SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 10, No. 1, pp. 227–267

Nonlinear Reduced Models for State and Parameter Estimation*

Albert Cohen†, Wolfgang Dahmen‡, Olga Mula§, and James Nichols¶

Abstract. State estimation aims at approximately reconstructing the solution u to a parametrized partial differential equation from m linear measurements when the parameter vector y is unknown. Fast numerical recovery methods have been proposed in Maday et al. [*Internat. J. Numer. Methods Engrg.*, 102 (2015), pp. 933–965] based on reduced models which are linear spaces of moderate dimension n that are tailored to approximate the solution manifold $\mathcal{M}$ where the solution sits. These methods can be viewed as deterministic counterparts to Bayesian estimation approaches and are proved to be optimal when the prior is expressed by approximability of the solution with respect to the reduced model [P. Binev et al., *SIAM/ASA J. Uncertain. Quantif.*, 5 (2017), pp. 1–29]. However, they are inherently limited by their linear nature, which bounds from below their best possible performance by the Kolmogorov width $d_m(\mathcal{M})$ of the solution manifold. In this paper, we propose to break this barrier by using simple nonlinear reduced models that consist of a finite union of linear spaces V_k, each having dimension at most m and leading to different estimators u_k^*. A model selection mechanism based on minimizing the PDE residual over the parameter space is used to select from this collection the final estimator u^*. Our analysis shows that u^* meets optimal recovery benchmarks that are inherent to the solution manifold and not tied to its Kolmogorov width. The residual minimization procedure is computationally simple in the relevant case of affine parameter dependence in the PDE. In addition, it results in an estimator y^* for the unknown parameter vector. In this setting, we also discuss an alternating minimization (coordinate descent) algorithm for joint state and parameter estimation that potentially improves the quality of both estimators.

Key words. state estimation, parameter estimation, reduced order modeling, optimal recovery

AMS subject classifications. 65M32, 65M12

DOI. 10.1137/20M1380818

1. Introduction.

1.1. Parametrized PDEs and inverse problems.

Parametrized partial differential equations are of common use to model complex physical systems. Such equations can generally be written in abstract form as

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

*Received by the editors November 17, 2020; accepted for publication (in revised form) November 15, 2021; published electronically February 8, 2022.
https://doi.org/10.1137/20M1380818
Funding: This work was supported by ERC Advanced grant BREAD; the Emergence Project of the Paris City Council "Models and Measures"; NSF grants DMS-1720297 and DMS-2012469; and the SmartState and Williams-Hedberg Foundation.

†Laboratoire Jacques-Louis Lions, Universite Pierre et Marie Curie, Paris, 75005 France (cohen@ann.jussieu.fr).
‡Department of Mathematics, University of South Carolina, Columbia, SC 29208 USA (wolfgang.anton.dahmen@googlemail.com).
§CEREMADE, Université Paris Dauphine-PSL, Paris, 75775 France (mula@ceremade.dauphine.fr).
¶Biological Data Science Institute, College of Science, Australian National University, Camberra, ACT 2601, Australia (james.nichols@anu.edu.au).

where $y = (y_1, \dots, y_d)$ is a vector of scalar parameters ranging in some domain $Y \subset \mathbb{R}^d$. We assume well-posedness; that is, for any $y \in Y$ the problem admits a unique solution $u = u(y)$ in some Hilbert space V. We may therefore consider the *parameter-to-solution map*

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

from Y to V, which is typically nonlinear, as well as the *solution manifold*

$$\mathcal{M} := \{u(y) \,:\, y \in Y\} \subset V \tag{1.3}$$

that describes the collection of all admissible solutions. Throughout this paper, we assume that Y is compact in $\mathbb{R}^d$ and that the map (1.2) is continuous. Therefore, $\mathcal{M}$ is a compact set of V. We sometimes refer to the solution $u(y)$ as the *state* of the system for the given parameter vector y.

The parameters are used to represent physical quantities such as diffusivity, viscosity, velocity, source terms, or the geometry of the physical domain in which the PDE is posed. In several relevant instances, y may be high or even countably infinite dimensional, that is, $d \gg 1$ or $d = \infty$.

In this paper, we are interested in *inverse problems* which occur when only a vector of *linear* measurements

$$z = (z_1, \dots, z_m) \in \mathbb{R}^m, \quad z_i = \ell_i(u), \quad i = 1, \dots, m, \tag{1.4}$$

is observed, where each $\ell_i \in V'$ is a known continuous linear functional on V. We also sometimes use the notation

$$z = \ell(u), \quad \ell = (\ell_1, \dots, \ell_m). \tag{1.5}$$

One wishes to recover from z the unknown state $u \in \mathcal{M}$ or even the underlying parameter vector $y \in Y$ for which $u = u(y)$. Therefore, in an idealized setting, one partially observes the result of the composition map

$$y \in Y \mapsto u \in \mathcal{M} \mapsto z \in \mathbb{R}^m \tag{1.6}$$

for the unknown y. More realistically, the measurements may be affected by additive noise

$$z_i = \ell_i(u) + \eta_i, \tag{1.7}$$

and the model itself might be biased, meaning that the true state u deviates from the solution manifold $\mathcal{M}$ by some amount. Thus, two types of inverse problems may be considered:

(i) State estimation: recover an approximation u^* of the state u from the observation $z = \ell(u)$. This is a linear inverse problem, in which the prior information on u is given by the manifold $\mathcal{M}$ which has a complex geometry and is not explicitly known.

(ii) Parameter estimation: recover an approximation y^* of the parameter y from the observation $z = \ell(u)$ when $u = u(y)$. This is a nonlinear inverse problem, for which the prior information available on y is given by the domain Y.

These problems become severely ill posed when Y has dimension $d > m$. For this reason, they are often addressed through Bayesian approaches [31, 39]: a prior probability distribution P_y being assumed on $y \in Y$ (thus inducing a push forward distribution P_u for $u \in \mathcal{M}$), the objective is to understand the *posterior* distributions of y or u conditioned by the observations z in order to compute plausible solutions y^* or u^* under such probabilistic priors. The accuracy of these solutions should therefore be assessed in some average sense.

In this paper, we do not follow this avenue: the only priors made on y and u are their membership to Y and $\mathcal{M}$. We are interested in developing practical estimation methods that offer uniform recovery guarantees under such deterministic priors in the form of upper bounds on the worst case error for the estimators over all $y \in Y$ or $u \in \mathcal{M}$. We also aim to understand whether our error bounds are optimal in some sense. Our primary focus will actually be on state estimation (i). Nevertheless, we present in section 4 several implications on parameter estimation (ii), which to our knowledge are new. For state estimation, error bounds have recently been established for a class of methods based on *linear reduced modeling*, as we recall next.

1.2. Reduced models: The PBDW method. In several relevant instances, the particular parametrized PDE structure allows one to construct linear spaces V_n of moderate dimension n that are specifically tailored to the approximation of the solution manifold $\mathcal{M}$, in the sense that

$$\operatorname{dist}(\mathcal{M}, V_n) = \max_{u \in \mathcal{M}} \min_{v \in V_n} \|u - v\| \leqslant \varepsilon_n, \tag{1.8}$$

where ε_n is a certified bound that decays with n significantly faster than when using for V_n classical approximation spaces of dimension n such as finite elements, algebraic or trigonometric polynomials, or spline functions. Throughout this paper,

$$\langle \cdot, \cdot \rangle_V^{1/2} =: \| \cdot \| = \| \cdot \|_V \tag{1.9}$$

denotes the norm of the Hilbert space V. The natural benchmark for such approximation spaces is the Kolmogorov n-width

$$d_n(\mathcal{M}) := \min_{\dim(E)=n} \operatorname{dist}(\mathcal{M}, E). \tag{1.10}$$

The space E_n that achieves the above minimum is thus the best possible reduced model for approximating all of $\mathcal{M}$; however, it is computationally out of reach.

One instance of computational reduced model spaces is generated by *sparse polynomial approximations* of the form

$$u_n(y) = \sum_{\nu \in \Lambda_n} u_\nu y^\nu, \quad y^\nu := \prod_{j \geqslant 1} y_j^{\nu_j}, \tag{1.11}$$

where Λ_n is a conveniently chosen set of multi-indices such that $\#(\Lambda_n) = n$. Such approximations can be derived, for example, by best n-term truncations of infinite Taylor or orthogonal polynomial expansions. We refer the reader to [13, 14], where convergence estimates of the form

$$\sup_{y \in Y} \|u(y) - u_n(y)\| \leqslant C n^{-s} \tag{1.12}$$

are established for some $s > 0$ even when $d = \infty$. Therefore, the space $V_n := \operatorname{span}\{u_\nu : \nu \in \Lambda_n\}$ approximates the solution manifold with accuracy $\varepsilon_n = Cn^{-s}$. In the particular approximation $u_n(y)$ of $u(y)$ for a given y, the parametric monomials y^ν are the scalar coefficients associated to the generators $u_\nu \in V$.

Another instance, known as *reduced basis approximation*, consists in using spaces of the form

$$V_n := \operatorname{span}\{u_1, \dots, u_n\}, \tag{1.13}$$

where $u_i = u(y^i) \in \mathcal{M}$ are instances of solutions corresponding to a particular selection of parameter values $y^i \in Y$ (see [33, 37, 38]). One typical selection procedure is based on a greedy algorithm: one picks y^k such that $u_k = u(y^k)$ is furthest away from the previously constructed space V_{k-1}, in the sense of maximizing a computable and tight a posteriori bound of the projection error $\|u(y) - P_{V_{k-1}} u(y)\|$ over a sufficiently fine discrete training set $\tilde{Y} \subset Y$. In turn, this method also delivers a computable upper estimate ε_k for $\operatorname{dist}(\mathcal{M}, V_k)$. It was proved in [3, 30] that the reduced basis spaces resulting from this greedy algorithm have near-optimal approximation property, in the sense that if $d_n(\mathcal{M})$ has a certain polynomial or exponential rate of decay as $n \to \infty$, then the same rate is achieved by $\operatorname{dist}(\mathcal{M}, V_n)$.

In both cases, these reduced models come in the form of a hierarchy $(V_n)_{n \geqslant 1}$, with computable decreasing error bounds $(\varepsilon_n)_{n \geqslant 1}$, where n corresponds to the level of truncation in the first case and the step of the greedy algorithm in the second case. Given a reduced model V_n, one way of tackling the state estimation problem is to replace the complex solution manifold $\mathcal{M}$ by the simpler prior class described by the cylinder

$$\mathcal{K} = \mathcal{K}(V_n, \varepsilon_n) = \{v \in V \ : \ \operatorname{dist}(v, V_n) \leqslant \varepsilon_n\} \tag{1.14}$$

that contains $\mathcal{M}$. The set $\mathcal{K}$ therefore reflects the approximability of $\mathcal{M}$ by V_n. This point of view leads to the *parametrized background data weak* (PBDW) method introduced in [1], also called the *one space method* and further analyzed in [2], that we recall below in a nutshell.

In the noiseless case, the knowledge of $z = (z_i)_{i=1,\dots,m}$ is equivalent to that of the orthogonal projection $w = P_W u$, where

$$W := \operatorname{span}\{\omega_1, \dots, \omega_m\} \tag{1.15}$$

and $\omega_i \in V$ are the Riesz representers of the linear functionals ℓ_i, that is,

$$\ell_i(v) = \langle \omega_i, v \rangle, \quad v \in V. \tag{1.16}$$

Thus, the data indicates that u belongs to the affine space $w + W^\perp$.

Combining this information with the prior class $\mathcal{K}$, the unknown state thus belongs to the ellipsoid

$$\mathcal{K}_w := \mathcal{K} \cap (w + W^\perp) = \{v \in \mathcal{K} \ : \ P_W v = w\}. \tag{1.17}$$

For this posterior class $\mathcal{K}_w$, the *optimal recovery* estimator u^* that minimizes the worst case error $\max_{u \in \mathcal{K}_w} \|u - u^*\|$ is therefore the center of the ellipsoid. It is proven in [2] that this center is equivalently given by

$$u^* = u^*(w) := \operatorname{argmin}\{\|v - P_{V_n} v\| \ : \ P_W v = w\}. \tag{1.18}$$

It can be computed from the data w in an elementary manner by solving a finite set of linear equations. The worst case performance for this estimator, both over $\mathcal{K}$ and $\mathcal{K}_w$, for any w, is thus given by the half-diameter of the ellipsoid which is the product of the width ε_n of $\mathcal{K}$ and the quantity

$$\mu_n = \mu(V_n, W) := \max_{v \in V_n} \frac{\|v\|}{\|P_W v\|}. \tag{1.19}$$

Note that μ_n is the inverse of the cosine of the angle between V_n and W. For $n \geqslant 1$, this quantity can be computed as the inverse of the smallest singular value of the $n \times m$ cross-Gramian matrix with entries $\langle \phi_i, \psi_j \rangle$ between any pair of orthonormal bases $(\phi_i)_{i=1,\dots,n}$ and $(\psi_j)_{j=1,\dots,m}$ of V_n and W, respectively. It is readily seen that one also has

$$\mu_n = \max_{w \in W^\perp} \frac{\|w\|}{\|P_{V_n^\perp} w\|}, \tag{1.20}$$

allowing us to extend the above definition to the case of the zero-dimensional space $V_n = \{0\}$ for which $\mu(\{0\}, W) = 1$.

Since $\mathcal{M} \subset \mathcal{K}$, the worst case error bound over $\mathcal{M}$ of the estimator, defined as

$$E_{wc} := \max_{u \in \mathcal{M}} \|u - u^*(P_W u)\|, \tag{1.21}$$

satisfies the error bound

$$E_{wc} \leqslant \max_{u \in \mathcal{K}} \|u - u^*(P_W u)\| = \mu_n \varepsilon_n. \tag{1.22}$$

Remark 1.1. *The estimation map $w \mapsto u^*(w)$ is linear with norm μ_n and does not depend on ε_n. It thus satisfies, for any individual $u \in V$ and $\eta \in W$,*

$$\|u - u^*(P_W u + \eta)\| \leqslant \mu_n(\mathrm{dist}(u, V_n) + \|\eta\|). \tag{1.23}$$

We may therefore account for an additional measurement noise and model bias: if the observation is $w = P_W u + \eta$ with $\|\eta\| \leqslant \varepsilon_{noise}$, and if the true states do not lie in $\mathcal{M}$ but satisfy $\mathrm{dist}(u, \mathcal{M}) \leqslant \varepsilon_{model}$, the guaranteed error bound (1.22) should be modified into

$$\|u - u^*(w)\| \leqslant \mu_n(\varepsilon_n + \varepsilon_{noise} + \varepsilon_{model}). \tag{1.24}$$

In practice, the noise component $\eta \in W$ typically results from a noise vector $\overline{\eta} \in \mathbb{R}^m$ affecting the observation z according to $z = \ell(u) + \overline{\eta}$. Assuming a bound $\|\overline{\eta}\|_2 \leqslant \overline{\varepsilon}_{noise}$ where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^m$, we thus receive the above error bound with $\varepsilon_{noise} := \|M\|\overline{\varepsilon}_{noise}$, where $M \in \mathbb{R}^{m \times m}$ is the matrix that transforms the representer basis $\omega = \{\omega_1, \dots, \omega_m\}$ into an orthonormal basis $\psi = \{\psi_1, \dots, \psi_m\}$ of W. Here estimation accuracy benefits from decreasing noise without increasing computational cost. This is in contrast to Bayesian methods, for which small noise level induces computational difficulties due to the concentration of the posterior distribution.

Remark 1.2. *To bring out the essential mechanisms, we have idealized (and technically simplified) the description of the PBDW method by omitting certain discretization aspects that are unavoidable in computational practice and should be accounted for. To start with, the snapshots u_i (or the polynomial coefficients u_ν) that span the reduced basis spaces V_n cannot be computed exactly but only up to some tolerance by a numerical solver. One typical instance is the finite element method, which yields an approximate parameter-to-solution map*

$$y \mapsto u_h(y) \in V_h, \tag{1.25}$$

where V_h is a reference finite element space ensuring a prescribed accuracy

$$\|u(y) - u_h(y)\| \leqslant \varepsilon_h, \quad y \in Y. \tag{1.26}$$

The computable states are therefore elements of the perturbed manifold

$$\mathcal{M}_h := \{u_h(y) \,:\, y \in Y\}. \tag{1.27}$$

The reduced model spaces V_n are low-dimensional subspaces of V_h and with certified accuracy

$$\operatorname{dist}(\mathcal{M}_h, V_n) \leqslant \varepsilon_n. \tag{1.28}$$

The true states do not belong to $\mathcal{M}_h$, and this deviation can therefore be interpreted as a model bias in the sense of the previous remark with $\varepsilon_{model} = \varepsilon_h$. The application of the PDBW also requires the introduction of the Riesz lifts ω_i in order to define the measurement space W. Since we operate in the space V_h, these can be defined as elements of this space satisfying

$$\langle \omega_i, v \rangle_V = \ell_i(v), \quad v \in V_h, \tag{1.29}$$

thus resulting in a measurement space $W \subset V_h$. For example, if V is the Sobolev spaces $H^1_0(D)$ for some domain D and V_h is a finite element subspace, the Riesz lifts are the unique solutions to the Galerkin problem

$$\int_D \nabla \omega_i \nabla v = \ell_i(v), \quad v \in V_h, \tag{1.30}$$

and can be identified by solving $n_h \times n_h$ linear systems. Measuring accuracy in V, i.e., in a metric dictated by the continuous PDE model, the idealization, to be largely maintained in what follows, also helps understanding how to properly adapt the background-discretization V_h to the overall achievable estimation accuracy. Other computational issues involving the space V_h will be discussed in section 3.4.

Note that $\mu_n \geqslant 1$ increases with n and that its finiteness imposes that $\dim(V_n) \leqslant \dim(W)$, that is, $m \geqslant n$. Therefore, one natural way to decide which space V_n to use is to take the value of $n \in \{0, \dots, m\}$ that minimizes the bound $\mu_n \varepsilon_n$. This choice is somehow crude since it might not be the value of n that minimizes the true reconstruction error for a given $u \in \mathcal{M}$, and for this reason it was referred to as a *poor man algorithm* in [3].

The PBDW approach to state estimation can be improved in various ways:

- One variant that is relevant to the present work is studied in [12] and consists in using reduced models of affine form

$$V_n = \overline{u}_n + \overline{V}_n, \tag{1.31}$$

where $\overline{V}_n$ is a linear space and $\overline{u}$ is a given offset. The optimal recovery estimator is again defined by the minimization property (1.18). Its computation amounts to the same type of linear systems, and the reconstruction map $w \mapsto u^*(w)$ is now affine. The error bound (1.22) remains valid with $\mu_n = \mu(\overline{V}_n, W)$ and ε_n a bound for $\operatorname{dist}(\mathcal{M}, V_n)$. Note that ε_n is also a bound for the distance of $\mathcal{M}$ to the linear space $\overline{V}_{n+1} := \overline{V}_n \oplus \mathbb{R}\overline{u}_n$ of dimension $n+1$. However, using instead this linear space could result in a stability constant $\mu_{n+1} = \mu(\overline{V}_{n+1}, W)$ that is much larger than μ_n, in particular when the offset $\overline{u}_n$ is close to $W^\perp$.
- Another variant proposed in [12] consists in using a large set $\mathcal{T}_N = \{u_i = u(y^i) : i = 1, \dots, N\}$ of precomputed solutions in order to train the reconstruction map $w \mapsto u^*(w)$ by minimizing the least-squares fit $\sum_{i=1}^N \|u_i - u^*(P_W u_i)\|^2$ over all linear or affine maps, which amounts to optimizing the choice of the space V_n in the PBDW method.
- Conversely, for a given reduced basis space V_n, it is also possible to optimize the choice of linear functionals $(\ell_1, \dots, \ell_m)$ giving rise to the data, among a dictionary $\mathcal{D}$, that represent a set of admissible measurement devices. The objective is to minimize the stability constant $\mu(V_n, W)$ for the resulting space W; see, in particular, [4], where a greedy algorithm is proposed for selecting the ℓ_i. We do not take this view in the present paper and think of the space W as fixed once and for all: the measurement devices are given to us and cannot be modified.

1.3. Objective and outline. The simplicity of the PBDW method and its above variants come together with a fundamental limitation of its performance: since the map $w \mapsto u^*(w)$ is linear or affine, the reconstruction necessarily belongs to an m- or $(m+1)$-dimensional space, and therefore the worst case performance is necessarily bounded from below by the Kolmogorov width $d_m(\mathcal{M})$ or $d_{m+1}(\mathcal{M})$. In view of this limitation, one principal objective of the present work is to develop *nonlinear* state estimation techniques which provably overcome the bottleneck of the Kolmogorov width $d_m(\mathcal{M})$.

In section 2, we introduce various benchmark quantities that describe the best possible performance of a recovery map in a worst case sense. We first consider an idealized setting where the state u is assumed to exactly satisfy the theoretical model described by the parametric PDE, that is, $u \in \mathcal{M}$. Then we introduce similar benchmarks quantities in the presence of model bias and measurement noise. All these quantities can be substantially smaller than $d_m(\mathcal{M})$.

In section 2, we discuss a nonlinear recovery method, based on a family of affine reduced models $(V_k)_{k=1,\dots,K}$, where each V_k has dimension $n_k \leqslant m$ and serves as a local approximation to a portion $\mathcal{M}_k$ of the solution manifold. Applying the PBDW method with each such space results in a collection of state estimators u_k^*. The value k for which the true state u belongs to $\mathcal{M}_k$ being unknown, we introduce a *model selection* procedure in order to pick a value k^* and define the resulting estimator $u^* = u_{k^*}^*$. We show that this estimator has performance comparable to the benchmark introduced in section 2. Such performances cannot be achieved

by the standard PBDW method due to the above described limitations.

Model selection is a classical topic of mathematical statistics [35], with representative techniques such as complexity penalization or cross-validation in which the data are used to select a proper model. Our approach differs from these techniques in that it exploits (in the spirit of *data assimilation*) the PDE model which is available to us by evaluating the distance to the manifold

$$\operatorname{dist}(v,\mathcal{M}) = \min_{y\in Y} \|v - u(y)\| \tag{1.32}$$

of the different estimators $v = u_k^*$ for $k = 1,\dots,K$ and picking the value k^* that minimizes it. In practice, the quantity (1.32) cannot be exactly computed, and we instead rely on a computable surrogate quantity $\mathcal{S}(v,\mathcal{M})$ expressed in terms of the residual to the PDE; see section 3.4.

One typical instance where such a surrogate is available is when (1.1) has the form of a linear operator equation

$$A(y)u = f(y), \tag{1.33}$$

where $A(y)$ is boundedly invertible from V to V', or more generally from $V \to Z'$, for a test space Z different from V, uniformly over $y \in Y$. Then $\mathcal{S}(v,\mathcal{M})$ is obtained by minimizing the residual

$$\mathcal{R}(v,y) = \|A(y)v - f(y)\|_{Z'} \tag{1.34}$$

over $y \in Y$. This task itself is greatly facilitated in the case where the operators $A(y)$ and source terms $f(y)$ have affine dependence in y and when the parameter domain Y is convex. One relevant example that has been often considered in the literature is the second order homogeneous boundary value problem with affine diffusion coefficient,

$$-\operatorname{div}(a\nabla u) = f(y) \quad \text{in } D,\ u|_{\partial D} = 0, \quad a = a(y) = \overline{a} + \sum_{j=1}^{d} y_j\psi_j, \tag{1.35}$$

where in this case $V = Z = H_0^1(D)$.

In section 4, we discuss the more direct approach for both state and parameter estimation based on minimizing $\mathcal{R}(v,y)$ over both $y \in Y$ and $v \in w + W^{\perp}$. The associated alternating minimization algorithm amounts to a simple succession of quadratic problems in the particular case of linear PDEs with affine parameter dependence. Such an algorithm is not guaranteed to converge to a global minimum (since the residual is not globally convex), and for this reason its limit may miss the optimality benchmark. On the other hand, using the estimator derived in section 3 as a "good initialization point" to this minimization algorithm leads to a limit state that has at least the same order of accuracy.

These various approaches are numerically tested in section 5 for the elliptic equation (1.35) for both the overdetermined regime $m \geqslant d$ and the underdetermined regime $m < d$.

1.4. Connections of the present work with other inverse problem approaches and model order reduction. One of the main results of the present paper is the development of a state estimation algorithm which has close to optimal reconstruction properties over the solution manifold, in the sense that we introduce in section 2. Reduced order models V_n play a prominent role since they are the main vehicle for making our strategy implementable in practice. We should note that the idea of using reduced models to solve inverse problems is actually not new in the literature, though. It can be traced back at least to the gappy POD method, first introduced in [17] by Everson and Sirovich. There, the authors address the problem of restoring a full image from partial pixel observations by using a least-squares strategy involving a reconstruction on linear spaces obtained by PCA. The same strategy was then brought to other fields, such as fluid and structural applications; see [18]. The introduction of a reduced model can be seen as an improvement with respect to working with one single background function, as is done in methods such as 3D-VAR; see [15, 16]. In contrast to the present work and the PBDW method in general, the gappy POD method is formulated on the Euclidean space $V = \mathbb{R}^{\mathcal{N}}$, with $\mathcal{N} \in \mathbb{N}$ typically much larger than m and n. It uses linear reduced models V_n obtained by PCA, and measurement observations are typically pointwise vector entries, that is, $\omega_i = e_i$, with $e_i \in \mathbb{R}^{\mathcal{N}}$ being the ith unit vector. For that particular choice of ambient space and reduced models, the linear PBDW method is very close to gappy POD. It is, however, not entirely equivalent since PBDW presents a certain component in $W \cap V_n^{\perp}$ which is missing in gappy POD. For the case of a general Hilbert space, linear PBDW is equivalent to the generalized empirical interpolation method when $m = n$; see [20, 21, 22]. It is also interesting to note that the reconstruction algorithm (1.18) corresponding to what we call linear PBDW in this paper has also been developed in [19] as an extension of compressed sensing to Hilbert spaces. There, the envisaged spaces V_n are related to Fourier and wavelet spaces rather than reduced models of parametric PDEs. Finally, reduced models are commonly used in order to significantly speed up forward simulations that could be needed in inversion tasks; see, e.g., [42], where a central issue is to judiciously switch between the high fidelity model, given in terms of a fine scale discretization, and the low fidelity reduced model.

In the above landscape of methods, our proposed piecewise affine extension of PBDW can be interpreted as a further generalization step which comes with optimal reconstruction guarantees. Our strategy is based on an offline partitioning of the manifold $\mathcal{M}$ in which, for each element of the partition, we compute reduced models. We then decide with a data-driven approach which reduced model is the most appropriate for the reconstruction. The idea of partitioning the manifold and working with different reduced order models for each partition is not new, but it has mostly been explored in works that focus on the forward problem of approximating the parameter-to-solution mapping $y \in Y \mapsto u \in \mathcal{M}$; see, e.g., [24, 25, 26, 27]. This strategy enters into the general topic of nonlinear forward model reduction for which little is known in terms of the performance guarantees. A first step towards a cohesive theory for nonlinear forward model reduction has recently been proposed in [23], in relation with the general concept of library widths [41].

2. Optimal recovery benchmarks. In this section, we describe the performance of the best possible recovery map

$$w \mapsto u^*(w) \tag{2.1}$$

in terms of its worst case error. We consider first the case of noiseless data and no model bias. In a subsequent step, we take such perturbations into account. While these best recovery maps cannot be implemented by a simple algorithm, their performance serves as the benchmark for the nonlinear state estimation algorithms discussed in the next section.

2.1. Optimal recovery for the solution manifold. In the absence of model bias and when a noiseless measurement $w = P_W u$ is given, our knowledge of u is that it belongs to the set

$$\mathcal{M}_w := \mathcal{M} \cap (w + W^\perp). \tag{2.2}$$

The best possible recovery map can be described through the following general notion.

Definition 2.1. *The Chebyshev ball of a bounded set $S \in V$ is the closed ball $B(v,r)$ of minimal radius that contains S. One denotes by $v = \mathrm{cen}(S)$ the Chebyshev center of S and $r = \mathrm{rad}(S)$ its Chebyshev radius.*

In particular, by Theorem 9 in [36], one has

$$\frac{1}{2}\mathrm{diam}(S) \leqslant \mathrm{rad}(S) \leqslant \frac{1}{\sqrt{2}}\mathrm{diam}(S), \tag{2.3}$$

where $\mathrm{diam}(S) := \sup\{\|u - v\| \,:\, u, v \in S\}$ is the diameter of S. Therefore, the recovery map that minimizes the worst case error over $\mathcal{M}_w$ for any given w, and therefore over $\mathcal{M}$, is defined by

$$u^*(w) = \mathrm{cen}(\mathcal{M}_w). \tag{2.4}$$

Its worst case error is

$$E^*_{\mathrm{wc}} = \sup\{\mathrm{rad}(\mathcal{M}_w) \,:\, w \in W\}. \tag{2.5}$$

In view of the equivalence (2.3), we can relate E^*_{wc} to the quantity

$$\delta_0 = \delta_0(\mathcal{M}, W) := \sup\{\mathrm{diam}(\mathcal{M}_w) \,:\, w \in W\} = \sup\{\|u - v\| \,:\, u, v \in \mathcal{M},\ u - v \in W^\perp\} \tag{2.6}$$

by the equivalence

$$\frac{1}{2}\delta_0 \leqslant E^*_{\mathrm{wc}} \leqslant \frac{1}{\sqrt{2}}\delta_0. \tag{2.7}$$

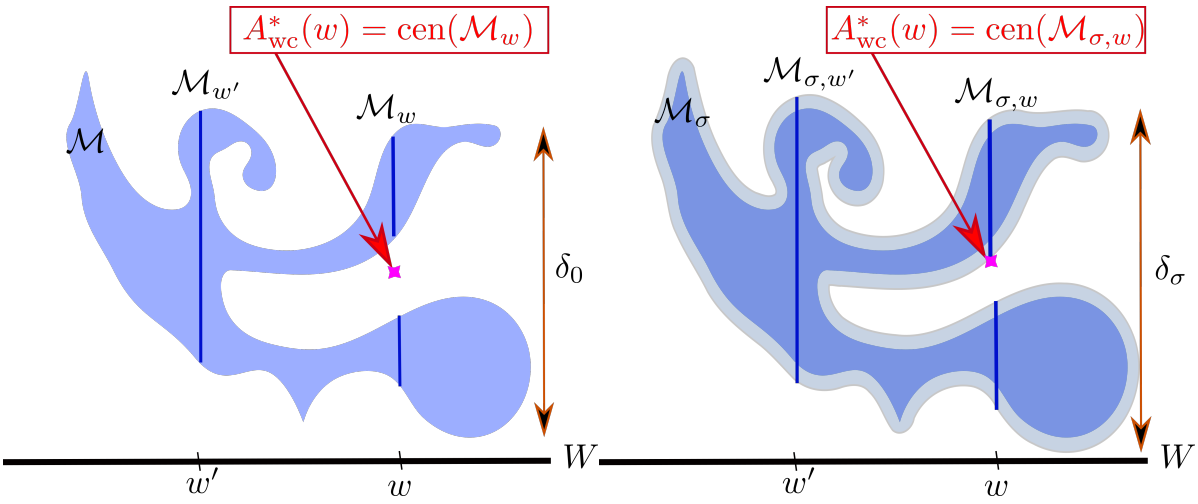
Note that injectivity of the measurement map P_W over $\mathcal{M}$ is equivalent to $\delta_0 = 0$. We provide in Figure 2.1a an illustration of the above benchmark concepts.

If $w = P_W u$ for some $u \in \mathcal{M}$, then any $u^* \in \mathcal{M}$ such that $P_W u^* = w$ meets the ideal benchmark $\|u - u^*\| \leqslant \delta_0$. Therefore, one way for finding such a u^* would be to minimize the distance to the manifold over all functions such that $P_W v = w$, that is, solve

$$\min_{v \in w + W^\perp} \mathrm{dist}(v, \mathcal{M}) = \min_{v \in w + W^\perp} \min_{y \in Y} \|u(y) - v\|. \tag{2.8}$$

This problem is computationally out of reach since it amounts to the nested minimization of two nonconvex functions in high dimension.

Computationally feasible algorithms such as the PBDW methods are based on a simplification of the manifold $\mathcal{M}$ which induces an approximation error. We introduce next a somewhat relaxed benchmark that takes this error into account.



(a) Perfect model (b) Model bias

Figure 2.1. *Illustration of the optimal recovery benchmark on a manifold in the two-dimensional Euclidean space.*

2.2. Optimal recovery under perturbations. In order to account for manifold simplification as well as model bias, for any given accuracy $\sigma > 0$, we introduce the σ-offset of $\mathcal{M}$,

$$\mathcal{M}_\sigma := \{v \in V \ : \ \mathrm{dist}(v, \mathcal{M}) \leqslant \sigma\} = \bigcup_{u \in \mathcal{M}} B(u, \sigma). \tag{2.9}$$

Likewise, we introduce the perturbed set

$$\mathcal{M}_{\sigma,w} = \mathcal{M}_\sigma \cap (w + W^\perp), \tag{2.10}$$

which, however, still excludes uncertainties in w. Our benchmark for the worst case error is now defined as (see Figure 2.1b for an illustration)

$$\delta_\sigma := \max_{w \in W} \mathrm{diam}(\mathcal{M}_{\sigma,w}) = \max\{\|u - v\| \ : \ u, v \in \mathcal{M}_\sigma, \ u - v \in W^\perp\}. \tag{2.11}$$

The map $\sigma \mapsto \delta_\sigma$ satisfies some elementary properties:

- Monotonicity and continuity. It is obviously nondecreasing:

$$\sigma \leqslant \tilde{\sigma} \implies \delta_\sigma \leqslant \delta_{\tilde{\sigma}}. \tag{2.12}$$

Simple finite-dimensional examples show that this map may have jump discontinuities. Take, for example, a compact set $\mathcal{M} \subset \mathbb{R}^2$ consisting of the two points $(0,0)$ and $(1/2,1)$, and $W = \mathbb{R}e_1$, where $e_1 = (1,0)$. Then $\delta_\sigma = 2\sigma$ for $0 \leqslant \sigma < \frac{1}{4}$, while $\delta_{\frac{1}{4}}(\mathcal{M}, W) = 1$. Using the compactness of $\mathcal{M}$, it is possible to check that $\sigma \mapsto \delta_\sigma$ is continuous from the right and in particular $\lim_{\sigma \to 0} \delta_\sigma(\mathcal{M}, W) = \delta_0$.

- Bounds from below and above: for any $u, v \in \mathcal{M}_{\sigma,w}$, and for any $\tilde{\sigma} \geqslant 0$, let $\tilde{u} = u + \tilde{\sigma} g$ and $\tilde{v} = v - \tilde{\sigma} g$ with $g = (u-v)/\|u-v\|$. Then $\|\tilde{u} - \tilde{v}\| = \|u-v\| + 2\tilde{\sigma}$ and $\tilde{u} - \tilde{v} \in W^\perp$, which shows that $\tilde{u}$, $\tilde{v} \in \mathcal{M}_{\sigma+\tilde{\sigma},w}$, and

$$\delta_{\sigma+\tilde{\sigma}} \geqslant \delta_\sigma + 2\tilde{\sigma}. \tag{2.13}$$

  In particular,

$$\delta_\sigma \geqslant \delta_0 + 2\sigma \geqslant 2\sigma. \tag{2.14}$$

  On the other hand, we obviously have the upper bound $\delta_\sigma \leqslant \operatorname{diam}(\mathcal{M}_\sigma) \leqslant \operatorname{diam}(\mathcal{M}) + 2\sigma$.
- The quantity

$$\mu(\mathcal{M}, W) := \frac{1}{2} \sup_{\sigma > 0} \frac{\delta_\sigma - \delta_0}{\sigma} \tag{2.15}$$

  may be viewed as a general stability constant inherent to the recovery problem, similar to $\mu(V_n, W)$, which is more specific to the particular PBDW method: in the special case where $\mathcal{M} = V_n$ and $V_n \cap W^\perp = \{0\}$, one has $\delta_0 = 0$ and $\frac{\delta_\sigma}{2\sigma} = \mu(V_n, W)$ for all $\sigma > 0$. Note that $\mu(\mathcal{M}, W) \geqslant 1$ in view of (2.14).

Regarding measurement noise, it suggests to introduce the quantity

$$\tilde{\delta}_\sigma := \max\{\|u - v\| \ : \ u, v \in \mathcal{M}, \ \|P_W u - P_W v\| \leqslant \sigma\}. \tag{2.16}$$

Thus, $\tilde{\delta}_\sigma$ accounts for a noise level σ in the measurement space W, which, by Remark 1.1, relates to the noise in the data measured in the Euclidean norm on $\mathbb{R}^m$.

The two quantities δ_σ and $\tilde{\delta}_\sigma$ are not equivalent; however, one has the framing

$$\delta_\sigma - 2\sigma \leqslant \tilde{\delta}_{2\sigma} \leqslant \delta_\sigma + 2\sigma. \tag{2.17}$$

To prove (2.17), we note that for any $u, v \in \mathcal{M}_\sigma$ such that $u - v \in W^\perp$, there exist $\tilde{u}, \tilde{v} \in \mathcal{M}$ at distance σ from u, v, respectively. Hence, $\|u-v\| \leqslant \|\tilde{u}-\tilde{v}\| + 2\sigma$ while $\|P_W(\tilde{u}-\tilde{v})\| \leqslant 2\sigma$, which gives the first inequality in (2.17). To prove the second inequality in (2.17), let $\tilde{u}_1$, $\tilde{u}_2 \in \mathcal{M}$ such that $\|P_W \tilde{u}_1 - P_W \tilde{u}_2\| \leqslant 2\sigma$. To construct nearby elements $u_i \in \mathcal{M}_\sigma$ sharing the same measurements, we write $\tilde{w}_i = P_W \tilde{u}_i$ and $\tilde{w}_i^\perp = P_{W^\perp} \tilde{u}_i$, $i = 1, 2$, so that $\|\tilde{w}_1 - \tilde{w}_2\| \leqslant 2\sigma$. For $i = 1, 2$, the function $u_i := (\tilde{w}_1 + \tilde{w}_2)/2 + \tilde{w}_i^\perp$ is at distance σ from $\tilde{u}_i$. Therefore, $u_i \in \mathcal{M}_\sigma$, $P_W(u_1 - u_2) = 0$, and

$$\|\tilde{u}_1 - \tilde{u}_2\|^2 = \|\tilde{w}_1 - \tilde{w}_2\|^2 + \|\tilde{w}_1^\perp - \tilde{w}_2^\perp\|^2 \leqslant (2\sigma)^2 + \|P_{W^\perp} u_1 - P_{W^\perp} u_2\|^2 \leqslant (2\sigma)^2 + \delta_\sigma^2,$$

which yields a stronger form of the upper inequality in (2.17). Finally, note that we have

$$\tilde{\delta}_{2\sigma} \leqslant 2\delta_\sigma, \quad \sigma \geqslant 0.$$

This is derived by inserting the inequality $2\sigma \leqslant \delta_\sigma$ from (2.14) in the second inequality of (2.17).

In the following analysis of reconstruction methods, we use the quantity δ_σ as a benchmark which, in view of this last observation, also accounts for the lack of accuracy in the measurement of $P_W u$. Our objective is therefore to design an algorithm that, for a given tolerance $\sigma > 0$, recovers from the measurement $w = P_W u$ an approximation to u with accuracy comparable to δ_σ. Such an algorithm requires that we are able to capture the solution manifold up to some tolerance $\varepsilon \leqslant \sigma$ by some reduced model.

3. Nonlinear recovery by reduced model selection.

3.1. Piecewise affine reduced models. Linear or affine reduced models, as used in the PBDW algorithm, are not suitable for approximating the solution manifold when the required tolerance ε is too small. In particular, when $\varepsilon < d_m(\mathcal{M})$ one would then need to use a linear space V_n of dimension $n > m$, therefore making $\mu(V_n, W)$ infinite.

One way out is to replace the single space V_n by a *family* of affine spaces

$$V_k = \overline{u}_k + \overline{V}_k, \quad k = 1, \dots, K, \tag{3.1}$$

each of them having dimension

$$\dim(V_k) = n_k \leqslant m, \tag{3.2}$$

such that the manifold is well captured by the union of these spaces, in the sense that

$$\operatorname{dist}\left(\mathcal{M}, \bigcup_{k=1}^{K} V_k\right) \leqslant \varepsilon \tag{3.3}$$

for some prescribed tolerance $\varepsilon > 0$. This is equivalent to saying that there exists a partition of the solution manifold

$$\mathcal{M} = \bigcup_{k=1}^{K} \mathcal{M}_k, \tag{3.4}$$

such that we have local certified bounds

$$\operatorname{dist}(\mathcal{M}_k, V_k) \leqslant \varepsilon_k \leqslant \varepsilon, \quad k = 1, \dots, K. \tag{3.5}$$

We may thus think of the family $(V_k)_{k=1,\dots,K}$ as a piecewise affine approximation to $\mathcal{M}$. We stress that, in contrast to the hierarchies $(V_n)_{n=0,\dots,m}$ of reduced models discussed in section 5, the spaces V_k do not have dimension k and are not nested. Most importantly, K is not limited by m while each n_k is.

The objective of using a piecewise reduced model in the context of state estimation is to have a joint control on the local accuracy ε_k as expressed by (3.5) and on the stability of the PBDW when using any individual V_k. This means that, for some prescribed $\mu > 1$, we ask that

$$\mu_k = \mu(\overline{V}_k, W) \leqslant \mu, \quad k = 1, \dots, K. \tag{3.6}$$

According to (1.22), the worst case error bound over $\mathcal{M}_k$ when using the PBDW method with a space V_k is given by the product $\mu_k\varepsilon_k$. This suggests to alternatively require from the collection $(V_k)_{k=1,\dots,K}$ that for some prescribed $\sigma > 0$, one has

$$\sigma_k := \mu_k\varepsilon_k \leqslant \sigma, \quad k = 1, \dots, K. \tag{3.7}$$

This leads us to the following definition.

Definition 3.1. *The family* $(V_k)_{k=1,\dots,K}$ *is* σ-admissible *if* (3.7) *holds. It is* (ε, μ)-admissible *if* (3.5) *and* (3.6) *are jointly satisfied.*

Obviously, any (ε, μ)-admissible family is σ-admissible with $\sigma := \mu\varepsilon$. In this sense, the notion of (ε, μ)-admissibility is thus more restrictive than that of σ-admissibility. The benefit of the first notion is in the uniform control on the size of μ which is critical in the presence of noise, as hinted at by Remark 1.1.

If $u \in \mathcal{M}$ is our unknown state and $w = P_W u$ is its observation, we may apply the PBDW method for the different V_k in the given family, which yields a corresponding family of estimators

$$u_k^* = u_k^*(w) = \operatorname{argmin}\{\operatorname{dist}(v, V_k) \,:\, v \in w + W^\perp\}, \quad k = 1, \dots, K. \tag{3.8}$$

If $(V_k)_{k=1,\dots,K}$ is σ-admissible, we find that the accuracy bound

$$\|u - u_k^*\| \leqslant \mu_k \operatorname{dist}(u, V_k) \leqslant \mu_k\varepsilon_k = \sigma_k \leqslant \sigma \tag{3.9}$$

holds whenever $u \in \mathcal{M}_k$.

Therefore, if in addition to the observed data w one had an oracle giving the information on which portion $\mathcal{M}_k$ of the manifold the unknown state sits, we could derive an estimator with worst case error

$$E_{wc} \leqslant \sigma. \tag{3.10}$$

This information is, however, not available, and such a worst case error estimate cannot be hoped for, even with an additional multiplicative constant. Indeed, as we shall see below, σ can be fixed arbitrarily small by the user when building the family $(V_k)_{k=1,\dots,K}$, while we know from section 2.1 that the worst case error is bounded from below by $E_{wc}^* \geqslant \frac{1}{2}\delta_0$, which could be nonzero. We will thus need to replace the ideal choice of k by a model selection procedure only based on the data w, that is, a map

$$w \mapsto k^*(w), \tag{3.11}$$

leading to a choice of estimator $u^* = u_{k^*}^*$. We shall prove further that such an estimator is able to achieve the accuracy

$$E_{wc} \leqslant \delta_\sigma, \tag{3.12}$$

that is, the benchmark introduced in section 2.2. Before discussing this model selection, we discuss the existence and construction of σ-admissible or (ε, μ)-admissible families.

3.2. Constructing admissible reduced model families. For any arbitrary choice of $\varepsilon > 0$ and $\mu \geqslant 1$, the existence of an (ε, μ)-admissible family results from the following observation: since the manifold $\mathcal{M}$ is a compact set of V, there exists a finite ε-cover of $\mathcal{M}$, that is, a family $\overline{u}_1, \ldots, \overline{u}_K \in V$, such that

$$\mathcal{M} \subset \bigcup_{k=1}^{K} B(\overline{u}_k, \varepsilon), \tag{3.13}$$

or equivalently, for all $v \in \mathcal{M}$, there exists a k such that $\|v - u_k\| \leqslant \varepsilon$. With such an ε cover, we consider the family of trivial affine spaces defined by

$$V_k = \{\overline{u}_k\} = \overline{u}_k + \overline{V}_k, \quad \overline{V}_k = \{0\}, \tag{3.14}$$

thus with $n_k = 0$ for all k. The covering property implies that (3.5) holds. On the other hand, for the zero-dimensional space, one has

$$\mu(\{0\}, W) = 1, \tag{3.15}$$

and therefore (3.6) also holds. The family $(V_k)_{k=1,\ldots,K}$ is therefore (ε, μ)-admissible, and also σ-admissible with $\sigma = \varepsilon$.

This family is, however, not satisfactory for algorithmic purposes for two main reasons. First, the manifold is not explicitly given to us and the construction of the centers $\overline{u}_k$ is by no means trivial. Second, asking for an ε-cover would typically require that K becomes extremely large as ε goes to 0. For example, assuming that the parameter to solution $y \mapsto u(y)$ has Lipschitz constant L,

$$\|u(y) - u(\tilde{y})\| \leqslant L|y - \tilde{y}|, \quad y, \tilde{y} \in Y, \tag{3.16}$$

for some norm $|\cdot|$ of $\mathbb{R}^d$, then an ε cover for $\mathcal{M}$ would be induced by an $L^{-1}\varepsilon$ cover for Y which has cardinality K growing like ε^{-d} as $\varepsilon \to 0$. Having a family of moderate size K is important for the estimation procedure since we intend to apply the PBDW method for all $k = 1, \ldots, K$.

In order to construct (ε, μ)-admissible or σ-admissible families of better controlled size, we need to split the manifold in a more economical manner than through an ε-cover and use spaces V_k of general dimensions $n_k \in \{0, \ldots, m\}$ for the various manifold portions $\mathcal{M}_k$. To this end, we combine standard constructions of linear reduced model spaces with an iterative splitting procedure operating on the parameter domain Y. Let us mention that various ways of splitting the parameter domain have already been considered in order to produce local reduced bases having both controlled cardinality and prescribed accuracy [32, 34, 6]. Here our goal is slightly different since we want to control both the accuracy ε and the stability μ with respect to the measurement space W.

We describe the greedy algorithm for constructing σ-admissible families and explain how it should be modified for (ε, μ)-admissible families. For simplicity, we consider the case where Y is a rectangular domain with sides parallel to the main axes, the extension to a more general bounded domain Y being done by embedding it in such a hyper-rectangle. We are given a prescribed target value $\sigma > 0$, and the splitting procedure starts from Y.

At step j, a disjoint partition of Y into rectangles $(Y_k)_{k=1,\dots,K_j}$ with sides parallel to the main axes has been generated. It induces a partition of $\mathcal{M}$ given by

$$\mathcal{M}_k := \{u(y) : y \in Y_k\}, \quad k = 1, \dots, K_j. \tag{3.17}$$

To each $k \in \{1, \dots, K_j\}$ we associate a hierarchy of affine reduced basis spaces

$$V_{n,k} = \overline{u}_k + \overline{V}_{n,k}, \quad n = 0, \dots, m, \tag{3.18}$$

where $\overline{u}_k = u(\overline{y}_k)$ with $\overline{y}_k$ the vector defined as the center of the rectangle Y_k. The nested linear spaces

$$\overline{V}_{0,k} \subset \overline{V}_{1,k} \subset \cdots \subset \overline{V}_{m,k}, \quad \dim(\overline{V}_{n,k}) = n \tag{3.19}$$

are meant to approximate the translated portion of the manifold $\mathcal{M}_k - \overline{u}_k$. For example, they could be reduced basis spaces obtained by applying the greedy algorithm to $\mathcal{M}_k - \overline{u}_k$ or spaces resulting from local n-term polynomial approximations of $u(y)$ on the rectangle Y_k. Each space $V_{n,k}$ has a given accuracy bound and stability constant

$$\operatorname{dist}(\mathcal{M}_k, V_{n,k}) \leqslant \varepsilon_{n,k} \quad \text{and} \quad \mu_{n,k} := \mu(\overline{V}_{n,k}, W). \tag{3.20}$$

We define the test quantity

$$\tau_k = \min_{n=0,\dots,m} \mu_{n,k}\varepsilon_{n,k}. \tag{3.21}$$

If $\tau_k \leqslant \sigma$, the rectangle Y_k is not split and becomes a member of the final partition. The affine space associated to $\mathcal{M}_k$ is

$$V_k = \overline{u}_k + \overline{V}_k, \tag{3.22}$$

where $V_k = V_{n,k}$ for the value of n that minimizes $\mu_{n,k}\varepsilon_{n,k}$. The rectangles Y_k with $\tau_k > \sigma$ are, on the other hand, split into a finite number of subrectangles in a way that we discuss below. This results in the new larger partition $(Y_k)_{k=1,\dots,K_{j+1}}$ after relabelling the Y_k. The algorithm terminates at the step j as soon as $\tau_k \leqslant \sigma$ for all $k = 1, \dots, K_j = K$, and the family $(V_k)_{k=1,\dots,K}$ is σ-admissible. In order to obtain an (ε, μ)-admissible family, we simply modify the test quantity τ_k by defining it instead as

$$\tau_k := \min_{n=0,\dots,m} \max\left\{\frac{\mu_{n,k}}{\mu}, \frac{\varepsilon_{n,k}}{\varepsilon}\right\} \tag{3.23}$$

and splitting the cells for which $\tau_k > 1$.

The splitting of one single rectangle Y_k can be performed in various ways. When the parameter dimension d is moderate, we may subdivide each sidelength at the midpoint, resulting in 2^d subrectangles of equal size. This splitting becomes too costly as d gets large, in which case it is preferable to make a choice of $i \in \{1, \dots, d\}$ and subdivide Y_k at the midpoint of the sidelength in the i-coordinate, resulting in only two subrectangles. In order to decide which

coordinate to pick, we consider the d possibilities and take the value of i that minimizes the quantity

$$\tau_{k,i} = \max\{\tau_{k,i}^-, \tau_{k,i}^+\}, \tag{3.24}$$

where $(\tau_{k,i}^-, \tau_{k,i}^+)$ are the values of τ_k for the two subrectangles obtained by splitting along the i-coordinate. In other words, we split in the direction that decreases τ_k most effectively. In order to be certain that all sidelengths are eventually split, we can mitigate the greedy choice of i in the following way: if Y_k has been generated by l consecutive refinements, and therefore has volume $|Y_k| = 2^{-l}|Y|$, and if l is even, we choose $i = (l/2 \bmod d)$. This means that at each even level we split in a cyclic manner in the coordinates $i \in \{1, \dots, d\}$.

Using such elementary splitting rules, we are ensured that the algorithm must terminate. Indeed, we are guaranteed that for any $\eta > 0$, there exists a level $l = l(\eta)$ such that any rectangle Y_k generated by l consecutive refinements has sidelengths smaller than 2η in each direction. Since the parameter-to-solution map is continuous, for any $\varepsilon > 0$, we can pick $\eta > 0$ such that

$$\|y - \tilde{y}\|_{\ell^\infty} \leqslant \eta \implies \|u(y) - u(\tilde{y})\| \leqslant \varepsilon, \quad y, \tilde{y} \in Y. \tag{3.25}$$

Applying this to $y \in Y_k$ and $\tilde{y} = \overline{y}_k$, we find that for $\overline{u}_k = u(\overline{y}_k)$

$$\|u - \overline{u}_k\| \leqslant \varepsilon, \quad u \in \mathcal{M}_k. \tag{3.26}$$

Therefore, for any rectangle Y_k of generation l, we find that the trivial affine space $V_k = \overline{u}_k$ has local accuracy $\varepsilon_k \leqslant \varepsilon$ and $\mu_k = \mu(\{0\}, W) = 1 \leqslant \mu$, which implies that such a rectangle would no longer be refined by the algorithm.

3.3. Reduced model selection and recovery bounds. We return to the problem of selecting an estimator within the family $(u_k^*)_{k=1,\dots,K}$ defined by (3.8). In an idealized version, the selection procedure picks the value k^* that minimizes the distance of u_k^* to the solution manifold, that is,

$$k^* = \operatorname{argmin}\{\operatorname{dist}(u_k^*, \mathcal{M}) \,:\, k = 1, \dots, K\}, \tag{3.27}$$

and takes for the final estimator

$$u^* = u^*(w) := u_{k^*}^*(w). \tag{3.28}$$

Note that k^* also depends on the observed data w. This estimation procedure is not realistic since the computation of the distance of a known function v to the manifold

$$\operatorname{dist}(v, \mathcal{M}) = \min_{y \in Y} \|u(y) - v\| \tag{3.29}$$

is a high-dimensional nonconvex problem, which requires that we explore the solution manifold. A more realistic procedure is based on replacing this distance by a surrogate quantity $\mathcal{S}(v, \mathcal{M})$ that is easily computable and satisfies a uniform equivalence

$$r \operatorname{dist}(v, \mathcal{M}) \leqslant \mathcal{S}(v, \mathcal{M}) \leqslant R \operatorname{dist}(v, \mathcal{M}), \quad v \in V, \tag{3.30}$$

for some constants $0 < r \leqslant R$. We then instead take for k^* the value that minimizes this surrogate, that is,

$$k^* = \operatorname{argmin}\{\mathcal{S}(u_k^*, \mathcal{M}) : k = 1, \dots, K\}. \tag{3.31}$$

Before discussing the derivation of $\mathcal{S}(v, \mathcal{M})$ and the relation of the constants r, R to the parametric model (1.1) in concrete cases, we establish a recovery bound in the absence of model bias and noise.

Theorem 3.2. *Assume that the family $(V_k)_{k=1,\dots,K}$ is σ-admissible for some $\sigma > 0$. Then the idealized estimator based on (3.27), (3.28) satisfies the worst case error estimate*

$$E_{wc} = \max_{u \in \mathcal{M}} \|u - u^*(P_W u)\| \leqslant \delta_\sigma, \tag{3.32}$$

where δ_σ is the benchmark quantity defined in (2.11). When using the estimator based on (3.31), the worst case error estimate is modified into

$$E_{wc} \leqslant \delta_{\kappa\sigma}, \quad \kappa = \frac{R}{r} > 1. \tag{3.33}$$

Proof. Let $u \in \mathcal{M}$ be an unknown state, and let $w = P_W u$. There exists $l = l(u) \in 1, \dots, K$, such that $u \in \mathcal{M}_l$, and for this value, we know that

$$\|u - u_l^*\| \leqslant \mu_l \varepsilon_l = \sigma_l \leqslant \sigma. \tag{3.34}$$

Since $u \in \mathcal{M}$, it follows that

$$\operatorname{dist}(u_l^*, \mathcal{M}) \leqslant \sigma. \tag{3.35}$$

On the other hand, for the value k^* selected by (3.31) and $u^* = u_{k^*}^*$, we have

$$\operatorname{dist}(u^*, \mathcal{M}) \leqslant r^{-1}\, \mathcal{S}(u^*, \mathcal{M}) \leqslant r^{-1}\, \mathcal{S}(u_l^*, \mathcal{M}) \leqslant \kappa \operatorname{dist}(u_l^*, \mathcal{M}) \leqslant \kappa\sigma. \tag{3.36}$$

It follows that u^* belongs to the offset $\mathcal{M}_{\kappa\sigma}$. Since $u \in \mathcal{M} \subset \mathcal{M}_\sigma \subseteq \mathcal{M}_{\kappa\sigma}$ and $u - u^* \in W^\perp$, we find that

$$\|u - u^*\| \leqslant \delta_{\kappa\sigma}, \tag{3.37}$$

which establishes the recovery estimate (3.33). The estimate (3.32) for the idealized estimator follows since it corresponds to having $r = R = 1$. ∎

Remark 3.3. *One possible variant of the selection mechanism, which is actually adopted in our numerical experiments, consists in picking the value k^* that minimizes the distance of u_k^* to the corresponding local portion $\mathcal{M}_k$ of the solution manifold or a surrogate $\mathcal{S}(u_k^*, \mathcal{M}_k)$ with equivalence properties analogous to (3.30). It is readily checked that Theorem 3.2 remains valid for the resulting estimator u^* with the same type of proof.*

In the above result, we do not obtain the best possible accuracy satisfied by the different u_k^* since we do not have an oracle providing the information on the best choice of k. We next show that this order of accuracy is attained in the particular case where the measurement map P_W is injective on $\mathcal{M}$ and the stability constant of the recovery problem defined in (2.15) is finite.

Theorem 3.4. *Assume that* $\delta_0 = 0$ *and that*

$$\mu(\mathcal{M}, W) = \frac{1}{2} \sup_{\sigma>0} \frac{\delta_\sigma}{\sigma} < \infty. \tag{3.38}$$

Then, for any given state $u \in \mathcal{M}$ *with observation* $w = P_W u$*, the estimator* u^* *obtained by the model selection procedure* (3.31) *satisfies the oracle bound*

$$\|u - u^*\| \leqslant C \min_{k=1,\dots,K} \|u - u_k^*\|, \quad C := 2\mu(\mathcal{M}, W)\kappa. \tag{3.39}$$

In particular, if $(V_k)_{k=1,\dots,K}$ *is* σ*-admissible, it satisfies*

$$\|u - u^*\| \leqslant C\sigma. \tag{3.40}$$

Proof. Let $l \in \{1, \dots, K\}$ be the value for which $\|u - u_l^*\| = \min_{k=1,\dots,K} \|u - u_k^*\|$. Reasoning as in the proof of Theorem 3.2, we find that

$$\operatorname{dist}(u^*, \mathcal{M}) \leqslant \kappa\beta, \quad \beta := \operatorname{dist}(u_l^*, \mathcal{M}), \tag{3.41}$$

and therefore

$$\|u - u^*\| \leqslant \delta_{\kappa\beta} \leqslant 2\mu(\mathcal{M}, W)\kappa \operatorname{dist}(u_l^*, \mathcal{M}), \tag{3.42}$$

which is (3.39). We then obtain (3.40) using the fact that $\|u - u_k^*\| \leqslant \sigma$ for the value of k such that $u \in \mathcal{M}_k$. ∎

We next discuss how to incorporate model bias and noise in the recovery bound, provided that we have a control on the stability of the PBDW method, through a uniform bound on μ_k, which holds when we use (ε, μ)-admissible families.

Theorem 3.5. *Assume that the family* $(V_k)_{k=1,\dots,K}$ *is* (ε, μ)*-admissible for some* $\varepsilon > 0$ *and* $\mu \geqslant 1$*. If the observation is* $w = P_W u + \eta$ *with* $\|\eta\| \leqslant \varepsilon_{noise}$*, and if the true state does not lie in* $\mathcal{M}$ *but satisfies* $\operatorname{dist}(u, \mathcal{M}) \leqslant \varepsilon_{model}$*, then the estimator based on* (3.31) *satisfies the estimate*

$$\|u - u^*(w)\| \leqslant \delta_{\kappa\rho} + \varepsilon_{noise}, \quad \rho := \mu(\varepsilon + \varepsilon_{noise}) + (\mu + 1)\varepsilon_{model}, \quad \kappa = \frac{R}{r}, \tag{3.43}$$

and the idealized estimator based on (3.27) *satisfies a similar estimate with* $\kappa = 1$*.*

Proof. There exists $l = l(u) \in \{1, \dots, K\}$ such that

$$\operatorname{dist}(u, \mathcal{M}_l) \leqslant \varepsilon_{model}, \tag{3.44}$$

and therefore

$$\operatorname{dist}(u, V_l) \leqslant \varepsilon_l + \varepsilon_{model}. \tag{3.45}$$

As already noted in Remark 1.1, we know that the PBDW method for this value of k has accuracy

$$\|u - u_l^*\| \leqslant \mu_l(\varepsilon_l + \varepsilon_{noise} + \varepsilon_{model}) \leqslant \mu(\varepsilon + \varepsilon_{noise} + \varepsilon_{model}). \tag{3.46}$$

Therefore,

$$\operatorname{dist}(u_l^*, \mathcal{M}) \leqslant \mu(\varepsilon + \varepsilon_{noise} + \varepsilon_{model}) + \varepsilon_{model} = \rho, \tag{3.47}$$

and in turn

$$\operatorname{dist}(u^*, \mathcal{M}) \leqslant \kappa\rho. \tag{3.48}$$

On the other hand, we define

$$v := u + \eta = u + w - P_W u = u + P_W(u^* - u), \tag{3.49}$$

so that

$$\operatorname{dist}(v, \mathcal{M}) \leqslant \|v - u\| + \varepsilon_{model} \leqslant \varepsilon_{noise} + \varepsilon_{model} \leqslant \rho. \tag{3.50}$$

Since $v - u^* \in W^\perp$, we conclude that $\|v - u^*\| \leqslant \delta_{\kappa\rho}$, from which (3.43) follows. ∎

While the reduced model selection approach provides us with an estimator $w \mapsto u^*(w)$ of a single plausible state, the estimated distance of some of the other estimates $u_k(w)$ may be of comparable size. Therefore, one could be interested in recovering a more complete estimate on a plausible *set* that may contain the true state u or even several states in $\mathcal{M}$ sharing the same measurement. This more ambitious goal can be viewed as a deterministic counterpart to the search for the entire posterior probability distribution of the state in a Bayesian estimation framework, instead of only searching for a single estimated state, for instance the expectation of this distribution. For simplicity, we discuss this problem in the absence of model bias and noise. Our goal is therefore to approximate the set

$$\mathcal{M}_w = \mathcal{M} \cap (w + W^\perp). \tag{3.51}$$

Given the family $(V_k)_{k=1,\dots,K}$, we consider the ellipsoids

$$\mathcal{E}_k := \{v \in w + W^\perp \quad \operatorname{dist}(v, V_k) \leqslant \varepsilon_k\}, \quad k = 1, \dots, K, \tag{3.52}$$

which have center u_k^* and diameter at most $\mu_k\varepsilon_k$. We already know that $\mathcal{M}_w$ is contained inside the union of the $\mathcal{E}_k$ which could serve as a first estimator. In order to refine this estimator, we would like to discard the $\mathcal{E}_k$ that do not intersect the associated portion $\mathcal{M}_k$ of the solution manifold.

For this purpose, we define our estimator of $\mathcal{M}_w$ as the union

$$\mathcal{M}_w^* := \bigcup_{k \in S} \mathcal{E}_k, \tag{3.53}$$

where S is the set of those k such that

$$\mathcal{S}(u_k^*, \mathcal{M}_k) \leqslant R\mu_k\varepsilon_k. \tag{3.54}$$

It is readily seen that $k \notin S$ implies that $\mathcal{E}_k \cap \mathcal{M}_k = \varnothing$. The following result shows that this set approximates $\mathcal{M}_w$ with an accuracy of the same order as the recovery bound established for the estimator $u^*(w)$.

Theorem 3.6. *For any state $u \in \mathcal{M}$ with observation $w = P_W u$, one has the inclusion*

$$\mathcal{M}_w \subset \mathcal{M}_w^*. \tag{3.55}$$

If the family $(V_k)_{k=1,\dots,K}$ is σ-admissible for some $\sigma > 0$, the Hausdorff distance between the two sets satisfies the bound

$$d_{\mathcal{H}}(\mathcal{M}_w^*, \mathcal{M}_w) = \max_{v \in \mathcal{M}_w^*} \min_{u \in \mathcal{M}_w} \|v - u\| \leqslant \delta_{(\kappa+1)\sigma}, \quad \kappa = \frac{R}{r}. \tag{3.56}$$

Proof. Any $u \in \mathcal{M}_w$ is a state from $\mathcal{M}$ that gives the observation $P_W u$. This state belongs to $\mathcal{M}_l$ for some particular $l = l(u)$, for which we know that u belongs to the ellipsoid $\mathcal{E}_l$ and that

$$\|u - u_l^*\| \leqslant \mu_l \varepsilon_l. \tag{3.57}$$

This implies that $\operatorname{dist}(u_l^*, \mathcal{M}_l) \leqslant \mu_l \varepsilon_l$, and therefore $\mathcal{S}(u_l^*, \mathcal{M}_l) \leqslant R\mu_l \varepsilon_l$. Hence, $l \in S$, which proves the inclusion (3.55). In order to prove the estimate on the Hausdorff distance, we take any $k \in S$, and notice that

$$\operatorname{dist}(u_k^*, \mathcal{M}_k) \leqslant \kappa \mu_k \varepsilon_k \leqslant \kappa\sigma, \tag{3.58}$$

and therefore, for all such k and all $v \in \mathcal{E}_k$, we have

$$\operatorname{dist}(v, \mathcal{M}_k) \leqslant (\kappa + 1)\mu_k \varepsilon_k. \tag{3.59}$$

Since $u - v \in W^\perp$, it follows that

$$\|v - u\| \leqslant \delta_{(\kappa+1)\sigma}, \tag{3.60}$$

which proves (3.56). ∎

Remark 3.7. *If we could take S to be exactly the set of those k such that $\mathcal{E}_k \cap \mathcal{M}_k \neq \varnothing$, the resulting $\mathcal{M}_w^*$ would still contain $\mathcal{M}_w$ but with a sharper error bound. Indeed, any $v \in \mathcal{M}_w^*$ belongs to a set $\mathcal{E}_k$ that intersects $\mathcal{M}_k$ at some $u \in \mathcal{M}_w$, so that*

$$d_{\mathcal{H}}(\mathcal{M}_w^*, \mathcal{M}_w) \leqslant 2\sigma. \tag{3.61}$$

In order to identify whether a k belongs to this smaller S, we need to solve the minimization problem

$$\min_{v \in \mathcal{E}_k} \mathcal{S}(v, \mathcal{M}_k) \tag{3.62}$$

and check whether the minimum is zero. As explained next, the quantity $\mathcal{S}(v, \mathcal{M}_k)$ is itself obtained by a minimization problem over $y \in Y_k$. The resulting double minimization problem is globally nonconvex, but it is convex separately in v and y, which allows one to apply simple alternating minimization techniques. These procedures (which are not guaranteed to converge to the global minimum) are discussed in section 4.2.

3.4. Residual based surrogates. The computational realization of the above concepts hinges on two main constituents, namely (i) the ability to evaluate bounds ε_n for $\mathrm{dist}(\mathcal{M}, V_n)$ as well as (ii) to have at hand computationally affordable surrogates $\mathcal{S}(v, \mathcal{M})$ for $\mathrm{dist}(v, \mathcal{M}) = \min_{u\in\mathcal{M}} \|v-u\|$. In both cases, one exploits the fact that errors in V are equivalent to residuals in a suitable dual norm. Regarding (i), the derivation of bounds ε_n has been discussed extensively in the context of reduced basis methods [37]; see also [29] for the more general framework discussed below. Substantial computational effort in an offline phase provides residual based surrogates for $\|u - u(y)\|$ permitting frequent parameter queries at an online stage needed, in particular, to construct reduced bases. This strategy becomes challenging, though, for high parameter dimensionality, and we refer the reader to [11] for remedies based on trading deterministic certificates against probabilistic ones at significantly reduced computational cost. Therefore, we focus here on task (ii).

One typical setting where a computable surrogate $\mathcal{S}(v, \mathcal{M})$ can be derived is when $u(y)$ is the solution to a parametrized operator equation of the general form

$$A(y)u(y) = f(y), \tag{3.63}$$

i.e., $\mathcal{P}(u, y) = f(y) - A(y)u$. Here we assume that for the given trial Hilbert space V we have identified a test Hilbert space Z such that $f(y)$ belongs to its dual Z' of Z and the operators $A(y)$ are boundedly invertible as mappings from V to the dual Z', *uniformly* in $y \in Y$. Moreover, we assume continuous dependence of $A(y)$ and $f(y)$ with respect to $y \in Y$.

It is well known that bounded invertibility is conveniently characterized in terms of a weak formulation of (3.63)

$$\mathcal{A}_y(u(y), v) = \mathcal{F}_y(v), \quad v \in Z, \tag{3.64}$$

with the parametrized bilinear form $\mathcal{A}_y(w, v) = \langle A(y)w, v\rangle_{Z',Z}$ and the linear form $\mathcal{F}_y(v) = \langle f(y), v\rangle_{Z',Z}$. Uniform bounded invertibility is then equivalent to the validity of continuity and inf-sup-conditions on the bilinear form $\mathcal{A}_y(w, v)$ for some inf-sup and continuity constants $0 < r \leqslant R < \infty$, respectively, for which one then has (see, e.g., [7])

$$\|A(y)\|_{V\to Z'} \leqslant R \quad \text{and} \quad \|A(y)^{-1}\|_{Z'\to V} \leqslant r^{-1}, \quad y \in Y. \tag{3.65}$$

This setting covers a wide range of problems, such as classical elliptic problems with $Z = V$, as well as saddle-point problems, other indefinite problems, and unsymmetric problems such as convection-diffusion problems, or space-time formulations of parabolic problems and wave equations, where in some cases Z has to be chosen different from V in order to ensure a moderate bound $\kappa = R/r$ of the *condition* of (3.64); see, e.g., [8, 9, 10, 28, 29, 40]. Specifically, (3.65) holds for perhaps the simplest instance of an elliptic operator $A(y) = -\mathrm{div}(a(y)\nabla)$ when the parametric field $a(y)$ is uniformly bounded away from 0 and ∞, as considered in section 5 below. The constants r, R then relate directly to $\min_{y\in Y} a(y)$, $\max_{y\in Y} a(y)$, respectively.

It follows from (3.65) that for any $v \in V$, one has the equivalence

$$r\|v - u(y)\|_V \leqslant \mathcal{R}(v, y) \leqslant R\|v - u(y)\|_V, \tag{3.66}$$

where

$$\mathcal{R}(v, y) := \|A(y)v - f(y)\|_{Z'} \tag{3.67}$$

is the residual of the PDE for a state v and parameter y.

Therefore, the quantity

$$\mathcal{S}(v, \mathcal{M}) := \min_{y \in Y} \mathcal{R}(v, y) \tag{3.68}$$

provides us with a surrogate of $\mathrm{dist}(v, \mathcal{M})$ that satisfies the required framing (3.30). Note that the bound $\kappa = R/r$ of the condition of (3.64) determines the tightness of the surrogate and thus enters all recovery estimates in the previous section.

One first advantage of this surrogate quantity is that for each given $y \in Y$, the evaluation of the residual $\|A(y)v - f(y)\|_{Z'}$ does not require one to compute the solution $u(y)$. Its second advantage is that the minimization in y is facilitated in the relevant case where $A(y)$ and $f(y)$ have affine dependence on y, that is,

$$A(y) = A_0 + \sum_{j=1}^{d} y_j A_j \quad \text{and} \quad f(y) = f_0 + \sum_{j=1}^{d} y_j f_j. \tag{3.69}$$

Indeed, $\mathcal{S}(v, \mathcal{M})$ then amounts to the minimization over $y \in Y$ of the function

$$\mathcal{R}(v, y)^2 := \left\| A_0 v - f_0 + \sum_{j=1}^{d} y_j (A_j v - f_j) \right\|_{Z'}^2, \tag{3.70}$$

which is a convex quadratic polynomial in y. Hence, a minimizer $y(v) \in Y$ of the corresponding constrained linear least-squares problem exists, rendering the surrogate $\mathcal{S}(v, \mathcal{M}) = \mathcal{R}(v, y(v))$ well-defined.

In all the above mentioned examples, the norm $\|\cdot\|_Z = \langle \cdot, \cdot \rangle_Z^{1/2}$ can be efficiently computed. For instance, in the simplest case of an $H_0^1(\Omega)$-elliptic problem one has $Z = V = H_0^1(\Omega)$ with

$$\langle v, z \rangle_Z = \int_\Omega \nabla v \cdot \nabla z dx. \tag{3.71}$$

The obvious obstacle is then, however, the computation of the *dual norm* $\|\cdot\|_{Z'}$ which in the particular example above is the $H^{-1}(\Omega)$-norm. A viable strategy is to use the Riesz lift $r_{\mathrm{Z}} : Z' \to Z$, defined by

$$\langle r_{\mathrm{Z}} g, z \rangle_Z = \langle g, z \rangle_{Z',Z} = g(z), \quad g \in Z',\ z \in Z, \tag{3.72}$$

which implies that $\|r_{\mathrm{Z}} g\|_Z = \|g\|_{Z'}$. Thus, $\mathcal{R}(v, y)^2$ is computed for a given $(v, y) \in V \times Y$ by introducing the lifted elements

$$e_j := r_{\mathrm{Z}}(A_j v - f_j), \quad j = 0, \ldots, d, \tag{3.73}$$

so that, by linearity,

$$\mathcal{R}(v, y)^2 = \left\| e_0 + \sum_{j=1}^{d} y_j e_j \right\|_Z^2. \tag{3.74}$$

Note that the above derivation is still idealized as the $d+1$ variational problems (3.73) are posed in the infinite-dimensional space Z. As already stressed in Remark 1.2, all computations take place in reference finite element spaces $V_h \subset V$ and $Z_h \subset Z$. We thus approximate the e_j by $e_{j,h} \in Z_h$, for $v \in V_h$, using the Galerkin approximation of (3.72). This gives rise to a computable least-squares functional

$$\mathcal{R}_h(v,y)^2 = \Big\| e_{0,h} + \sum_{j=1}^{d} y_j e_{j,h} \Big\|_Z^2, \quad y \in Y. \tag{3.75}$$

The practical distance surrogate is then defined through the corresponding constrained least-squares problem

$$\mathcal{S}_h(v, \mathcal{M}_h) := \min_{y \in Y} \mathcal{R}_h(v,y), \tag{3.76}$$

which can be solved by standard optimization methods. As indicated earlier, the recovery schemes can be interpreted as taking place in a fixed discrete setting, with $\mathcal{M}$ replaced by $\mathcal{M}_h$, comprised of approximate solutions in a large finite element space $V_h \subset V$, and measuring accuracy only in this finite-dimensional setting. One should note, though, that the approach allows one to disentangle discretization errors from recovery estimates, even with regard to the underlying continuous PDE model. In fact, given any target tolerance ε_h, using a posteriori error control in Z, the spaces V_h, Z_h can be chosen large enough to guarantee that

$$\big|\mathcal{R}(v,y) - \mathcal{R}_h(v,y)\big| \leqslant \varepsilon_h \|v\|, \quad v \in V_h. \tag{3.77}$$

Accordingly, one has $\big|\mathcal{S}_h(v,\mathcal{M}_h) - \mathcal{S}(v,\mathcal{M})\big| \leqslant \varepsilon_h \|v\|$, so that recovery estimates remain meaningful with respect to the continuous setting as long as ε_h remains sufficiently dominated by the thresholds $\varepsilon_k, \sigma, \varepsilon_{\text{noise}}$ appearing in the above results. For notational simplicity, we therefore continue working in the continuous setting.

4. Joint parameter and state estimation.

4.1. An estimate for y.

Searching for a parameter $y \in Y$, which explains an observation $w = P_W u(y)$, is a nonlinear inverse problem. As shown next, a quantifiable estimate for y can be obtained from a state estimate $u^*(w)$ combined with a residual minimization.

For any state estimate $u^*(w)$ which we compute from w, the most plausible parameter is the one associated to the metric projection of $u^*(w)$ into $\mathcal{M}$, that is,

$$y^* \in \operatorname*{argmin}_{y \in Y} \|u(y) - u^*(w)\|.$$

Note that y^* depends on w but we omit the dependence in the notation in what follows. Finding y^* is a difficult task since it requires solving a nonconvex optimization problem.

However, as we have already noticed, a near metric projection of u^* to $\mathcal{M}$ can be computed through a simple convex problem in the case of affine parameter dependence (3.69), minimizing the residual $\mathcal{R}(v,y)$ given by (3.70). Our estimate for the parameter is therefore

$$y^* \in \operatorname*{argmin}_{y \in Y} \mathcal{R}(u^*, y), \tag{4.1}$$

and it satisfies, in view of (3.66),

$$\|u^* - u(y^*)\| \leqslant r^{-1}\mathcal{R}(u^*, y^*) \leqslant \kappa \operatorname{dist}(u^*, \mathcal{M}), \quad \kappa = R/r. \tag{4.2}$$

Hence, if we use, for instance, the state estimate $u^*(w)$ from (3.28), we conclude by Theorem 3.2 that $u(y^*)$ deviates from the true state $u(y)$ by

$$\begin{aligned} \|u(y) - u(y^*)\| &\leqslant \|u(y) - u^*(w)\| + \|u^*(w) - u(y^*)\| \\ &\leqslant (1+\kappa)\|u(y) - u^*(w)\| \\ &\leqslant (1+\kappa)\delta_{\kappa\sigma}, \end{aligned} \tag{4.3}$$

where $\delta_{\kappa\sigma}$ is the benchmark quantity defined in (2.11). If, in addition, $P_W : \mathcal{M} \to W$ is also injective so that $\delta_0 = 0$, and if W and $\mathcal{M}$ are favorably oriented, as detailed in the assumptions of Theorem 3.4, one even obtains

$$\|u(y) - u(y^*)\| \leqslant (2\mu(\mathcal{M}, W) + 1)\kappa\sigma. \tag{4.4}$$

To derive from such bounds estimates for the deviation of y^* from y, more information on the underlying PDE model is needed. For instance, for the second order parametric family of elliptic PDEs (1.35) and strictly positive right-hand side f, it is shown in [5] that the parameter-to-solution map is injective. If, in addition. the parameter dependent diffusion coefficient $a(y)$ belongs to $H^1(\Omega)$, one has a quantitative *inverse stability* estimate of the form

$$\|a(y) - a(y')\|_{L^2(\Omega)} \leqslant C\|u(y) - u(y')\|^{1/6}. \tag{4.5}$$

Combining this, for instance, with (4.3) yields

$$\|a(y) - a(y^*)\|_{L^2(\Omega)} \leqslant C(1+\kappa)^{1/6}\delta_{\kappa\sigma}^{1/6}. \tag{4.6}$$

Under the favorable assumptions of Theorem 3.4, one obtains a bound of the form

$$\|a(y) - a(y^*)\|_{L^2(\Omega)} \lesssim \sigma^{1/6}. \tag{4.7}$$

Finally, in relevant situations (Karhunen–Loève expansions) the functions ψ_j in the expansion of $a(y)$ form an L^2-orthogonal system. The above estimates translate then into estimates for a weighted ℓ_2-norm,

$$\Big(\sum_{j\geqslant 1} c_j(y_j - y_j^*)^2\Big)^{1/2} \lesssim \sigma^{1/6}, \tag{4.8}$$

where $c_j = \|\psi_j\|_{L^2}^2$.

4.2. Alternating residual minimization. The state estimate $u^*(w)$ is defined by selecting among the potential estimates $u_k^*(w)$ the one that sits closest to the solution manifold, in the sense of the surrogate distance $\mathcal{S}(v, \mathcal{M})$. Finding the element in $w + W^\perp$ that is closest to $\mathcal{M}$ would provide a possibly improved state estimate and as pointed out in the previous section

also an improved parameter estimate. As explained earlier, it would help in addition with improved set estimators for $\mathcal{M}_w$.

Adhering to the definition of the residual $\mathcal{R}(v,y)$ from (3.67), we are thus led to consider the double minimization problem

$$\min_{(v,y)\in(w+W^\perp)\times Y} \mathcal{R}(v,y) = \min_{v\in w+W^\perp} \mathcal{S}(v,\mathcal{M}). \tag{4.9}$$

We first show that a global minimizing pair (u^*,y^*) meets the optimal benchmarks introduced in section 2. In the unbiased and noiseless case, the value of the global minimum is 0, attained by the exact parameter y and state $u(y)$. Any global minimizing pair (u^*,y^*) will thus satisfy $P_W u^* = w$ and $u^* = u(y^*) \in \mathcal{M}$. In other words, the state estimate u^* belongs to $\mathcal{M}_w$ and therefore meets the optimal benchmark

$$\|u-u^*\| \leqslant \delta_0. \tag{4.10}$$

In the case of model bias and noise of amplitude ε_{noise}, the state u satisfies

$$\operatorname{dist}(u,\mathcal{M}) \leqslant \varepsilon_{model} \quad \text{and} \quad \|w - P_W u\| \leqslant \varepsilon_{noise}. \tag{4.11}$$

It follows that there exists a parameter y such that $\|u-u(y)\| \leqslant \varepsilon_{model}$ and a state $\tilde{u} \in w+W^\perp$ such that $\|u-\tilde{u}\| \leqslant \varepsilon_{noise}$. For this state and parameter, one thus has

$$\mathcal{R}(\tilde{u},y) \leqslant R\|u(y)-\tilde{u}\| \leqslant R(\varepsilon_{model}+\varepsilon_{noise}). \tag{4.12}$$

Any global minimizing pair (u^*,y^*) will thus satisfy

$$\|u^*-u(y^*)\| \leqslant \frac{1}{r}\mathcal{R}(u^*,y^*) \leqslant \kappa(\varepsilon_{model}+\varepsilon_{noise}), \quad \kappa := \frac{R}{r}. \tag{4.13}$$

Therefore, u^* belongs to the set $\mathcal{M}_{\varepsilon,w}$, as defined by (2.10), with $\varepsilon := \kappa(\varepsilon_{model}+\varepsilon_{noise})$, and so does $\tilde{u}$ since $\|\tilde{u}-u(y)\| \leqslant \varepsilon_{model}+\varepsilon_{noise} \leqslant \varepsilon$. In turn, the state estimate u^* meets the perturbed benchmark

$$\|u^*-u\| \leqslant \varepsilon_{noise} + \|u^*-\tilde{u}\| \leqslant \varepsilon_{noise} + \delta_\varepsilon \leqslant \frac{3\delta_\varepsilon}{2} \tag{4.14}$$

since $\varepsilon_{noise} \leqslant \varepsilon \leqslant \delta_\varepsilon/2$, having used (2.14) in the last step.

From a numerical perspective, the search for a global minimizing pair is a difficult task due to the fact that $(v,y) \mapsto \mathcal{R}(v,y)$ is generally not a convex function. However, it should be noted that in the case of affine parameter dependence (3.69), the residual $\mathcal{R}(v,y)$ given by (3.70) is a convex function in each of the two variables v,y separately, keeping the other one fixed. More precisely, $(v,y) \mapsto \mathcal{R}(v,y)^2$ is a quadratic convex function in each variable. This suggests the following alternating minimization procedure. Starting with an initial guess $u^0 \in w+W^\perp$, we iteratively compute for $j=0,1,2,\dots,$

$$y^{j+1} \in \operatorname*{argmin}_{y\in Y} \mathcal{R}(u^j,y), \tag{4.15}$$

$$u^{j+1} \in \operatorname*{argmin}_{v\in w+W^\perp} \mathcal{R}(v,y^{j+1}). \tag{4.16}$$

Each problem has a simply computable minimizer, as discussed in the next section, and the residuals are nonincreasing:

$$\mathcal{R}(u^j, y^j) \geqslant \mathcal{R}(u^j, y^{j+1}) \geqslant \mathcal{R}(u^{j+1}, y^{j+1}) \geqslant \cdots . \tag{4.17}$$

Of course, one cannot guarantee in general that (u^j, y^j) converges to a global minimizer, and the procedure may stagnate at a local minimum.

The above improvement property still tells us that if we initialize the algorithm by taking $u^0 = u^* = u^*(w)$ to be the state estimate from (3.28) and $y^0 \in \operatorname{argmin}_{y\in Y}\mathcal{R}(u^*, y)$, then we are ensured at step k that

$$\mathcal{R}(u^j, y^j) \leqslant \mathcal{R}(u^*, y^*), \tag{4.18}$$

and therefore, by the same arguments as in the proof of Theorem 3.5, one finds that

$$\|u - u^j\| \leqslant \delta_{\kappa\rho} + \varepsilon_{noise}, \tag{4.19}$$

with κ and ρ as in (3.43). In other words, the new estimate u^j satisfies at least the same accuracy bound as u^*. The numerical tests performed in section 5.3 reveal that it can be significantly more accurate.

4.3. Computational issues. We now explain how to efficiently compute the steps in (4.15) and (4.16). We continue to consider a family of linear parametric PDEs with affine parameter dependence (3.69), admitting a uniformly stable variational formulation over the pair trial and test spaces V, Z; see (3.64)–(3.65).

Minimization of (4.15). Problem (4.9) requires minimizing $\mathcal{R}(v, y)$ for a fixed $v \in w + W^\perp$ over $y \in Y$. According to (3.74), it amounts to solving a linear least-squares problem constrained to $y \in Y$,

$$\min_{y\in Y} \left\| e_0 + \sum_{j=1}^{d} y_j e_j \right\|_Z^2, \tag{4.20}$$

where the $e_j \in Z$ are the Riesz lifts $r_Z(A_j v - f_j)$, $j = 0, \ldots, d$, defined in (3.73). As indicated earlier, the numerical solution of (4.20) (for $e_j = e_{j,h} \in Z_h \subset Z$) is standard when Y is convex.

Minimization of (4.16). Problem (4.16) is of the form

$$\min_{v\in w+W^\perp} \mathcal{R}(v, y)^2 = \min_{v\in w+W^\perp} \|A(y)v - f(y)\|_{Z'}^2 \tag{4.21}$$

for a fixed $y \in Y$. A naive approach for solving (4.21) would consist in working in a closed subspace of $\widetilde{W}^\perp \subseteq W^\perp$ of sufficiently large dimension. We would then optimize over $v \in w + \widetilde{W}^\perp$. However, this would lead to a large quadratic problem of size $\dim \widetilde{W}^\perp$ which would involve $\dim \widetilde{W}^\perp$ Riesz representer computations. We next propose an alternative strategy involving the solution of only $m + 3$ variational problems. To that end, we assume in what follows that V is continuously embedded in Z', which is the case for all the examples of interest, mentioned earlier in the paper.

The proposed strategy is based on two isomorphisms from V to Z that preserve inner products in a sense to be explained next. We make again heavy use of the Riesz isometry defined in (3.72) and consider the two isomorphisms

$$T = T(y) := r_{\mathrm{Z}} A(y) : V \to Z, \quad S = S(y) := A(y)^{-*} r_{\mathrm{V}}^{-1} : V \to Z, \tag{4.22}$$

where $r_Z : Z' \to Z$ and $r_V : V' \to V$ are the previously introduced Riesz lifts. One then observes that, by standard duality arguments, they preserve inner products, in the sense that for $u, v \in V$,

$$\langle Tu, Sv\rangle_Z = \langle r_{\mathrm{Z}} A(y)u, A(y)^{-*} r_{\mathrm{V}}^{-1} v\rangle_Z = \langle u, v\rangle_V, \tag{4.23}$$

where we have used self-adjointness of Riesz isometries. In these terms, the objective functional $\mathcal{R}(v,y)^2$ takes the form

$$\|A(y)v - f(y)\|_{Z'}^2 = \|Tv - r_{\mathrm{Z}} f(y)\|_Z^2. \tag{4.24}$$

We can use (4.23) to reformulate (4.21) as

$$\min_{v \in w + W^\perp} \mathcal{R}(v,y)^2 = \min_{v \in w + W^\perp} \|Tv - r_{\mathrm{Z}} f(y)\|_Z^2 = \min_{z \in Tw + S(W)^\perp} \|z - r_{\mathrm{Z}} f(y)\|_Z^2, \tag{4.25}$$

where we have used that $T(W^\perp) = S(W)^\perp$ to obtain the last equality. Note that the unique solution $z^* \in Z$ to the right-hand side gives a solution $v^* \in V$ to the original problem through the relationship $Tv^* = z^*$. The minimizer z^* can be obtained by an appropriate orthogonal projection onto $S(W)$. This indeed amounts to solving a fixed number of $m+3$ variational problems without compromising accuracy by choosing a perhaps too moderate dimension for a subspace $\widetilde{W}^\perp$ of $W^\perp$.

More precisely, we have $z^* = Tw + \tilde{z}$, where $\tilde{z} \in S(W)^\perp$ minimizes $\|\tilde{z} + Tw - r_{\mathrm{Z}} f(y)\|_Z^2$, and therefore

$$\tilde{z} = P_{S(W)^\perp}(r_{\mathrm{Z}} f(y) - Tw) = r_{\mathrm{Z}} f(y) - Tw - P_{S(W)}(r_{\mathrm{Z}} f(y) - Tw). \tag{4.26}$$

This shows that

$$z^* = z^*(y) := f(y) - P_{S(W)}(r_{\mathrm{Z}} f(y) - Tw). \tag{4.27}$$

Thus, a single iteration of the type (4.21) requires assembling z^* followed by solving the variational problem

$$\langle Tv^*, z\rangle_Z = (A(y)v^*)(z) = \langle z^*, z\rangle_Z, \quad z \in Z, \tag{4.28}$$

that gives v^*. Assembling z^* involves

(i) evaluating Tw, which means solving the Riesz lift $\langle Tw, z\rangle_Z = (A(y)w)(z)$, $z \in Z$;
(ii) computing the Riesz lift $r_{\mathrm{Z}} f(y)$ by solving $\langle r_{\mathrm{Z}} f(y), z\rangle_Z = (f(y))(z)$, $z \in Z$;
(iii) computing the projection $P_{S(W)}(r_{\mathrm{Z}} f(y) - Tw)$, which requires computing the transformed basis functions $Sw_i = A(y)^{-*} r_{\mathrm{V}}^{-1} w_i$, which are solutions to the variational problems

$$(A(y)^* S w_i)(v) = \langle w_i, v\rangle_V, \ v \in V, \quad i = 1, \dots, m. \tag{4.29}$$

Of course, these variational problems are solved only approximately in appropriate large but finite-dimensional spaces $V_h \subset V, Z_h \subset Z$ along the remarks at the end of the previous section. While approximate Riesz lifts involve symmetric variational formulations which are well-treated by Galerkin schemes, the problems involving the operator $A(y)$ or $A(y)^*$ may in general require an unsymmetric variational formulation where $Z \neq V$ and Petrov–Galerkin schemes on the discrete level. For each of the examples (such as a time-space variational formulation of parabolic or convection diffusion equations), stable discretizations are known; see, e.g., [8, 10, 28, 29, 40].

A particularly important strategy for unsymmetric problems is to write the PDE first as an equivalent system of first order PDEs permitting a so-called ultraweak formulation where the (infinite-dimensional) trial space V is actually an L_2-space and the required continuous embedding $V \subset Z'$ holds. The mapping r_{V} is then just the identity, and so-called discontinuous Petrov–Galerkin methods offer a way of systematically finding appropriate test spaces in the discrete case with uniform inf-sup stability [9]. In this context, the mapping T from (4.22) plays a pivotal role in the identification of "optimal test spaces" and is referred to as the "trial-to-test-map."

Of course, in the case of problems that admit a symmetric variational formulation, i.e., $V = Z$, things simplify even further. To exemplify this, consider a parametric family of elliptic PDEs (1.35). In this case, one has (assuming homogeneous boundary conditions) $V = Z = H^1_0(D)$ so that $r_{\mathrm{Z}} = r_{\mathrm{V}} = \Delta^{-1}$. Due to the self-adjointness of the underlying elliptic operators $A(y)$ in this case, the problems (4.29) are of the same form as in (4.28), which can be treated on the discrete level by standard Galerkin discretizations.

5. Numerical illustration.

In this section, we illustrate the construction of nonlinear reduced models and demonstrate the mechanism of model selection using the residual surrogate methods outlined in section 3.4.

In our tests, we consider the elliptic problem mentioned in section 1.3 on the unit square $D =]0,1[^2$ with homogeneous Dirichlet boundary conditions and a parameter dependence in the diffusivity field a. Specifically, we consider the problem

$$-\operatorname{div}(a(y)\nabla u) = f, \tag{5.1}$$

with $f = 1$ on D, with $u_{|\partial D} = 0$. The classical variational formulation uses the same trial and test space $V = Z = H^1_0(D)$. We perform space discretization by the Galerkin method using $\mathbb{P}_1$ finite elements to produce solutions $u_h(y)$, with a triangulation on a regular grid of mesh size $h = 2^{-7}$.

5.1. Test 1: Predetermined splittings.

In this first test, we examine the reconstruction performance with localized reduced bases on a manifold having a predetermined splitting. Specifically, we consider two partitions of the unit square into subdomains $\{D_{1,\ell}\}_{\ell=1}^4$ and $\{D_{2,\ell}\}_{\ell=1}^4$, with

$$D_{1,1} := \left]0,\frac{3}{4}\right[\times\left]0,\frac{3}{4}\right[\quad D_{1,2} := \left]0,\frac{3}{4}\right[\times\left]\frac{3}{4},1\right[\quad D_{1,3} := \left]\frac{3}{4},1\right[\times\left]0,\frac{3}{4}\right[\quad D_{1,4} := \left]\frac{3}{4},1\right[\times\left]\frac{3}{4},1\right[,$$
$$D_{2,1} := \left]\frac{1}{4},1\right[\times\left[\frac{1}{4},1\right[\quad D_{2,2} := \left]\frac{1}{4},1\right[\times\left]0,\frac{1}{4}\right[\quad D_{2,3} := \left]0,\frac{1}{4}\right[\times\left]\frac{1}{4},1\right[\quad D_{2,4} := \left]0,\frac{1}{4}\right[\times\left]0,\frac{1}{4}\right[.$$

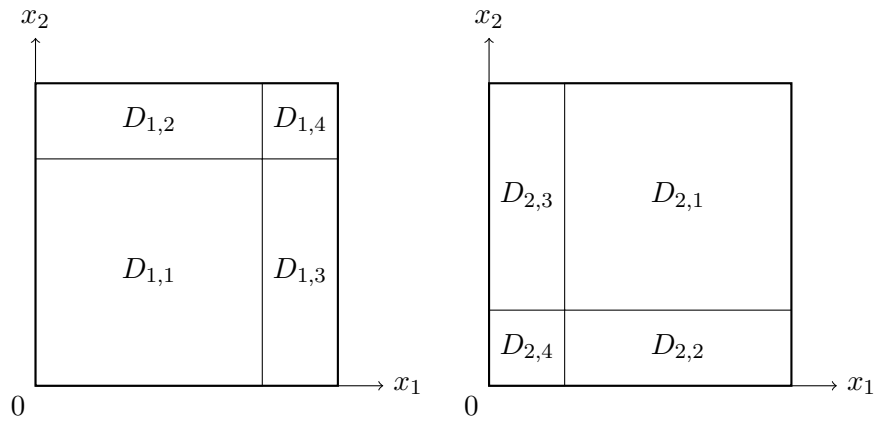


Figure 5.1. *Left: the partition of the unit square used in a_1. Right: the partition used in a_2.*

The partitions are symmetric along the line $x_1 + x_2 = 1$ as illustrated in Figure 5.1. This will play a role in the understanding of the results below. We next define two parametric diffusivity fields

$$a_1(y) := \overline{a} + c\sum_{\ell=1}^{4} \chi_{D_{1,\ell}} y_\ell \quad \text{and} \quad a_2(y) := \overline{a} + c\sum_{\ell=1}^{4} \chi_{D_{2,\ell}} y_\ell\,, \tag{5.2}$$

where the vector of parameters $y = (y_1, \dots, y_4)$ ranges in $Y = [-1, 1]^4$ and $\chi_{D_{1,\ell}}$ is the indicator function of $D_{1,\ell}$ (similarly for $\chi_{D_{2,\ell}}$). The fields $a_1(y)$ and $a_2(y)$ are mirror images of each other along $x_1 + x_2 = 1$. In the numerical tests that follow, we take $\overline{a} = 1$ and $c = 0.9$.

We denote by $u_1(y)$ the solution to the elliptic problem (5.1) with diffusivity field $a_1(y)$ and then label by $\mathcal{M}_1 := \{u_1(y) : y \in Y\}$ the resulting solution manifold. Strictly speaking, we should write $\mathcal{M}_{h,1}$, as our solutions are finite-dimensional approximations; however, we suppress the h, as there should be little ambiguity going forward. Similarly, $\mathcal{M}_2$ is the set of all solutions $u_2(y)$ of (5.1) over Y where the diffusivity field is given by a_2. We take their union $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$ to be our global solution manifold that has the obvious predetermined splitting available to us.

For our computations, we generate training and test sets. For the training, we draw $N_{\text{tr}} = 5000$ independent samples $\widetilde{Y}_{\text{tr}} = (y_j^{\text{tr}})_{j=1}^{N_{\text{tr}}}$ that are uniformly distributed over Y. The collection of solutions $\widetilde{\mathcal{M}}_1 := \{u_1(y_j^{\text{tr}})\}_{j=1}^{N_{\text{tr}}}$ and $\widetilde{\mathcal{M}}_2 := \{u_2(y_j^{\text{tr}})\}_{j=1}^{N_{\text{tr}}}$ are used as *training sets* for $\mathcal{M}_1$ and $\mathcal{M}_2$. The training set for the full manifold $\mathcal{M}$ is $\widetilde{\mathcal{M}} = \widetilde{\mathcal{M}}_1 \cup \widetilde{\mathcal{M}}_2$. Since we use the same parameter points y_j^{tr} for both sets, any solution in $\widetilde{\mathcal{M}}_1$ has a corresponding solution in $\widetilde{\mathcal{M}}_2$ that is its symmetric image along the line $x_1 + x_2 = 1$. To test our reconstruction methods, we generate $N_{\text{te}} = 2000$ independent points in Y that are distinct from the training set. The corresponding test solution sets are $\widetilde{\mathcal{T}}_1$ and $\widetilde{\mathcal{T}}_2$. All computations are done by solving (5.1) in the finite element space.

Given an unknown $u \in \mathcal{M}$, we want to recover it from its observation $w = P_W u$. For the measurement space W, we take a collection of $m = 8$ measurement functionals $\ell_i(u) = \langle \omega_i, u\rangle = |B_i|^{-1} \int u\, \chi_{B_i}$ that are local averages in a small area B_i which are boxes of width

$2h = 2^{-6}$, each placed randomly in the unit square. The measurement space is then $W = \operatorname{span}\{\omega_1, \dots, \omega_m\}$. Note that the sensor locations (as well as their nature) affect the "visibility" of the observed states and hence the value of the stability constants $\mu(V_n, W)$. Searching for a judicious sensor placement in the spirit of [4] could help keeping them small in favor of a more accurate recovery. However, in many applications the sensor locations are fixed. Random locations, on the one hand, are to account for this fact and, on the other hand, avoid "accidentally favorable" symmetries.

Since we are only given w, we do not know whether the function to reconstruct is in $\mathcal{M}_1$ or $\mathcal{M}_2$, and we consider two possibilities for reconstruction:

- *Affine method.* We use affine reduced models $V_{n,0} = \bar{u}_0 + \bar{V}_{n,0}$ generated for the full manifold $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$. In our example, we take $\bar{u}_0 = u(y = 0)$ and $\bar{V}_{n,0}$ is computed by the greedy selection algorithm over $\widetilde{\mathcal{M}} - \bar{u}_0$. Of course the spaces $V_{n,0}$ with n sufficiently large have high potential for approximation of the full manifold $\mathcal{M}$ and obviously also for the subsets $\mathcal{M}_1$ and $\mathcal{M}_2$ (see Figure 5.3). Yet, we can expect some bad artefacts in the reconstruction with this space since the true solution will be approximated by snapshots, some coming from the wrong part of the manifold and thus associated to the wrong partition of D. In addition, we can only work with $n \leqslant m = 8$ and this may not be sufficient regarding the approximation power. Our estimator $u_0^*(w)$ uses the space $V_{n_0^*,0}$, where the dimension n_0^* is the one that reaches $\tau_0 = \min_{1 \leqslant n \leqslant m} \mu_{n,0}\varepsilon_{n,0}$ as defined in (3.20) and (3.21). Figure 5.3 shows the product $\mu_{n,0}\varepsilon_{n,0}$ for $n = 1, \dots, m$, and we see that $n_0^* = 4$.
- *Nonlinear method.* We generate affine reduced bases spaces $V_{n,1} = \bar{u}_1 + \bar{V}_{n,1}$ and $V_{n,2} = \bar{u}_2 + \bar{V}_{n,2}$, each one specific for $\mathcal{M}_1$ and $\mathcal{M}_2$. Similarly as for the affine method, we take as offsets $\bar{u}_i = u_i(y = 0) = \bar{u}_0$, for $i = 1, 2$, and we run two separate greedy algorithms over $\widetilde{\mathcal{M}}_1 - \bar{u}_1$ and $\widetilde{\mathcal{M}}_2 - \bar{u}_2$ to build $\bar{V}_{n,1}$ and $\bar{V}_{n,2}$. We select the dimensions n_k^* that reach $\tau_k = \min_{n=1,\dots,m} \mu_{n,k}\varepsilon_{n,k}$ for $k = 1, 2$. From Figure 5.3, we deduce that $n_1^* = 3$ and $n_2^* = 3$. This yields two estimators $u_1^*(w)$ and $u_2^*(w)$. We can expect better results than the affine approach if we can detect well in which part of the manifold the target function is located. The main question is thus whether our model selection strategy outlined in section 3.3 is able to detect well from the observed data w whether the true u lies in $\mathcal{M}_1$ or $\mathcal{M}_2$. For this, we compute the surrogate manifold distances

$$\mathcal{S}(u_k^*(w), \mathcal{M}_k) := \min_{y \in Y} \mathcal{R}_k(u_k^*(w), y), \quad k = 1, 2, \tag{5.3}$$

where

$$\mathcal{R}_k(u_k^*(w), y) := \|\operatorname{div}(a_k(y)\nabla u_k^*(w)) + f\|_{V'}$$

is the residual of $u_k^*(w)$ related to the PDE with diffusion field $a_k(y)$. To solve problem (5.3), we follow the steps given in section 3.4. The final estimator is $u^* = u_{k^*}^*$, where

$$k^* = \operatorname*{argmin}_{k=1,2} \mathcal{S}(u_k^*(w), \mathcal{M}_k).$$

Table 1 quantifies the quality of the model selection approach. It displays how many times our model selection strategy yields the correct result $k^* = 1$ or incorrect result $k^* = 2$ for

Table 1
Performance of model selection and oracle selection.

Test function from:

Surrogate selection	$\widetilde{\mathcal{T}}_1 \subset \mathcal{M}_1$	$\widetilde{\mathcal{T}}_2 \subset \mathcal{M}_2$
$k^* = 1$	1625	386
$k^* = 2$	375	1614
Success rate	81.2 %	80.7 %

Test function from:

Oracle selection	$\widetilde{\mathcal{T}}_1 \subset \mathcal{M}_1$	$\widetilde{\mathcal{T}}_2 \subset \mathcal{M}_2$
$k^* = 1$	1962	9
$k^* = 2$	38	1991
Success rate	98.1 %	99.5 %

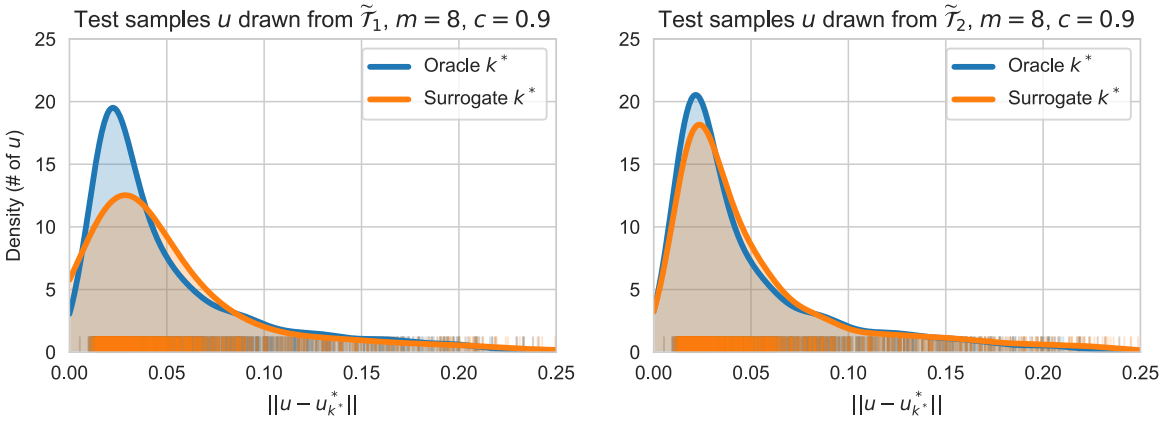


Figure 5.2. *Kernel density estimate (KDE) plot of the u_1^* and u_2^*.*

the functions from the test set $\widetilde{\mathcal{T}}_1 \subset \mathcal{M}_1$ (and vice versa for $\widetilde{\mathcal{T}}_2$). Recalling that these tests sets have $N_{te} = 2000$ snapshots, we conclude that the residual gives us the correct manifold portion roughly 80% of the time. We can compare this performance with the one given by the oracle estimator (see Table 1)

$$k^*_{\text{oracle}} = \operatorname*{argmin}_{k=1,2} \|u - u_k^*(w)\|.$$

In this case, we see that the oracle selection is very efficient since it gives us the correct manifold portion roughly 99% of the time. Figure 5.2 completes the information given in Table 1 by showing the distribution of the values of the residuals and oracle errors. The distributions give visual confirmation that both the model and oracle selection tend to pick the correct model by giving residual/error values which are lower in the right manifold portion. Figure 5.3 gives information on the value of inf-sup constants and residual errors leading to the choice of the dimension n^* for the reduced models. Last, but not least, Table 2 summarizes the reconstruction errors.

5.2. Test 2: Constructing σ-admissible families. In this example, we examine the behavior of the splitting scheme to construct σ-admissible families outlined in section 3.2.

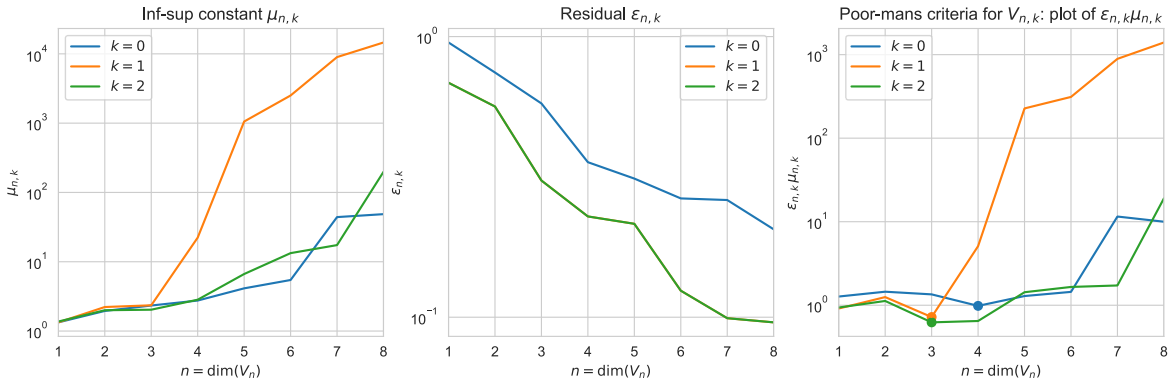


Figure 5.3. *Inf-sup constants $\mu_{n,k}$ and residual errors $\varepsilon_{n,k}$, leading to the choice of the dimension n^* for V_k.*

Table 2
Reconstruction errors with the different methods.

	Test function from:	
Average error	$\widetilde{\mathcal{T}}_1 \subset \mathcal{M}_1$	$\widetilde{\mathcal{T}}_2 \subset \mathcal{M}_2$
Affine method u_0^*	6.047e-02	6.661e-02
Nonlinear with oracle model selection	5.057e-02	4.855e-02
Nonlinear with surrogate model selection	5.522e-02	5.201e-02

	Test function from:	
Worst case error	$\widetilde{\mathcal{T}}_1 \subset \mathcal{M}_1$	$\widetilde{\mathcal{T}}_2 \subset \mathcal{M}_2$
Affine method u_0^*	4.203e-01	4.319e-01
Nonlinear with oracle model selection	2.786e-01	2.641e-01
Nonlinear with surrogate model selection	4.798e-01	2.660e-01

The manifold $\mathcal{M}$ is given by the solutions to (5.1) associated to the diffusivity field

$$a(y) = \bar{a} + \sum_{\ell=1}^{d} c_\ell \chi_{D_\ell} y_\ell, \quad y \in Y, \tag{5.4}$$

where $\bar{a} = 1$, χ_{D_ℓ} is the indicator function on the set D_ℓ, and parameters range uniformly in $Y = [-1, 1]^d$. We study the impact of the intrinsic dimensionality of the manifold by considering two cases for the partition of the unit square D, a 2×2 uniform grid partition resulting in $d = 4$ parameters and a 4×4 grid partition of D resulting in $d = 16$ parameters. We also study the impact of coercivity and anisotropy on our reconstruction algorithm by examining the different manifolds generated by taking $c_\ell = c_1 \ell^{-r}$ with $c_1 = 0.9$ or 0.99 and $r = 1$ or 2. The value $c_1 = 0.99$ corresponds to a severe degeneration of coercivity, and the rate $r = 2$ corresponds to a more pronounced anisotropy.

We use two different measurement spaces, one with $m = \dim(W) = 4$ evenly spaced local averages and the other with $m = 16$ evenly spaced local averages. The measurement locations are shown diagrammatically in Figure 5.4. The local averages are exactly as in the last test, taken in squares of sidelength 2^{-6}. Note that the two values $m = 4$ and $m = 16$ which we consider for the dimension of the measurement space are the same as the parameter dimensions $d = 4$ and $d = 16$ of the manifolds. This allows us to study different regimes:

- When $m < d$, we have a highly ill-posed problem since the intrinsic dimension of the manifold is larger than the dimension of the measurement space. In particular,

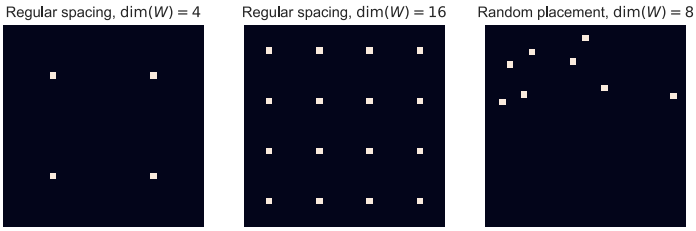


Figure 5.4. *Measurement locations for Tests* 2 *and* 3.

we expect that the fundamental barrier $\delta_0(\mathcal{M})$ is strictly positive. Thus, we cannot expect very accurate reconstructions even with the splitting strategy.

- When $m \geqslant d$, the situation is more favorable and we can expect that the reconstruction involving manifold splitting brings significant accuracy gains.

As in the previous case, the training set $\widetilde{\mathcal{M}}$ is generated by a subset $\widetilde{Y}_{\rm tr} = \{y_j^{\rm tr}\}_{j=1,\dots,N_{\rm tr}}$ of $N_{\rm tr} = 5000$ samples taken uniformly on Y. We build the σ-admissible families outlined in section 3.2 using a dyadic splitting, and the splitting rule is given by (3.21). For example, our first split of Y results in two rectangular cells Y_1 and Y_2, and the corresponding collections of parameter points $\widetilde{Y}_1 \subset Y_1$ and $\widetilde{Y}_2 \subset Y_2$, as well as split collections of solutions $\widetilde{\mathcal{M}}_1$ and $\widetilde{\mathcal{M}}_2$. On each $\widetilde{\mathcal{M}}_k$ we apply the greedy selection procedure, resulting in V_k, with computable values μ_k and ε_k. The coordinate direction in which we split Y is precisely the direction that gives us the smallest resulting $\sigma = \max_{k=1,2} \mu_k \varepsilon_k$, so we need to compute greedy reduced bases for each possible splitting direction before deciding which results in the lowest σ. Subsequent splittings are performed in the same manner, but at each step we first chose cell $k_{\rm split} = \operatorname{argmax}_{k=1,\dots,K} \mu_k \varepsilon_k$ to be split.

After $K-1$ splits, the parameter domain is divided into $Y = \bigcup_{k=1}^K Y_k$ disjoint subsets Y_k and we have computed a family of K affine reduced spaces $(V_k)_{k=1,\dots,K}$. For a given $w \in W$, we have K possible reconstructions $u_1^*(w), \dots u_K^*(w)$ and we select a value k^* with the surrogate based model selection outlined in section 3.4. The test is done on a test set of $N_{\rm te} = 1000$ snapshots which are different from the ones used for the training set $\widetilde{\mathcal{M}}$.

In Figure 5.5 we plot the reconstruction error, averaged over the test set, as a function of the number of splits K for all the different configurations: we consider the two different diffusivity fields $a(y)$ with $d = 4$ and $d = 16$ parameters, the two measurement spaces of dimensions $m = 4$ and $m = 16$, and the four different ellipticity/coercivity regimes of c_ℓ in $a(y)$. We also plot the error when taking for k^* the *oracle value* that corresponds to the value of k that contains the parameter y which gave rise to the snapshot and measurement.

Our main findings can be summarized as follows:

(i) The error decreases with the number of splits. As anticipated, the splitting strategy is more effective in the overdetermined regime $m \geqslant d$.

(ii) Degrading coercivity has a negative effect on the estimation error, while anisotropy has a positive effect. In our computations, a larger r in c_ℓ corresponds to a higher degree of anisotropy and in turn to a reduced width of the solution manifold $\mathcal{M}$ in dimensions associated to the less active coordinates. Hence, it is no surprise that the approximation errors from our algorithm are lower for these higher anisotropy

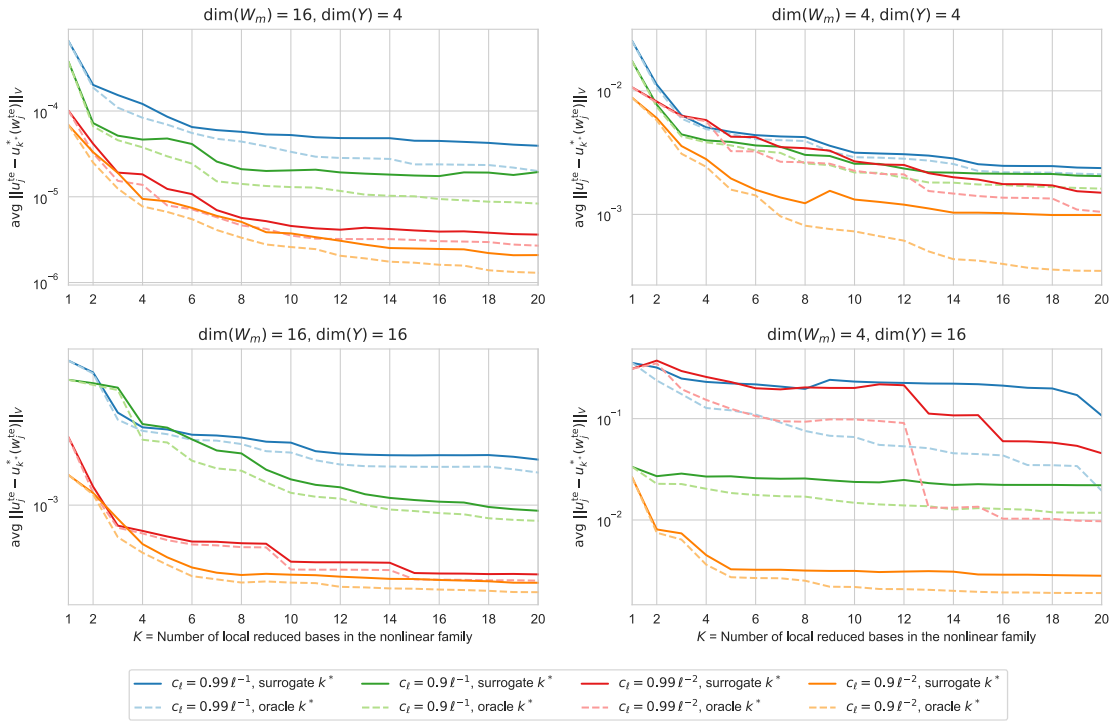


Figure 5.5. *Average of errors* $\|u_j^{\mathrm{te}} - u_{k*}^*(w_j^{\mathrm{te}})\|$ *for different choices of* k^*.

examples.

(iii) Choosing k^* by the surrogate based model selection yields error curves that are above yet close to those obtained with the oracle choice. The largest discrepancy is observed when coercivity degrades.

Figure 5.6 presents the error bounds $\sigma_K := \max_{k=1,\dots,K} \mu_k \varepsilon_k$ which are known to be upper bounds for the estimation error when choosing the oracle value for k^* at the given step K of the splitting procedure. We observe that these worst upper bounds have behavior similar to that of the averaged error curves depicted in Figure 5.5. In Figure 5.7, for the particular configuration $\dim(Y) = \dim(W) = 16$, we demonstrate that σ_K indeed acts as an upper bound for the worst case error of the oracle estimator.

5.3. Test 3: Improving the state estimate by alternating residual minimization. The goal of this test is to illustrate how the alternating residual minimization outlined in section 4.2 allows one to improve the accuracy of the state estimate. We use the same setting as Test 2; in particular, we consider the solution manifold $\mathcal{M}$ of (5.1) with the random field $a(y)$ defined as in (5.4). Again we consider the cases where the D_ℓ from (5.4) are from the 2×2 and 4×4 grid, resulting in $d = \dim(Y) = 4$ and 16, respectively. Our test uses all three measurement regimes presented in Figure 5.4, with $m = 4$ and 16 evenly spaced local average functions and $m = 8$ randomly placed local averages confined to the upper half. We use the coercivity/anisotropy regime $c_\ell = 0.9\,\ell^{-1}$.

In this test, we compare three different candidates for u^0, the starting point of the alternating minimization procedure:

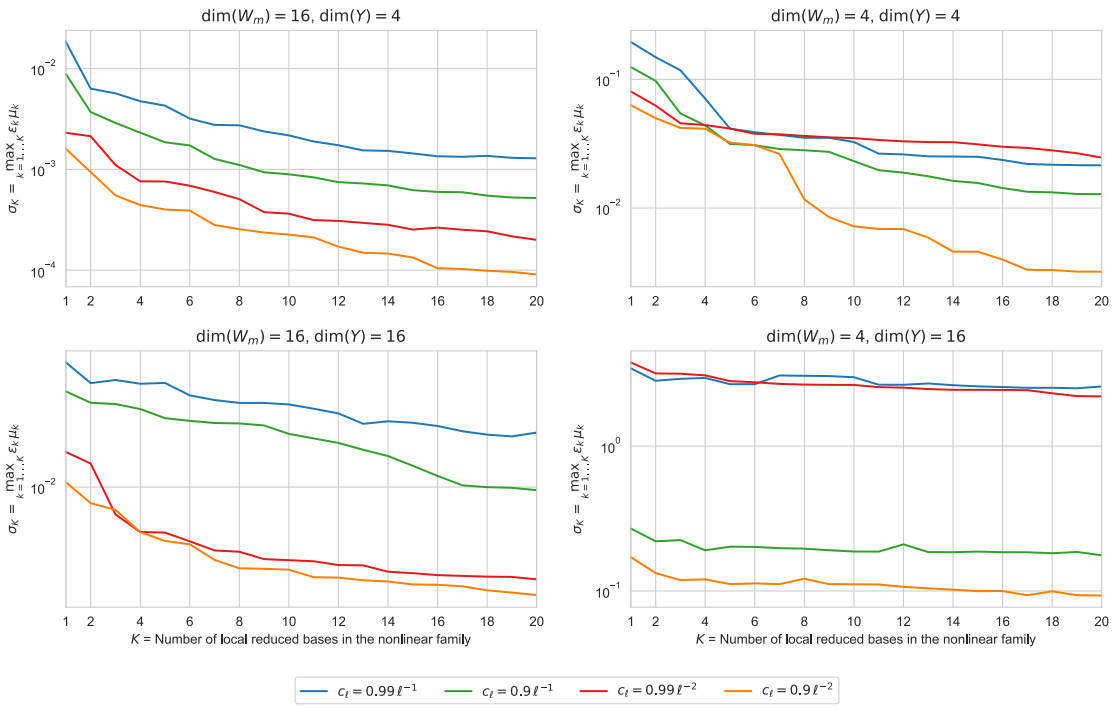


Figure 5.6. *Error bounds of local linear families, given by* $\sigma_K = \max_{k=1,\dots,K} \mu_k \varepsilon_k$.

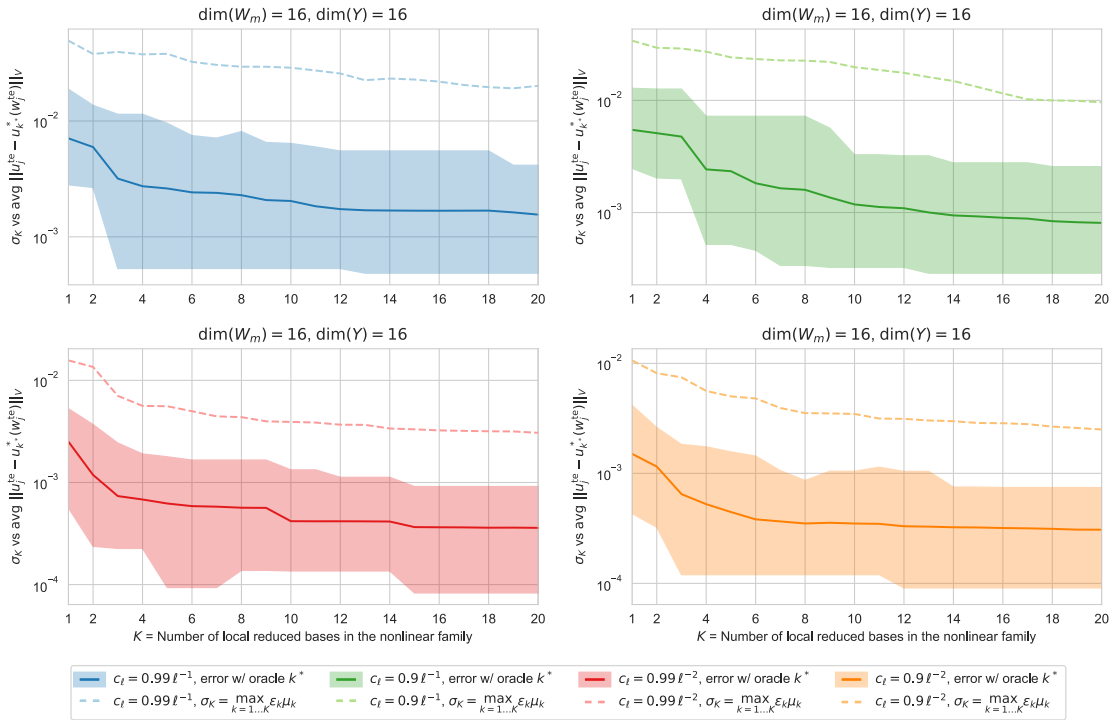


Figure 5.7. *Comparison between* σ_K *(dashed curve), the averaged oracle error (full curve), and the range from maximum to minimum oracle error (shaded region).*

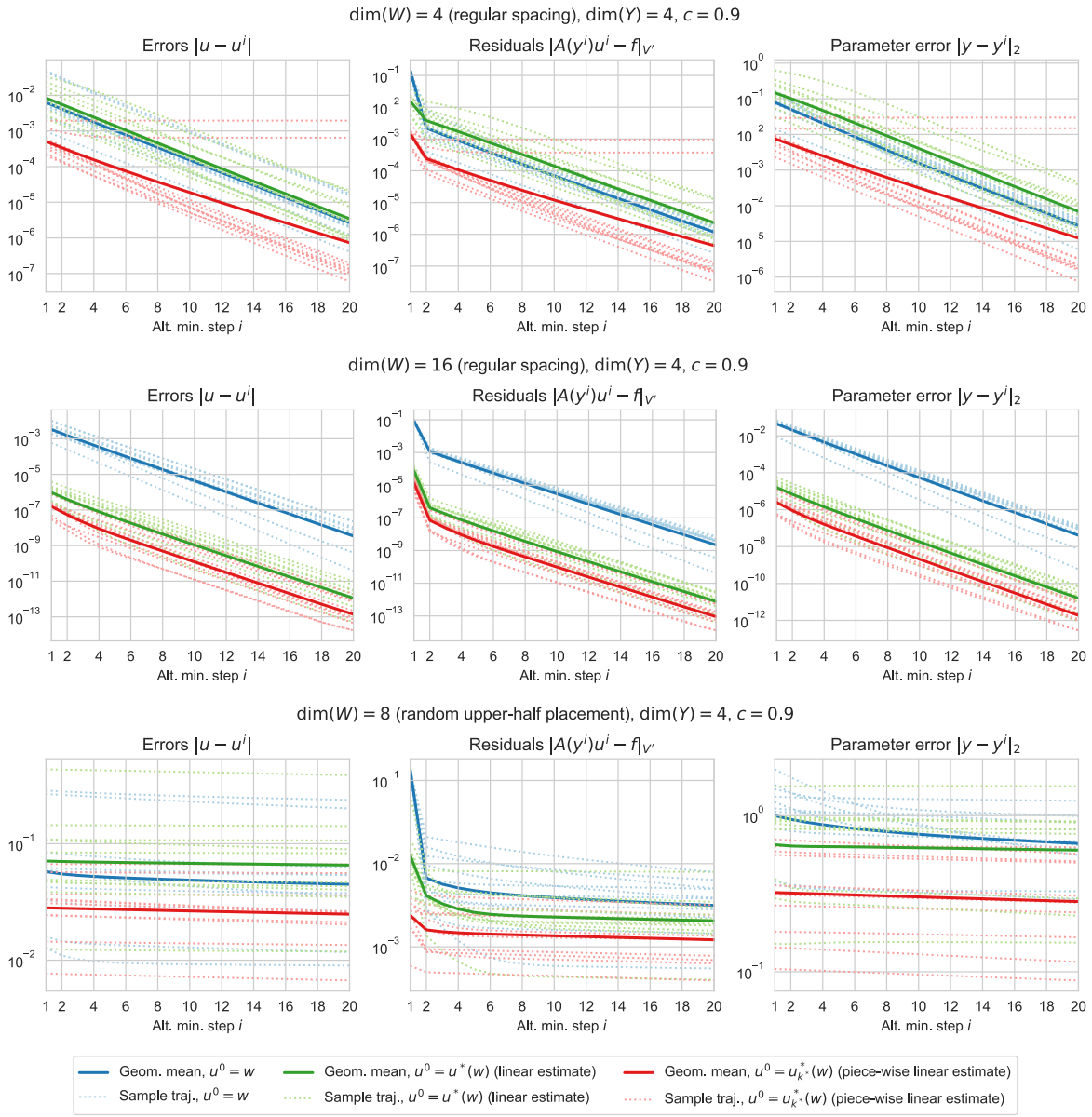


Figure 5.8. *State error (left plots), residual error (middle plots), and parameter error (right plots) during the iterations of the joint state-parameter estimation. The dashed lines show individual iterations for a target snapshot. The full lines are their geometric average. The color green, blue, or red corresponds to different starting guesses for* u^0.

- $u^0 = w$, the measurement vector without any further approximation, or equivalently the reconstruction of minimal H_0^1 norm among all functions that agree with the observations.
- $u^0 = u^*(w)$, the PBDW state estimation using the greedy basis over the whole man-

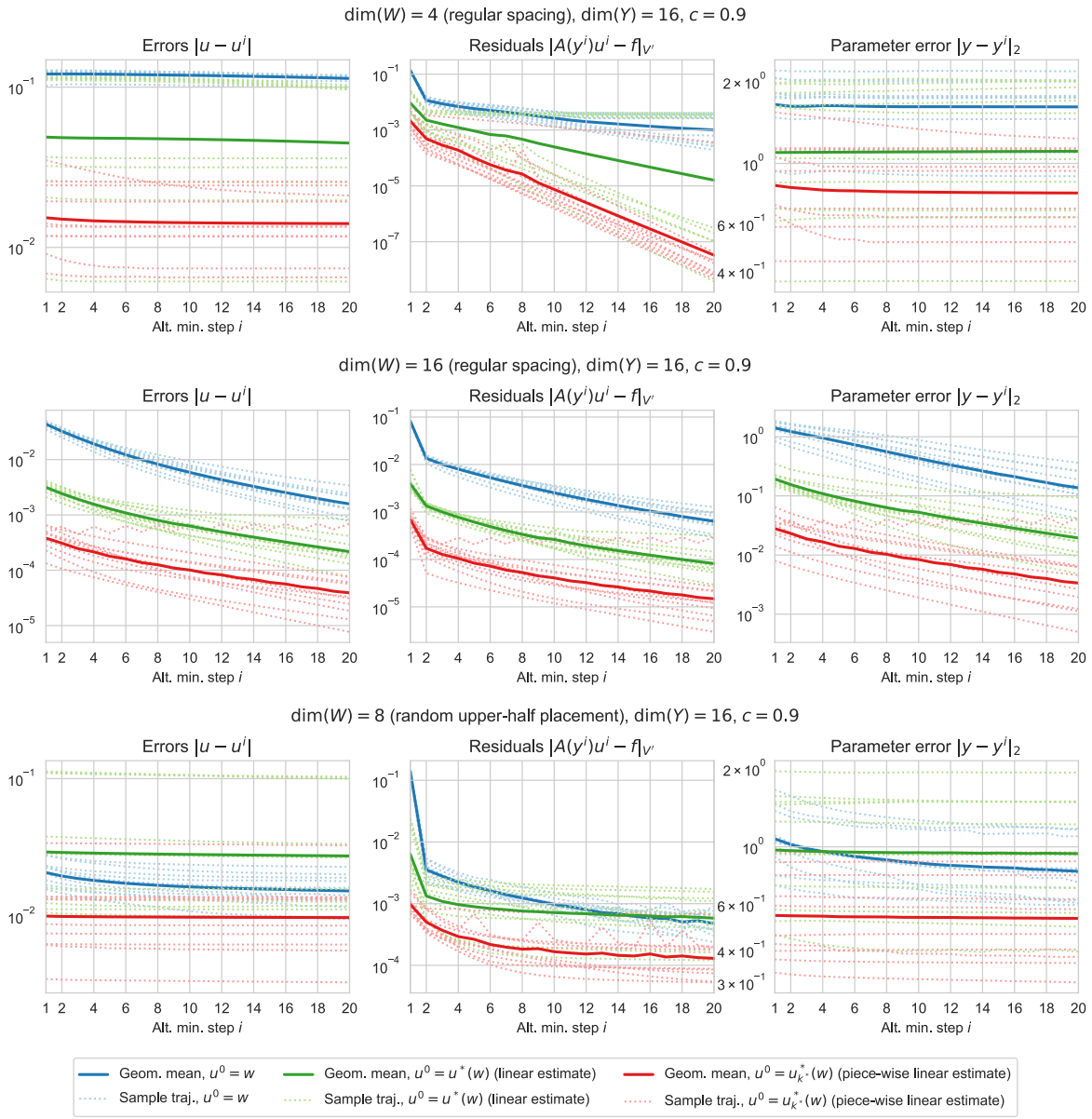


Figure 5.9. *State error (left plots), residual error (middle plots), and parameter error (right plots) during the iterations of the joint state-parameter estimation. The dashed lines show individual iterations for a target snapshot. The full lines are their geometric average. The color green, blue, or red corresponds to different starting guesses for* u^0.

ifold, thus starting the minimization from a "lifted" candidate that we hope is closer to the manifold $\mathcal{M}$ and should thus offer better performance.

- $u^0 = u^*_{k*}(w)$, the surrogate-chosen local linear reconstruction from the same family of local linear models from section 5.2 (where k^* is the index of the chosen local linear

model). In this last case, we take $K = 20$ local linear models, i.e., where Y has been split 19 times.

Furthermore, in the third case, we restrict our parameter range to be the local parameter range chosen by the surrogate; that is, we alter the step outlined in (4.15) to be

$$y^{j+1} = \operatorname*{argmin}_{y \in Y_{k*}} \mathcal{R}(u^j, y),$$

where y^{j+1} denotes the parameter found at the $(j+1)$th step of the procedure. The hope is that we have correctly chosen the local linear model and restricted parameter range from which the true solution comes thanks to our local model selection. The alternating minimization will thus have a better starting position and then a faster convergence rate due to the restricted parameter range.

In our test, we use the same training set $\widetilde{\mathcal{M}}$ as in the previous test, with $N_{\text{tr}} = 5000$ samples, in order to generate the reduced basis spaces. We consider a set of $N_{\text{te}} = 10$ snapshots, distinct from any snapshots in $\widetilde{\mathcal{M}}$, and perform the alternate minimization for each of the snapshots in the test set.

Figures 5.8 and 5.9 above display the state error trajectories $j \mapsto \|u - u^j\|$ for each snapshot (dashed lines), as well as their geometric average (full lines), in different colors depending on the initialization choice. Similarly, we display the residual trajectories $j \mapsto \mathcal{R}(u^j, y^j) = \|A(y^j)u^j - f\|_{V'}$ and parameter error trajectories $j \mapsto |y - y^j|_2$. Our main findings can be summarized as follows:

(i) In all cases, there is a substantial gain in taking $u^0 = u^*_{k*}(w)$, the surrogate-chosen local linear reconstruction, as the starting point. In certain cases, the iterative procedure initiated from the two other choices w or $u^*(w)$ stagnates at an error level that is even higher than $\|u - u^*_{k*}(w)\|$.

(ii) The state error, residual error, and parameter error decrease to zero in the overdetermined configurations where $(\dim(W), \dim(Y))$ is $(4,4)$, $(16,16)$, or $(16,4)$, with equally spaced measurement sites. In the underdetermined configurations $(4,16)$, the state and parameter errors stagnate, while the residual error decreases to zero, which reflects the fact that there are several $(y,u) \in Y \times w + W^\perp$ satisfying $\mathcal{R}(y,u) = 0$, making the fundamental barrier δ_0 strictly positive.

(iii) The state error, residual error, and parameter error do not decrease to zero in the overdetermined configuration $(\dim(W), \dim(Y)) = (8,4)$ where the measurement sites are concentrated on the upper-half of the domain. This case is interesting since, while we may expect that there is a unique pair $(y, u(y)) \in Y \times w + W^\perp$ reaching the global minimal value $\mathcal{R}(y,u) = 0$, the algorithm seems to get trapped in local minima.

REFERENCES

[1] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano, *A parametrized-background data-weak approach to variational data assimilation: Formulation, analysis, and application to acoustics*, Internat. J. Numer. Methods Engrg., 102 (2015), pp. 933–965, https://doi.org/10.1002/nme.4747.

[2] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Data assimilation in reduced modeling*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 1–29, https://doi.org/10.1137/15M1025384.

[3] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced basis methods*, SIAM J. Math. Anal., 43 (2011), pp. 1457–1472, https://doi.org/10.1137/100795772.

[4] P. Binev, A. Cohen, O. Mula, and J. Nichols, *Greedy algorithms for optimal measurements selection in state estimation using reduced models*, SIAM J. Uncertain. Quantif., 43 (2018), pp. 1101–1126, https://doi.org/10.1137/17M1157635.

[5] A. Bonito, A. Cohen, R. DeVore, G. Petrova, and G. Welper, *Diffusion coefficients estimation for elliptic partial differential equations*, SIAM J. Math. Anal, 49 (2017), pp. 1570–1592, https://doi.org/10.1137/16M1094476.

[6] A. Bonito, A. Cohen, R. DeVore, D. Guignard, P. Jantsch, and G. Petrova, *Nonlinear methods for model reduction*, ESAIM Math. Model. Numer. Anal., 55 (2021), pp. 507–531.

[7] D. Braess, *Finite Elements, Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press, Cambridge, UK, 1997.

[8] D. Broersen and R. P. Stevenson, *A Petrov-Galerkin discretization with optimal test space of a mild-weak formulation of convection-diffusion equations in mixed form*, IMA J. Numer. Anal., 35 (2015), pp. 39–73, https://doi.org/10.1093/imanum/dru003.

[9] C. Carstensen, L. Demkowicz, and J. Gopalakrishnan, *Breaking spaces and forms for the DPG method and applications including Maxwell equations*, Comput. Math. Appl., 72 (2016), pp. 494–522, https://doi.org/10.1016/j.camwa.2016.05.004.

[10] A. Cohen, W. Dahmen, and G. Welper, *Adaptivity and variational stabilization for convection-diffusion equations*, ESAIM Math. Model. Numer. Anal., 46 (2012), pp. 1247–1273, https://doi.org/10.1051/m2an/2012003.

[11] A. Cohen, W. Dahmen, R. DeVore, and J. Nichols, *Reduced basis greedy selection using random training sets*, ESAIM Math. Model. Numer. Anal., 54 (2020), pp. 1509–1524, https://doi.org/10.1051/m2an/2020004.

[12] A. Cohen, W. Dahmen, R. DeVore, J. Fadili, O. Mula, and J. Nichols, *Optimal reduced model algorithms for data-based state estimation*, SIAM J. Numer. Anal., 58 (2020), pp. 3355–3381, https://doi.org/10.1137/19M1255185, 2020.

[13] A. Cohen and R. DeVore, *Approximation of high dimensional parametric PDEs*, Acta Numer., 24 (2015), pp. 1–159, https://doi.org/10.1017/S0962492915000033.

[14] A. Cohen, R. DeVore, and C. Schwab, *Analytic regularity and polynomial approximation of parametric stochastic elliptic PDEs*, Anal. Appl., 9 (2011), pp. 11–47, https://doi.org/10.1142/S0219530511001728.

[15] A. C. Lorenc, *A global three-dimensional multivariate statistical interpolation scheme*, Mon. Weather Rev., 109 (1981), pp. 701–721.

[16] C. Lorenc, *Analysis methods for numerical weather prediction*, Q. J. R. Meteorol. Soc., 112 (1986), pp. 1177–1194.

[17] R. Everson and L. Sirovich, *Karhunen–Loeve procedure for gappy data*, J. Opt. Soc. Amer. A, 12 (1995), pp. 1657–1664.

[18] K. Willcox, *Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition*, Comput. & Fluids, 35 (2006), pp. 208–226.

[19] B. Adcock, A. C. Hansen, and C. Poon, *Beyond consistent reconstructions: Optimality and sharp bounds for generalized sampling, and application to the uniform resampling problem*, SIAM J. Math. Anal., 45 (2013), pp. 3132–3167, https://doi.org/10.1137/120895846.

[20] 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, Springer INdAM Ser. 4, Springer, Milan, 2013, pp. 221–235.

[21] Y. Maday, O. Mula, A. T. Patera, and M. Yano, *The generalized empirical interpolation method: Stability theory on Hilbert spaces with an application to the Stokes equation*, Comput. Methods Appl. Mech. Engrg., 287 (2015), pp. 310–334.

[22] Y. Maday, O. Mula, and G. Turinici, *Convergence analysis of the generalized empirical interpolation method*, SIAM J. Numer. Anal., 54 (2016), pp. 1713–1731, https://doi.org/10.1137/140978843.

[23] A. Bonito, A. Cohen, R. DeVore, D. Guignard, P. Jantsch, and G. Petrova, *Nonlinear methods for model reduction*, ESAIM Math. Model. Numer. Anal., 55 (2021), pp. 507–531.

[24] D. Amsallem, M. J. Zahr, and C. Farhat, *Nonlinear model order reduction based on local reduced-order bases*, Internat. J. Numer. Methods Engrg., 92 (2012), pp. 891–916.
[25] B. Peherstorfer, D. Butnaru, K. Willcox, and H.-J. Bungartz, *Localized discrete empirical interpolation method*, SIAM J. Sci. Comput., 36 (2014), pp. A168–A192, https://doi.org/10.1137/130924408.
[26] K. Carlberg, *Adaptive h-refinement for reduced-order models*, Internat. J. Numer. Methods Engrg., 102 (2015), pp. 1192–1210.
[27] D. Amsallem and B. Haasdonk, *PEBL-ROM: Projection-error based local reduced-order models*, Adv. Model. Simul. Eng. Sci., 3 (2016), 6.
[28] W. Dahmen, C. Huang, C. Schwab, and G. Welper, *Adaptive Petrov–Galerkin methods for first order transport equations*, SIAM J. Numer. Anal., 50 (2012), pp. 2420–2445, https://doi.org/10.1137/110823158.
[29] W. Dahmen, C. Plesken, and G. Welper, *Double greedy algorithms: Reduced basis methods for transport dominated problems*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 623–663, https://doi.org/10.1051/m2an/2013103.
[30] R. DeVore, G. Petrova, and P. Wojtaszczyk, *Greedy algorithms for reduced bases in Banach spaces*, Constr. Approx., 37 (2013), pp. 455–466, https://doi.org/10.1007/s00365-013-9186-2.
[31] M. Dashti and A. M. Stuart, *The Bayesian approach to inverse problems*, in Handbook of Uncertainty Quantification, R. Ghanem, D. Higdon, and H. Owhadi, eds., Springer, Cham, 2017, pp. 311–428, https://doi.org/10.1007/978-3-319-12385-1_7.
[32] J. L. Eftang, A. T. Patera, and E. M. Ronquist, *An "hp" certified reduced basis method for parametrized elliptic partial differential equations*, SIAM J. Sci. Comput., 32 (2010), pp. 3170–3200, https://doi.org/10.1137/090780122.
[33] Y. Maday, A. T. Patera, and G. Turinici, *Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic partial differential equations*, C. R. Math. Acad. Sci. Paris, 335 (2002), pp. 289–294, https://doi.org/10.1016/S1631-073X(02)02466-4.
[34] Y. Maday and B. Stamm, *Locally adaptive greedy approximations for anisotropic parameter reduced basis spaces*, SIAM J. Sci. Comput., 35 (2013), pp. A2417–A2441, https://doi.org/10.1137/120873868.
[35] P. Massart, *Concentration Inequalities and Model Selection*, Springer, Berlin, 2007, https://doi.org/10.1007/978-3-540-48503-2.
[36] N. A. Routledge, *A result in Hilbert space*, Quart. J. Math. Oxford Ser. (2), 3 (1952), pp. 12–18.
[37] G. Rozza, D. B. P. Huynh, and A. T. Patera, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: Application to transport and continuum mechanics*, Arch. Comput. Methods Eng., 15 (2008), pp. 229–275, https://doi.org/10.1007/s11831-008-9019-9.
[38] S. Sen, *Reduced-basis approximation and a posteriori error estimation for many-parameter heat conduction problems*, Numer. Heat Tr. B-Fund, 54 (2008), pp. 369–389. https://doi.org/10.1080/10407790802424204.
[39] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559, https://doi.org/10.1017/S0962492910000061.
[40] R. P. Stevenson and J. Westerdiep, *Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations*, IMA J. Numer. Anal., 41 (2021), pp. 28–47, https://doi.org/10.1093/imanum/drz069.
[41] V. Temlyakov, *Nonlinear Kolmogorov widths*, Math. Notes, 63 (1998), pp. 785–795.
[42] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, SIAM Rev., 60 (2018), pp. 550–591, https://doi.org/10.1137/16M1082469.