and smoother. *Quarterly Journal of the Royal Meteorological Society*, 127(572):661–683, 2001.
[31] J. S. Lim. *Two-dimensional signal and image processing*. Prentice Hall, 1990.
[32] A. Lorenc. A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, 109(4):701–721, 1981.
[33] A. C. Lorenc. Analysis methods for numerical weather prediction. *Royal Meteorological Society, Quarterly Journal*, 112:1177–1194, 1986.
[34] Y. Maday and O. Mula. A generalized empirical interpolation method: Application of reduced basis techniques to data assimilation. In *Analysis and Numerics of Partial Differential Equations*, pages 221–235. Springer, 2013.
[35] Y. Maday, O. Mula, A. Patera, and M. Yano. The generalized empirical interpolation method: stability theory on hilbert spaces with an application to the stokes equation. *Computer Methods in Applied Mechanics and Engineering*, 287:310–334, 2015.
[36] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *International Journal for Numerical Methods in Engineering*, 2014.
[37] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. PBDW state estimation: Noisy observations; configuration-adaptive background spaces; physical interpretations. *ESAIM: Proceedings and Surveys*, 50:144–168, 2015.
[38] S. C. Malik and S. Arora. *Mathematical analysis*. New Age International, 1992.
[39] S. Müller. *Complexity and stability of kernel-based reconstructions*. PhD thesis, Dissertation, Georg-August-Universitttingen, Institut fr Numerische und Angewandte Mathematik, Lotzestrasse 16-18, D-37083 Gttingen, 2009.
[40] M. Perego, A. Veneziani, and C. Vergara. A variational approach for estimating the compliance of the cardiovascular tissue: An inverse fluid-structure interaction problem. *SIAM Journal on Scientific Computing*, 33(3):1181–1211, 2011.
[41] A. Pinkus. *N-widths in Approximation Theory*. Springer Science & Business Media, 1985.
[42] T. Poggio and C. Shelton. On the mathematical foundations of learning. *American Mathematical Society*, 39(1):1–49, 2002.
[43] C. Prud'homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80, 2002.
[44] C. Prud'homme, D. V. Rovas, K. Veroy, and A. T. Patera. A mathematical and computational framework for reliable real-time solution of parametrized partial differential equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 36(05):747–771, 2002.
[45] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations: An Introduction*, volume 92. Springer, 2015.
[46] J. Rice. *Mathematical statistics and data analysis*. Nelson Education, 2006.
[47] G. Rozza, D. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):229–275, 2008.
[48] R. Schaback and H. Wendland. Adaptive greedy techniques for approximate solution of large rbf systems. *Numerical Algorithms*, 24(3):239–254, 2000.
[49] R. Ştefănescu and I. M. Navon. Pod/deim nonlinear model order reduction of an adi implicit shallow water equations model. *Journal of Computational Physics*, 237:95–114, 2013.
[50] R. Ştefănescu, A. Sandu, and I. M. Navon. Pod/deim reduced-order strategies for efficient four dimensional variational data assimilation. *Journal of Computational Physics*, 295:569–595, 2015.
[51] T. Taddei, J. D. Penn, and A. T. Patera. Experimental a posteriori error estimation by monte carlo sampling of observation functionals. Technical report, MIT, 2016. submitted to International Journal for Numerical Methods in Engineering (August 2016).
[52] O. Talagrand. Assimilation of observations, an introduction. *Journal meteorological society of Japan Series 2*, 75:81–99, 1997.
[53] V. Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.
[54] P. Vermeulen and A. Heemink. Model-reduced variational data assimilation. *Monthly weather review*, 134(10):2888–2899, 2006.
[55] G. Wahba. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 364–372, 1978.
[56] G. Wahba. *Spline models for observational data*, volume 59. Siam, 1990.
[57] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in computational Mathematics*, 4(1):389–396, 1995.

[58] H. Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.
[59] K. Willcox. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers & fluids*, 35(2):208–226, 2006.
[60] E. G. Williams. *Fourier acoustics: sound radiation and nearfield acoustical holography*. Academic press, 1999.
[61] D. Wirtz and B. Haasdonk. A vectorial kernel orthogonal greedy algorithm. *Proceedings of DWCAA12*, 6:83–100, 2013.
[62] D. Xiao, F. Fang, A. G. Buchan, C. C. Pain, I. M. Navon, J. Du, and G. Hu. Non-linear model reduction for the navier–stokes equations using residual deim method. *Journal of Computational Physics*, 263:1–18, 2014.