Coarse reduced model selection for nonlinear state estimation

J. A. Nichols[1]

(Received 14 January 2021; revised 13 December 2021)

Abstract

State estimation is the task of approximately reconstructing a solution $\mathfrak{u}$ of a parametric partial differential equation when the parameter vector $\mathfrak{y}$ is unknown and the only information is $\mathfrak{m}$ linear measurements of $\mathfrak{u}$. Cohen et al. [arXiv:2009.02687, Nov. 2020] proposed a method to use a family of linear reduced spaces as a generalised nonlinear reduced model for state estimation. A computable *surrogate distance* is used to evaluate which linear estimate lies closest to a true solution of the PDE problem. In this article we propose a strategy of coarse computation of the surrogate distance while maintaining a fine mesh reduced model, as the computational cost of the surrogate distance is large relative to the reduced modelling task. We demonstrate numerically that the error induced by the coarse distance is dominated by other approximation errors.

DOI:10.21914/anziamj.v62.16169, Published 2022-02-07, as part of the Proceedings of the 19th Biennial Computational Techniques and Applications Conference. ISSN 1445-8810. (Print two pages per sheet of paper.)

Contents

1 Introduction

Complex physical systems are often modelled by a parametric partial differential equation (PDE). We consider the general problem of the form

$$A(y)u = f(y)\,, \tag{1}$$

where $A(y) : \mathcal{H} \to \mathcal{H}'$ is an elliptic second order operator and $\mathcal{H}$ is an appropriate Hilbert space. The problem is defined on a physical domain $D \subset \mathbb{R}^{2,3}$, and the parameter y is within a parameter domain $\mathcal{Y}$, typically a subset of $\mathbb{R}^d$ where d might be large. Each parameter $y \in \mathcal{Y}$ is assumed to result in a unique solution $u(x, y)$ but we mostly suppress the spatial dependence $x \in D$ and write $u(y)$. We denote all $\mathcal{H}$-norms by $\|\cdot\| := \|\cdot\|_{\mathcal{H}}$ and inner-products $\langle\cdot,\cdot\rangle := \langle\cdot,\cdot\rangle_{\mathcal{H}}$, and use an explicit subscript if they are from a different space.

In practical modelling applications it is often computationally expensive to produce a high precision numerical solution to the PDE problem (1). However, to our advantage the mapping $u : y \to \mathcal{H}$ is typically smooth and compact, hence the set of solutions over all parameter values will be a smooth manifold,

possibly of finite intrinsic dimension. We define the *solution manifold* as follows, assuming from here that it is compact,

$$\mathcal{M} := \{\mathfrak{u}(\mathfrak{y}) : \mathfrak{y} \in \mathcal{Y}\}.$$

The methods proposed in this article extend on *reduced basis* approximations. A reduced basis is a linear subspace $\mathsf{V} \subset \mathcal{H}$ of moderate dimension $\mathfrak{n} := \dim(\mathsf{V}) \leqslant \dim(\mathcal{M}) \leqslant \dim(\mathcal{H})$. We use the *worst-case error* as a benchmark, defined as

$$\varepsilon(\mathsf{V}, \mathcal{M}) = \max_{\mathfrak{u} \in \mathcal{M}} \operatorname{dist}(\mathfrak{u}, \mathsf{V}) = \max_{\mathfrak{u} \in \mathcal{M}} \|\mathfrak{u} - \mathsf{P}_{\mathsf{V}}\mathfrak{u}\|, \tag{2}$$

where P_{V} is the orthogonal projection operator onto V. There are a variety of methods to construct a reduced basis with desirable worst-case error performance, and here we concentrate on *greedy methods* that select points in $\mathcal{M}$ which become the basis for V. We discuss these methods further in Section 1.1.

A reduced basis can be used to accelerate the *forward problem*. One can numerically solve the PDE problem for a given parameter $\mathfrak{y}$ by using V directly in the Galerkin method, making the numerical problem vastly smaller while retaining a high level of accuracy. A thorough treatment of the development of reduced basis approximations is given by Hesthaven, Rozza, and Stamm [5].

In this article we are concerned with *inverse problems*. In this setting it is assumed that there is some unknown true state $\mathfrak{u}$ (which could correspond to the state of some physical system), and we do not know the parameter vector $\mathfrak{y}$ that gives this solution. Instead, we make do with a handful of $\mathfrak{m}$ linear measurements $\ell_i(\mathfrak{u})$. These measurements are used to make some kind of accurate reconstruction of $\mathfrak{u}$ (*state estimation*) or a guess of the true parameter $\mathfrak{y}$ (*parameter estimation*).

The *parametrized background data weak* (PBDW) approach introduced by Maday et al. [6] gives a straightforward procedure for finding an estimator $\mathfrak{u}^*$ of the true state $\mathfrak{u}$, using only the linear measurement information and a

reduced basis V. However, one limitation is that the estimator error $\|u^* - u\|$ is bounded from below by the *Kolmogorov* n*-width* given by

$$d_n(\mathcal{M}) = \inf_{\dim(V)=n} \sup_{u \in \mathcal{M}} \operatorname{dist}(u, V), \tag{3}$$

where the infimum is taken over all n dimensional linear spaces in $\mathcal{H}$. The n-width is known to converge slowly for many parametric PDE problems.

We review methods for constructing a family of *local linear* reduced models and a nonlinear estimator u^* using a surrogate distance model selection procedure. We propose the use of coarser finite element meshes to perform this selection. This coarse selection strategy is motivated by the observation of Cohen et al. [3] that the model selection is by far the most computationally costly component in the nonlinear estimation routine. Finally, in Section 4 we examine numerical examples of surrogate distances over different mesh widths, and see that they make insubstantial impacts to the nonlinear estimator.

1.1 Linear reduced models

A reduced basis is a linear space of the form $V = \operatorname{span}(u^1, \ldots, u^n) \subset \mathcal{H}$, where the $u^i = u(y^i) \in \mathcal{M}$. The parameter values y^i are typically chosen in some iterative greedy procedure to try and minimise $\varepsilon(V, \mathcal{M})$ at each step.

We define a greedy procedure as follows: given V of dimension n and a finite subset $\widetilde{\mathcal{Y}}$ of $\mathcal{Y}$, to produce an $(n+1)$-dimensional reduced space we find the parameter $y^{n+1} \in \widetilde{\mathcal{Y}}$ which gives us the largest $\operatorname{dist}(u(y^{n+1}), V) = \|u(y^{n+1}) - P_V u(y^{n+1})\|$. This parameter y^{n+1} can be found through a brute-force search of the finite set $\widetilde{\mathcal{Y}}$. We then augment the space V with $u^{n+1} = u(y^{n+1})$. This simple strategy, in some cases [1], yields a reduced basis that is optimal with regards to the Kolmogorov n-width of $\mathcal{M}$. In this setting the quantity

$$\varepsilon_{\text{est}}(V, \mathcal{M}) := \max_{y \in \widetilde{\mathcal{Y}}} \|u(y) - P_V u(y)\| \tag{4}$$

serves as a reasonable and calculable estimate of $\varepsilon(V, \mathcal{M})$, as shown by Cohen et al. [4].

In practice, we also allow V to be an *affine* space, with an offset $\bar{u}$, such that

$$V = \bar{u} \oplus \operatorname{span}(u^1, \ldots, u^n)\,.$$

Typically we take $\bar{u}$ to be an approximate barycenter of $\mathcal{M}$. From now on we use the term *reduced space* or *model* rather than basis.

With V chosen, we now consider the state estimation problem. We have m pieces of linear data $\ell_i(u)$ for $i = 1, \ldots, m$ of the unknown state u, and the $\ell_i \in \mathcal{H}'$. We assume we know the *form* of the functionals ℓ_i and hence the Riesz representers ω_i for which $\langle \omega_i, u \rangle = \ell_i(u)$. These ω_i define a *measurement space* and the measurement vector of u, respectively,

$$W := \operatorname{span}(\omega_1, \ldots, \omega_m) \quad \text{and} \quad w = P_W u\,.$$

Note here we assume no noise in our measurements, but allowing for random noise is straightforward and has been considered by Maday et al. [6] and Binev et al. [2].

The PBDW approach, developed by Maday et al. [6], seeks a reconstruction candidate or estimator $u^*(w)$ that is close to u, but that agrees with the measurement data, that is $P_W u^*(w) = P_W u = w$. Maday et al. [6] define an estimator

$$u^*(w) = \operatorname*{arg\,min}_{v \in w + W_\perp} \operatorname{dist}(v, V)\,,$$

which can be calculated through a set of normal equations of size $n \times m$ using the cross-Gramian matrix of the bases of W and V. Given only the measurement information w, the measurement space W, and the reduced space V, this estimator $u^*(w)$ is an optimal choice [2]. The estimator lies in the subspace $u^*(w) \in V \oplus W$.

We require that $W^\perp \cap V = \{0\}$ for this reconstruction algorithm to be well posed, as otherwise there are infinitely many candidates for u^*. This in turn requires that $n = \dim(V) \leqslant \dim(W) = m$, since $\dim(V) > \dim(W)$ implies that there must be at least one vector in V that is perpendicular to W. This

dimensionality requirement is reflected in the error analysis. We define an inf-sup constant

$$\mu(V, W) := \max_{v \in V} \frac{\|v\|}{\|P_W v\|}, \tag{5}$$

which is the inverse of the cosine of the angle between V and W and we have $\mu(V, W) \in [1, \infty]$. For μ to be finite, we require $W^\perp \cap V = \{0\}$. The inf-sup constant plays the role of a stability constant for our linear estimator as we have the well known bound

$$E_{wc} = \max_{u \in \mathcal{M}} \|u - u^*(P_{W_m} u)\| \leqslant \mu(V, W)\, \varepsilon(V, \mathcal{M}),$$

as demonstrated by Binev et al. [2]. From the definitions we have $\varepsilon(V, \mathcal{M}) \geqslant d_{n+1}(\mathcal{M})$, hence this reconstruction error can at best be the $(n+1)$-width of $\mathcal{M}$.

2 Nonlinear reduced models

A fundamental drawback of linear reduced models is the slow decay of the Kolmogorov n-width for a wide variety of PDE problems. To circumvent this limitation, a framework for non-linear reduced models and their use for state estimation was presented by Cohen et al. [3]. The proposal involves determining a partition of the manifold

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

and producing a family of affine reduced space approximations V_k to each portion $\mathcal{M}_k$. Each space has dimension $n_k = \dim(V_k)$, requiring $n_k < m$ for well-posedness.

Given any target $\varepsilon > 0$ and $\mu \geqslant 1$ it is possible with large enough K to determine a partition and family of reduced spaces V_k that satisfy

$$\varepsilon_k := \varepsilon(V_k, \mathcal{M}_k) \leqslant \varepsilon \quad \text{and} \quad \mu_k := \mu(V_k, W) \leqslant \mu \quad \text{for all } k = 1, \ldots, K, \tag{6}$$

in which case we say the family $(V_k)_{k=1}^K$ is (μ, ε)-admissible. A slightly looser criteria on the partition can also be satisfied: given some $\sigma > 0$, we say the family $(V_k)_{k=1}^K$ is σ-*admissible* if $\mu_k \varepsilon_k \leqslant \sigma$ for all $k = 1, \dots, K$. The existence of both (μ, ε) and σ-admissible families follow from the compactness of $\mathcal{M}$; Cohen et al. [3] give a full demonstration.

In practice one may construct a σ-admissible family in the following way. Say we are given a partition of the parameter space $(\mathcal{Y}_k)_{k=1}^{K-1}$ with $\cup_{k=1}^{K-1} \mathcal{Y}_k = \mathcal{Y}$, and the associated partition of the manifold $(\mathcal{M}_k)_{k=1}^{K-1}$. We have a reduced space V_k approximating $\mathcal{M}_k$, produced by a greedy algorithm on a finite sample set $\tilde{\mathcal{Y}}_k \subset \mathcal{Y}_k$. With each V_k we have an associated error estimate $\sigma_{\text{est},k} = \mu_k(V, W)\varepsilon_{\text{est},k}(V, \mathcal{M})$. We pick the largest $\sigma_{\text{est},k}$, with index $\tilde{k}$ say, and we split the cell $\mathcal{Y}_{\tilde{k}}$ in half for each parameter coordinate direction $i \in \{1, \dots, d\}$, resulting in two reduced spaces $V^+_{\tilde{k},i}$ and $V^-_{\tilde{k},i}$ for each split direction. We take the split direction i to be the one with the smallest maximum error $\max(\sigma^+_{\tilde{k},i}, \sigma^-_{\tilde{k},i})$, and we enrich the family $(V_k)_{k=1}^{K-1}$ with the two new reduced spaces, making sure to remove $V_{\tilde{k}}$ from the collection. More details of the splitting procedure are provided by Cohen et al. [3].

2.1 Surrogate reduced model selection

Based on a measurement w, each affine reduced space has an associated reconstruction candidate that is found through the PBDW method

$$u_k^*(w) := \arg\min\{\text{dist}(v, V_k) : P_W v = w\} \quad \text{for } k = 1, \dots, K. \tag{7}$$

If we happen to know that the true state u originates from some $\mathcal{M}_k$, then we would best use u_k^* as our estimator. In this scenario we would have an error bound of $\|u - u_k^*(P_W u)\| \leqslant \varepsilon_k \mu_k < \sigma$. This information about the true state u is not available in practice, so we require some other method to determine which candidate u_k^* to choose.

Consider a surrogate distance $\mathcal{S}(v, \mathcal{M})$ from v to $\mathcal{M}$ that satisfies the uniform bound for $0 < r \leqslant R$,

$$r\,\text{dist}(v, \mathcal{M}) \leqslant \mathcal{S}(v, \mathcal{M}) \leqslant R\,\text{dist}(v, \mathcal{M}). \tag{8}$$

If this surrogate distance is computable, then we can use it to find a *surrogate selected* estimator by choosing

$$k^* := \arg\min\{\mathcal{S}(u_k^*, \mathcal{M}) : k = 1, \dots, K\}, \text{ and define } u^*(w) := u_{k^*}^*(w)\,, \quad (9)$$

noting that as there is a dependence on w we write $k^*(w)$. We define an error benchmark

$$\delta_\sigma := \sup\{\|u - v\| : \operatorname{dist}(u, \mathcal{M}), \operatorname{dist}(v, \mathcal{M}) \leqslant \sigma, P_W u = P_W v\}\,.$$

This quantity takes in to account errors from model bias, and we have the following result.

Theorem 1. *Given a σ-admissible family of affine reduced spaces $(V_k)_{k=1}^K$, the estimator based on the surrogate selection (9) has worst-case error bounded above by*

$$\max_{u \in \mathcal{M}} \|u - u^*(P_W u)\| \leqslant \delta_{\kappa\sigma}\,, \quad (10)$$

where $\kappa = R/r$ depends only on the uniform bounds of the surrogate distance.

The proof is detailed in Theorem 3.2 by Cohen et al. [3]. Note that even given some optimal nonlinear reconstruction algorithm, our best possible error would be δ_0, and it is not bounded from below by $d_n(\mathcal{M})$. We remark also that $\delta_\sigma \geqslant \sigma$.

3 Affine elliptic operators

Say the operator $A(y)$ in (1) is uniformly bounded in y with uniformly bounded inverse. That is, for some $0 < r \leqslant R < \infty$ we have

$$\|A(y)\|_{\mathcal{H} \to \mathcal{H}'} \leqslant R \quad \text{and} \quad \|A(y)^{-1}\|_{\mathcal{H}' \to \mathcal{H}} \leqslant r^{-1}\,, \quad y \in \mathcal{Y}\,. \quad (11)$$

Then we can show that for any $v \in \mathcal{H}$ the residual of the PDE

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

satisfies the uniform bound $r\|v-u(y)\| \leqslant \mathcal{R}(v,y) \leqslant R\|v-u(y)\|$. We define

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

and then we arrive at a surrogate distance that satisfies the uniform bounds of (8). Using this surrogate in the selection (9) to define $u^*(w)$, we have a nonlinear reconstruction algorithm with the error guarantees of (10).

We now make the further assumption that the operator $A(y)$ and source term $f(y)$ have affine dependence on y. That is,

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

The residual is calculated using representers in $\mathcal{H}$. We define e_j as member of $\mathcal{H}$ that satisfies

$$\langle e_j, z\rangle = \langle A_j(y)v - f_j, z\rangle_{\mathcal{H}',\mathcal{H}} \quad \text{for all } z\in\mathcal{H}\,, \tag{12}$$

and now write $e(y) := e_0 - \sum_{j=1}^{d} y_j e_j$ to denote the representer of the overall residual problem. The residual is equal to $\mathcal{R}(v,y) = \|e(y)\|$, and determining the surrogate distance is a quadratic minimisation problem

$$\mathcal{S}(v,\mathcal{M})^2 = \min_{y\in\mathcal{Y}} \|e(y)\|^2 = \min_{y\in\mathcal{Y}} \Big\|e_0 + \sum_{j=1}^{d} y_j e_j\Big\|^2.$$

This leads to a constrained least squares problem that can be solved using standard optimisation routines, using the values $\langle e_i, e_j\rangle$ for $0 \leqslant i,j \leqslant d$.

Technically it is possible to have a $A(y)$ and $f(y)$ with nonlinear dependence on y, but this comes at the cost of having to solve a nonlinear least squares problem to evaluate $\mathcal{S}(v,\mathcal{M})$. In certain model cases this is a convex optimisation problem, for example the PDE examined in Section 4 if $a(x,y)$ is uniformly convex in y.

3.1 Finite element residual evaluation

In practice these calculations take place in a finite element space $\mathcal{H}_h$ that is the span of polynomial elements on a triangulation $\mathcal{T}_h$ of width $h > 0$. In this setting the residual is $\mathcal{R}_h(v, y) = \|e_h(y)\|$ where $e_h(y) = e_{0,h} - \sum_{j=1}^{d} y_j e_{j,h}$ and the $e_{j,h} \in \mathcal{H}_h$ satisfy the variational problem

$$\langle e_{j,h}, z\rangle = \langle A_j(y)v - f_j, z\rangle_{\mathcal{H}',\mathcal{H}} \quad \text{for all } \in \mathcal{H}_h\,. \tag{13}$$

Naturally we define $\mathcal{S}_h(v, \mathcal{M}) := \min_{y\in\mathcal{Y}} \mathcal{R}_h(v, y)$. Note that when we subtract (12) from (13) we obtain $\langle e_h(y) - e(y), z\rangle = 0$ for all $z \in \mathcal{H}_h$, meaning that $e_h(y) - e(y) \perp \mathcal{H}_h$.

We write $y^* = \arg\min_{y\in\mathcal{Y}} \mathcal{R}(v, y)$ to be the minimiser selected in $\mathcal{S}(v, \mathcal{M})$, and y_h^* the equivalent for $\mathcal{S}_h(v, \mathcal{M})$. In general $y^* \neq y_h^*$, but we have

$$\begin{aligned} |\mathcal{S}(v, \mathcal{M}) - \mathcal{S}_h(v, \mathcal{M})| &= |\mathcal{R}(v, y^*) - \mathcal{R}_h(v, y_h^*)| \\ &\leqslant |\mathcal{R}(v, y^*) - \mathcal{R}_h(v, y^*)| + |\mathcal{R}(v, y_h^*) - \mathcal{R}_h(v, y_h^*)| \\ &\leqslant \|e(y^*) - e_h(y^*)\| + \|e(y_h^*) - e_h(y_h^*)\|\,, \end{aligned} \tag{14}$$

where in the last step we used

$$|\mathcal{R}(v, y) - \mathcal{R}_h(v, y)| = \big|\|e(y)\| - \|e_h(y)\|\big| \leqslant \|e(y) - e_h(y)\|\,.$$

Thus the convergence of $\mathcal{S}_h$ to $\mathcal{S}$ depends on the finite element convergence of solutions e_h to e. This convergence is determined by the regularity of $e(y)$, which depends on the smoothness of the data $A(y)v - h$ and the so called Riesz lift implied in the variational problem (12).

Recall that $A(y)$ in (1) is a second order symmetric elliptic operator. If we assume homogeneous Dirichlet boundary conditions on ∂D and that $f(y) \in L^2(D)$, then a natural choice for our ambient space is the Sobolev space $\mathcal{H} = H_0^1(D)$, with $\|\cdot\| = \|\cdot\|_{H_0^1(D)}$. In this setting the solutions $e(y)$ of (12) are the weak solutions of the Poisson problem with homogeneous Dirichlet boundary conditions

$$\nabla_x^2 e(y) = A(y)v - f(y) \text{ on } D, \quad \text{with } e(y) = 0 \text{ on } \partial D, \tag{15}$$

where we have written ∇_x^2 to denote the Laplacian in the spatial variables, recalling that we have the unwritten spatial dependence in $e(y) = e(x, y)$.

In our surrogate model selection, $e(y)$ depends on $u_k^*(w)$. This estimator $u_k^*(w)$ lies in $V_k \oplus W$. As V_k is the span of some solutions selected from $\mathcal{M}_k$, the smoothness of $u_k^*(w)$ will depend on the smoothness of all solutions of the PDE problem and of the measurement functionals ω_i.

The convergence of e_h to e in the weak form of (15) is well known in a wide variety of settings, for example we have the classical result

$$\|e_h(y) - e(y)\| \leqslant ch\|A(y)v - f(y)\|_{L^2(D)}, \tag{16}$$

which is applicable in a wide variety of situations. Under these circumstances we thus have that $|\mathcal{S}(v, \mathcal{M}) - \mathcal{S}_h(v, \mathcal{M})| \sim h$.

3.2 Coarse surrogate evaluation

Say we construct a family of reduced spaces $(V_k)_{k=1}^K$ where each family $V_k = \overline{u}_k + \text{span}(u_{h'}^1, \ldots, u_{h'}^{n_k})$, and the solutions $u_{h'}^i$ are numerically calculated with respect to a triangulation $\mathcal{T}_{h'}$ with mesh-width h'. Our estimators $u_k^*(w)$ will be in $\mathcal{H}_{h'}$.

We consider $\mathcal{T}_{h'}$ our *fine* mesh, and nominate another triangulation $\mathcal{T}_h$ with $h' < h$ to be a *coarse* mesh. We use this coarse mesh to compute the surrogate distance, noting that $\mathcal{S}_h(v, \mathcal{M})$ may be orders of magnitude faster to compute than the fine mesh equivalent. This will necessarily introduce an inaccuracy in the surrogate selection; however, we maintain the high fidelity of the fine mesh reduced space approximations V_k.

If $\mathcal{T}_{h'}$ is a fine mesh that contains $\mathcal{T}_h$ in the sense that $\mathcal{H}_h \subset \mathcal{H}_{h'}$, then the variational problem (13) is straightforward as we can calculate the inner-product $\langle A(y)v - f(y), z\rangle$ based on the known relationships between the basis elements of $\mathcal{H}_h$ and $\mathcal{H}_{h'}$. Furthermore, the error $\|e_h(y) - e_{h'}(y)\|$, and hence $|\mathcal{S}_h(v, \mathcal{M}) - \mathcal{S}_{h'}(v, \mathcal{M})|$, is bounded above by a constant times $\|e(y) - e_h(y)\|$ through the same reasoning as in (14). If $\mathcal{T}_{h'}$ does not contain $\mathcal{T}_h$, then the

same error estimates still apply, but we will require interpolation schemes to construct the variational problem.

4 Numerical tests

We present two separate studies of the surrogate as evaluated in finite element spaces. On the unit square $D = [0,1]^2$ we consider the PDE

$$-\nabla_x \cdot (a(x,y)\nabla_x u(x,y)) = 1\,, \quad \text{with } u(x,y) = 0 \text{ on } \partial D\,,$$

where our parametric diffusivity field is given by $a(x,y) = 1+\sum_{j=1}^{16} c_j y_j \chi_{D_j}(x)$. Here the D_j denote squares of side length $1/4$ that subdivide the unit square in to 16 portions, and χ_{D_j} is the indicator function on D_j. The parameter range is the hypercube $\mathcal{Y} = [-1,1]^{16}$, and the coefficients are $c_j = 0.9j^{-1}$ or $c_j = 0.99j^{-1}$, meaning that $a(x,y) > 0$, but when $c_j = 0.99j^{-1}$ the $a(x,y)$ become closer to zero and the PDE problem can lose ellipticity.

We perform a space discretisation by the Galerkin method using $\mathbb{P}_1$ finite elements to produce solutions $u_{h'}(y)$, with a fine triangulation on a regular grid of mesh size $h' = 2^{-7}$. In the tests that follow, we evaluate $\mathcal{S}_h(v,\mathcal{M})$ with coarse meshes of size $h = 2^{-s}$ with $s = 2,\dots,6$.

We generate training sets to compute the reduced models, and test sets on which we test the reconstruction algorithms. The training set $\widetilde{\mathcal{M}}^{\mathrm{tr}}$ is the collection of PDE solutions for $N_{\mathrm{tr}} = 1000$ random samples $\widetilde{\mathcal{Y}}^{\mathrm{tr}} = \{y_j^{\mathrm{tr}}\}_{j=1,\dots,N_{\mathrm{tr}}}$ drawn independently and uniformly on $\mathcal{Y} = [-1,1]^{16}$. The test set $\widetilde{\mathcal{M}}^{\mathrm{te}}$ is created from $N_{\mathrm{te}} = 100$ independent parameter samples that are distinct from the training set samples.

In our test the measurement space W is a collection of $m = 8$ measurement functionals $\langle \omega_i, u\rangle = \ell_i(u) = |B_i|^{-1}\int u\chi_{B_i}$ that are local averages in small areas B_i which are boxes of width $2h = 2^{-6}$, each placed randomly in the unit square.

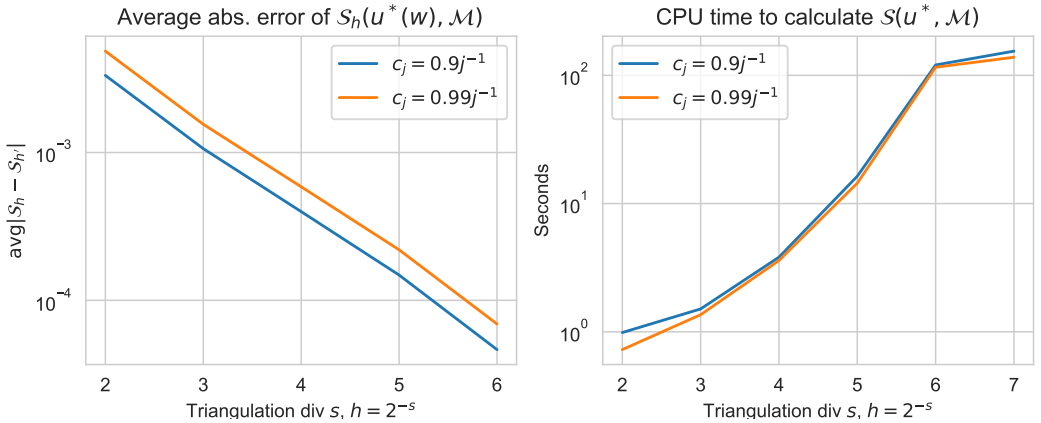

Figure 1: Left: average absolute error of the surrogate $\mathcal{S}_h$; Right: the CPU wall time for the computation of $\mathcal{S}_h$ for all N_{te} test points

In our first test we compute a single reduced space V using the greedy procedure on $\widetilde{\mathcal{M}}^{tr}$. For each test candidate $u \in \widetilde{\mathcal{M}}^{te}$ we calculate $u^*(P_W u)$. In Figure 1 we plot the difference between the coarse and fine surrogate distances $|\mathcal{S}_h(u^*, \mathcal{M}) - \mathcal{S}_{h'}(u^*, \mathcal{M})|$. This error is significantly dominated by the calculated value of the estimator error (4), $\sigma_{est} := \mu(V, W)\, \varepsilon_{est}(V, \mathcal{M})$ as we found $\sigma_{est} \approx 10^0$ for $c_j = 0.9j^{-1}$ and $\sigma_{est} \approx 10^1$ for $c_j = 0.99j^{-1}$. This dominates the errors presented in Figure 1 significantly.

We observe a linear relationship in the right plot of Figure 1. The linear best fit of $\log(\mathrm{avg}_{\widetilde{\mathcal{M}}^{te}}|\mathcal{S}_h - \mathcal{S}_{h'}|)$ and $\log(h) = -s$ has slope 1.51, which implies that on average

$$|\mathcal{S}_h - \mathcal{S}_{h'}| \sim h^{1.51}.$$

This exponent slightly exceeds the theoretically guaranteed exponent of one given in Section 3.1, and is likely due to the extra regularity of f in this example.

For the second test we examine the impact of using the coarse surrogate $\mathcal{S}_h(u_k^*, \mathcal{M})$ for model selection. We build the σ-admissible families as outlined in Section 2 using the greedy splitting of $\mathcal{Y} = [-1, 1]^{16}$. At each point the cells $\mathcal{Y}_k$ are

Table 1: Agreement of coarse model selection out of $N^{te} = 100$ test points $u \in \widetilde{\mathcal{M}}^{te}$.

	$c_j = 0.9j^{-1}$					$c_j = 0.99j^{-1}$				
Mesh width h	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}
$\#\{k^*_h(w) = k^*_{h'}(w)\}$	97	100	100	100	100	88	94	96	98	100
$\#\{k^*_h(w) = k^*_{\text{true}}(u)\}$	74	77	77	77	77	58	59	61	65	64

rectangular, which we split in half in coordinate directions. We split the parameter space seven times, resulting in $K = 8$ local reduced spaces.

For each test candidate $u \in \widetilde{\mathcal{M}}^{te}$ we have K possible reconstructions $u^*_1(w), \ldots, u^*_K(w)$. We use the coarse surrogate in the model selection (9), writing $k^*_h(w) = k^*_h(P_W u)$ to make the dependence on h and w clear, and we inspect how often it agrees with the *fine* selection $k^*_{h'}(w)$ for all test points u. We also compare this to the 'true' selection $k^*_{\text{true}}(u)$ for which $u \in \mathcal{M}_{k^*_{\text{true}}}$.

Table 1 demonstrates that $k^*_h(w)$ agrees with $k^*_{h'}$ the vast majority of time from $h = 2^{-4}$ onwards, for both cases of c_j. We also see that the fine selection $k^*_{h'}$ agrees with the true selection 77 times out of 100 for $c_j = 0.9j^{-1}$ and 64 times for $c_j = 0.99j^{-1}$. That is, it picks the estimator $u^*_k(w)$ that is trained on the portion of manifold $\mathcal{M}_k$ that u originated from. Figure 2 plots the histogram of selections $k^*_{h'}$ and $k^*_{\text{true}}(u)$, recalling that they are a number in $\{1, \ldots, 8\}$. We see broadly similar patterns in the reduced model selection. Given the CPU time savings that we see in Figure 1, we conclude that model selection through a coarse surrogate distance is a worthwhile numerical strategy.

References

[1] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. "Convergence rates for greedy algorithms in reduced

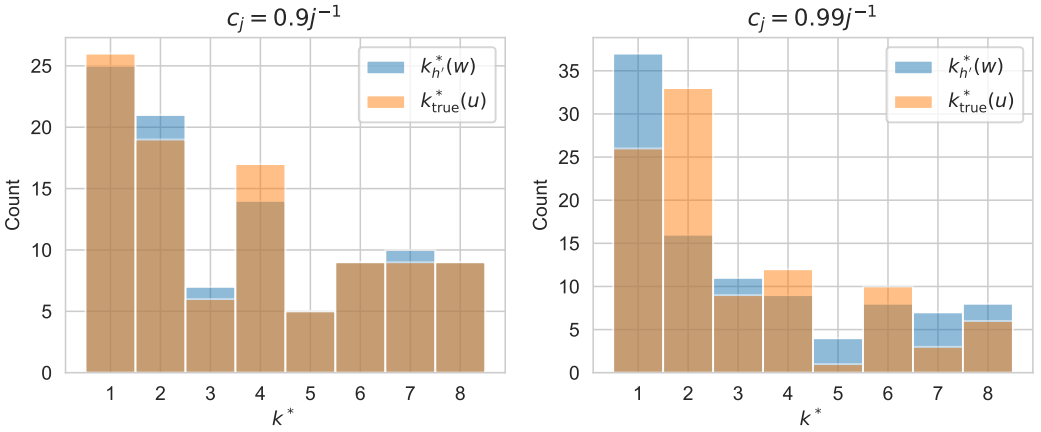

Figure 2: A histogram of the number of fine surrogate selections $\mathrm{k}^*_{\mathrm{h}'}$ and true reduced model selection $\mathrm{k}^*_{\mathrm{true}}(\mathrm{u})$.

basis methods". In: *SIAM J. Math. Anal.* 43.3 (2011), pp. 1457–1472. DOI: 10.1137/100795772. (Cit. on p. C195).

[2] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. "Data assimilation in reduced modeling". In: *SIAM/ASA J. Uncert. Quant.* 5.1 (2017), pp. 1–29. DOI: 10.1137/15M1025384 (cit. on pp. C196, C197).

[3] A. Cohen, W. Dahmen, O. Mula, and J. Nichols. "Nonlinear reduced models for state and parameter estimation". In: *arXiv:2009.02687 [cs, math]* (2020). URL: http://arxiv.org/abs/2009.02687 (visited on 01/07/2021) (cit. on pp. C195, C197, C198, C199).

[4] A. Cohen, D. Wolfgang, R. DeVore, and J. Nichols. "Reduced basis greedy selection using random training sets". In: *ESAIM: Math. Model. Num. Anal.* 54 (2020), pp. 1509–1524. DOI: 10.1051/m2an/2020004. (Cit. on p. C195).

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

Mathematics. Springer, 2016. DOI: 10.1007/978-3-319-22470-1. (Cit. on p. C194).

[6] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano. "A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics". In: *Int. J. Num. Meth. Eng.* 102.5 (2015), pp. 933–965. DOI: 10.1002/nme.4747. (Cit. on pp. C194, C196).

Author address

1. **J. A. Nichols**, Biological Data Science Institute, Australian National University, Acton ACT 0200, AUSTRALIA
mailto:james.nichols@anu.edu.au