

Model Reduction from Partial Observations

C. Herzet⁽¹⁾, P. Héas⁽¹⁾, A. Drémeau⁽²⁾

⁽¹⁾ INRIA, Campus de Beaulieu, 35000 Rennes, France

⁽²⁾ ENSTA Bretagne, 42 rue François Verny, 29200 Brest, France

This paper deals with model-order reduction of parametric partial differential equations (PPDE). More specifically, we consider the problem of finding a good approximation subspace of the solution manifold of the PPDE when only partial information on the latter is available. We assume that two sources of information are available: *i*) a “rough” prior knowledge, taking the form of a manifold containing the target solution manifold; *ii*) partial linear measurements of the solutions of the PPDE (the term partial refers to the fact that observation operator cannot be inverted). We provide and study several tools to derive good approximation subspaces from these two sources of information. We first identify the best worst-case performance achievable in this setup and propose simple procedures to approximate the corresponding optimal approximation subspace. We then provide, in a simplified setup, a theoretical analysis relating the achievable reduction performance to the choice of the observation operator and the prior knowledge available on the solution manifold.

1 Introduction

Our contribution takes place within the context of model reduction for parametric partial differential equations:

$$\text{PDE}(\mathbf{h}, \theta) = 0, \quad (1)$$

where \mathbf{h} belongs to a Hilbert space \mathcal{H} and $\theta \in \Theta$ is a parameter. When the solution $\mathbf{h}(\theta)$ of (1) has to be evaluated for many different values $\theta \in \Theta$, the computational effort may become prohibitive. To circumvent this issue, model reduction intends to simplify the resolution of (1) by (typically) constraining \mathbf{h} to belong to some low-dimensional subspace $S \subset \mathcal{H}$. As a matter of fact, the choice of S should be made so that each element of the solution manifold $\mathcal{M} = \{\mathbf{h}(\theta) \in \mathcal{H} : \theta \in \Theta\}$ is well-approximated by

some element of S . Many techniques have been proposed in the literature to identify such subspaces: Taylor [1] or Hermite [2] expansions, proper orthogonal decomposition (POD) [3], balanced truncation [4], reduced basis techniques [5], etc.

All the methods mentioned above presuppose some refined knowledge of the solution manifold \mathcal{M} . For example, it is typically assumed that the solution manifold \mathcal{M} can be finely sampled [1–4] or, at least, that the set of parameters Θ defining \mathcal{M} is perfectly known [5]. Unfortunately, in practice a refined knowledge of \mathcal{M} may not always be available. Nevertheless, in many situations one may have access to some partial^a measurements of the elements of \mathcal{M} . The main question addressed in this paper is therefore as follows: can we benefit from these partial measurements to (complement our prior knowledge and) compute a good approximation subspace for \mathcal{M} ?

In order to provide a precise answer to this question, we assume in the rest of this paper that we have the following two ingredients at our disposal:

- i) a prior manifold $\mathcal{M}_{\text{prior}}$* , which collects all the knowledge we have “a priori” about \mathcal{M} . The only constraint we impose on $\mathcal{M}_{\text{prior}}$ is to be such that

$$\mathcal{M} \subseteq \mathcal{M}_{\text{prior}}, \quad (2)$$

i.e., the prior manifold must contain all the elements of the target manifold \mathcal{M} .

- ii) a set of partial observations of the elements of \mathcal{M}* : we assume that we collect, $\forall \mathbf{h} \in \mathcal{M}$, a set of noiseless linear measurements:

$$\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m, \quad (3)$$

where $\{\mathbf{w}_j\}_{j=1}^m$ is an orthonormal basis (ONB) of some subspace W and $\langle \cdot, \cdot \rangle$ denotes the inner product in \mathcal{H} .

The prior information typically derives from some physical considerations and/or constraints we may have about the system under study. We give an example of construction of $\mathcal{M}_{\text{prior}}$ in Section 4. The nature of the observations available in practice depends on the experimental setup. In this work, we make the assumption that the measurements can be seen as the outputs of a noiseless linear operator. The noisy setting is not considered hereafter

^aHere, the term “partial” refers to the fact that the measurement operator cannot be inverted.

and left for future work. We note that if $m = \dim(\mathcal{H})$, the observation operator can be inverted to exactly recover the elements of \mathcal{M} , leading us back to the standard setup. In the sequel, we will thus assume that $m < \dim(\mathcal{H})$, so that some ambiguity on \mathbf{h} subsists upon the observation of $\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m$.

Although of clear practical interest, only a few contributions have tackled the problem of model reduction from partial measurements. To the best of our knowledge, the first paper dealing with this question is due to Everson and Sirovich [6]. The authors proposed a methodology, dubbed ‘‘Gappy POD’’, constructing an approximation subspace when only some elements of each solution $\mathbf{h}(\theta)$ are observed. This approach has been applied with success to *e.g.*, oceanography in [7] or fluid mechanics in [8–10]. Unfortunately, Gappy POD requires some diversity in the observation operator to work properly: as noted in [9], this method is doomed to produce poor approximation subspaces as soon as some directions of \mathcal{H} are never observed.^b This is for example the case when all the elements of the solution manifold are observed through the *same* partial observation operator as in (3). In order to circumvent this issue, prior information about \mathcal{M} can be included in the reduction process. This approach was recently used in [11–15]. In [11], the authors suggested to iteratively enrich the approximation subspace by using ‘‘a posteriori’’ estimates of some elements of \mathcal{M} (the term ‘‘a posteriori’’ refers here to the fact that the estimates stem from the combination of partial observations and some prior knowledge on \mathcal{M}). In [12], the authors of the present work refined this approach in a Bayesian framework: they proposed to include the uncertainty inherent to the a posteriori estimates in the reduction process. An efficient implementation of this Bayesian reduction strategy relying on adaptive a posteriori sampling is proposed in [13]. The works [14, 15] are also closely related to this Bayesian perspective of model reduction: the authors determine an optimal low-dimensional approximation of the a posteriori distribution for solving efficiently inverse problems governed by PPDEs. In this paper, we present a rigorous formulation and justification to these approaches in a deterministic framework. In particular, we provide elements of answer to the following questions:

- i)* What is the best performance which can be achieved by combining the information provided by the prior manifold $\mathcal{M}_{\text{prior}}$ and the collected observations $\{\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m\}_{\mathbf{h} \in \mathcal{M}}$?
- ii)* Can we characterize this ideal performance as a function of the choice

^bIn fact, each direction of ‘‘large variation’’ of \mathcal{M} should be observed in at least one observation.

of the prior manifold $\mathcal{M}_{\text{prior}}$ and the measurement subspace W ?

- iii) How to compute a good approximation subspace from $\mathcal{M}_{\text{prior}}$ and $\{\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m\}_{\mathbf{h} \in \mathcal{M}}$ in practice?

As suggested by our above discussion, distinct from, but related to model reduction from partial measurements, is the question of deriving a good estimate of some unknown $\mathbf{h}(\theta)$ by exploiting both a reduced-order model and some collected data. This paradigm has recently been explored in several papers [11, 16–18]. In [16], the authors proposed a reduced version of a Kalman filter and showed that the error on the estimate delivered by the latter can be bounded by a function of the data residual. In [17] (resp. [11]), Maday *et al.* considered a data assimilation problem from noiseless (resp. noisy) linear observations, where the prior model is defined as a low-dimensional approximation subspace of \mathcal{M} . The same setup was discussed in [18] by Binev *et al.* and extended to priors defined as an intersection of degenerate ellipsoids. Some of the theoretical considerations exposed in the present work are built upon the arguments derived in that paper.

The rest of the paper is organized as follows. In Section 3 we show that the worst-case optimal performance achievable in our partially-informed setup is characterized by the “Kolmogorov width” of some well-defined set. In Section 4, we propose a simple practical scheme to approximate this worst-case optimal performance and illustrate its performance on the well-known “thermal-block problem” in Section 5. Finally, in Section 6 we provide a theoretical result relating the achievable reduction performance to the choice of the prior and observation operators. The main steps of the proof of this result are exposed in Section 7 whereas the technical details are postponed to Appendix A. In Appendix B, we discuss the weak and algebraic formulations of the PPDE used in our simulations. In order to ease the reading of our work, we also provide a summary of the main notations and expressions useful to understand the paper in Appendix C.

2 Notational Conventions

Except if otherwise stated, the notational conventions used in this paper are as follows. Italic lowercase boldface letters (as *e.g.*, \mathbf{h} , \mathbf{v} , \mathbf{w} , etc.) denote elements of the Hilbert space \mathcal{H} . Uppercase italic letters (*e.g.*, S , U , V , W , etc.) are used for subspaces of \mathcal{H} . Lowercase (*e.g.*, \mathbf{n} , \mathbf{x} , etc.) and uppercase boldface (*e.g.*, \mathbf{X} , \mathbf{S} , etc.) letters respectively stand for vector and matrix notations. Lowercase italic (*e.g.*, a , b , α , β , etc.) and uppercase normal (*e.g.*,

N, T, etc.) letters denote scalars. For matrices and vectors, the superscript \cdot^T means transposition. The element located at row i and column j of a matrix \mathbf{X} is denoted x_{ij} .

The inner product and the induced norm associated to \mathcal{H} are denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ respectively. The space orthogonal to a subspace X with respect to the inner product $\langle \cdot, \cdot \rangle$ is written as X^\perp . The subspace induced by a set $\{\mathbf{v}_i\}_{i=1}^n$ is denoted $\text{span}(\{\mathbf{v}_i\}_{i=1}^n)$. The distance between an element $\mathbf{h} \in \mathcal{H}$ and a closed subspace S is defined as

$$\text{dist}(\mathbf{h}, S) \triangleq \min_{\mathbf{h}' \in S} \|\mathbf{h} - \mathbf{h}'\|,$$

and the projection of \mathbf{h} onto S as

$$P_S(\mathbf{h}) \triangleq \arg \min_{\mathbf{h}' \in S} \|\mathbf{h} - \mathbf{h}'\|.$$

The notation $P_S(V)$ stands for the set $\{P_S(\mathbf{h}) : \mathbf{h} \in V\}$. The operator \oplus is used to denote the direct sum between two subsets of \mathcal{H} , *e.g.*, $S \oplus V = \{\mathbf{s} + \mathbf{v} : \mathbf{s} \in S, \mathbf{v} \in V\}$. Finally, $B_\epsilon = \{\mathbf{h} \in \mathcal{H} : \|\mathbf{h}\| \leq \epsilon\}$ is the $\|\cdot\|$ -ball of radius ϵ .

3 Worst-case Optimal Model Reduction

In this section, we tackle the problem of finding a good approximation subspace from a worst-case perspective. In Section 3.1, we first discuss the nature of the information provided by the prior manifold (2) and the collected observations (3), and identify the set of manifolds compatible with the latter. In Section 3.2, we then characterize the best worst-case performance achievable in our partially-informed setup.

3.1 Feasible and Posterior Manifolds

The prior manifold and the partial measurements provide some valuable information about the unknown manifold \mathcal{M} since they both define a set of constraints (discussed below) which are known to be satisfied by the latter. However, \mathcal{M} is usually not the only manifold verifying these constraints. In the sequel, we will denote the set of manifolds satisfying the constraints imposed by $\mathcal{M}_{\text{prior}}$ and the partial measurements $\{\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m\}_{\mathbf{h} \in \mathcal{M}}$ as Ξ_{feas} ; the elements of Ξ_{feas} will be referred to as “feasible” manifolds.

In the rest of this subsection, we give a precise characterization of Ξ_{feas} and emphasize that the largest^c element of this set is defined by

$$\mathcal{M}_{\text{post}} = \mathcal{M}_{\text{prior}} \cap (\cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}), \quad (4)$$

where

$$H_{\mathbf{h}} \triangleq \left\{ \mathbf{h}' = \mathbf{h} + \mathbf{w}^\perp : \mathbf{w}^\perp \in W^\perp \right\}.$$

This observation will turn out to be crucial in the next subsection to characterize the worst-case optimal approximation subspace.

Let us first discuss the constraints defining Ξ_{feas} . First, it is clear that any manifold $\tilde{\mathcal{M}}$ compatible with our prior assumption (2) should be such that

$$\tilde{\mathcal{M}} \subseteq \mathcal{M}_{\text{prior}}. \quad (5)$$

Secondly, a manifold $\tilde{\mathcal{M}}$ compatible with the received observations should reproduce exactly the same set of measurements as those obtained from \mathcal{M} when measured with the same observation operator. More specifically, any manifold $\tilde{\mathcal{M}}$ compatible with the received observations should be such that

$$\left\{ \left\{ \langle \mathbf{w}_j, \mathbf{h} \rangle \right\}_{j=1}^m \right\}_{\mathbf{h} \in \tilde{\mathcal{M}}} = \left\{ \left\{ \langle \mathbf{w}_j, \mathbf{h} \rangle \right\}_{j=1}^m \right\}_{\mathbf{h} \in \mathcal{M}}. \quad (6)$$

We thus define the set of feasible manifolds as

$$\Xi_{\text{feas}} = \left\{ \tilde{\mathcal{M}} : (5) \text{ and } (6) \text{ hold} \right\}. \quad (7)$$

In other words, Ξ_{feas} represents the set of manifolds which are compatible with both our prior assumption (2) and the set of collected observations. It is obvious from (5)-(7) that $\mathcal{M} \in \Xi_{\text{feas}}$.

We now prove the following lemma:

Lemma 1. $\mathcal{M}_{\text{post}}$ is the largest element of Ξ_{feas} .

Proof: We first show that

$$\begin{cases} \tilde{\mathcal{M}} \subseteq \mathcal{M}_{\text{prior}}, \\ \tilde{\mathcal{M}} \subseteq \cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}, \end{cases} \quad (8)$$

^cThe term ‘‘largest’’ must be understood as follows: if $\tilde{\mathcal{M}} \in \Xi_{\text{feas}}$ then $\tilde{\mathcal{M}} \subseteq \mathcal{M}_{\text{post}}$.

are necessary conditions for $\tilde{\mathcal{M}} \in \Xi_{\text{feas}}$ and, (8) together with

$$\mathcal{M} \subseteq \tilde{\mathcal{M}}, \quad (9)$$

is a sufficient condition for $\tilde{\mathcal{M}} \in \Xi_{\text{feas}}$. The proof of the main result will derive straightforwardly from these conditions.

The necessity of $\tilde{\mathcal{M}} \subseteq \mathcal{M}_{\text{prior}}$ is obvious from the definition of Ξ_{feas} . The necessity of $\tilde{\mathcal{M}} \subseteq \cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}$ can be shown as follows. First, note that the set of elements of \mathcal{H} leading to a given set of observations $\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m$ is an affine subspace defined as

$$\begin{aligned} H_{\mathbf{h}} &= \{\mathbf{h}' : \langle \mathbf{w}_j, \mathbf{h}' \rangle = \langle \mathbf{w}_j, \mathbf{h} \rangle \text{ for } j = 1, \dots, m\}, \\ &= \{\mathbf{h}' = \mathbf{h} + \mathbf{w}^\perp : \mathbf{w}^\perp \in W^\perp\}. \end{aligned}$$

Hence, if $\tilde{\mathcal{M}} \not\subseteq \cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}$, then $\exists \mathbf{h}' \in \tilde{\mathcal{M}}$ such that

$$\{\langle \mathbf{w}_j, \mathbf{h}' \rangle\}_{j=1}^m \notin \left\{ \{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m \right\}_{\mathbf{h} \in \mathcal{M}},$$

and therefore (6) cannot be satisfied. The necessity of $\tilde{\mathcal{M}} \subseteq \cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}$ is thus obtained by contraposition.

We now show that (8)-(9) is sufficient for $\tilde{\mathcal{M}} \in \Xi_{\text{feas}}$. Since the first condition in (8) is identical to (5), we only need to show that (8)-(9) implies (6). First, from our previous discussion, we have that $\tilde{\mathcal{M}} \subseteq \cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}$ implies that

$$\left\{ \{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m \right\}_{\mathbf{h} \in \tilde{\mathcal{M}}} \subseteq \left\{ \{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m \right\}_{\mathbf{h} \in \mathcal{M}}.$$

Moreover, if $\mathcal{M} \subseteq \tilde{\mathcal{M}}$ holds then we also have

$$\left\{ \{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m \right\}_{\mathbf{h} \in \mathcal{M}} \subseteq \left\{ \{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m \right\}_{\mathbf{h} \in \tilde{\mathcal{M}}}.$$

Combining the last two inclusions, we obtain (6).

We finally prove the statement of Lemma 1 by exploiting the necessary and sufficient conditions defined above. Since any $\tilde{\mathcal{M}} \in \Xi_{\text{feas}}$ must satisfy (8), we have

$$\tilde{\mathcal{M}} \subseteq \mathcal{M}_{\text{prior}} \cap (\cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}) = \mathcal{M}_{\text{post}}.$$

It thus remains to show that $\mathcal{M}_{\text{post}} \in \Xi_{\text{feas}}$. Now, $\mathcal{M}_{\text{post}}$ verifies (8) by definition. Moreover, (9) also holds because \mathcal{M} is included in both $\mathcal{M}_{\text{prior}}$

and $\cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}$. Hence, $\mathcal{M}_{\text{post}} \in \Xi_{\text{feas}}$ by virtue of the sufficiency of (8)-(9). \square

$\mathcal{M}_{\text{post}}$ summarizes the uncertainty about the unknown manifold \mathcal{M} by gathering the information provided by the prior model and the partial measurements. In particular, it is the smallest subset of \mathcal{H} containing all the manifolds compatible with the prior constraint (2) and the received observations. In the sequel, we will thus refer to $\mathcal{M}_{\text{post}}$ as the ‘‘posterior’’ manifold because of its analogy with the posterior probabilities defined in a Bayesian framework: both characterize the uncertainty remaining on some quantity of interest upon the combination of some prior and observation models [19].

3.2 Worst-case Optimal Model Reduction

In the model-reduction literature, an ideal figure of merit to assess the reducibility of a manifold \mathcal{M} is its ‘‘Kolmogorov i -width’’:

$$\kappa_i(\mathcal{M}) = \inf_{S: \dim(S)=i} \left(\sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S) \right). \quad (10)$$

It is clear from its definition that $\kappa_i(\mathcal{M})$ provides the best worst-case error achievable by any approximation subspace of dimension i . If the infimum of (10) can be attained^d, a worst-case optimal approximation subspace is thus given by

$$S_i^{\text{perf}} \in \arg \min_{S: \dim(S)=i} \left(\sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S) \right). \quad (11)$$

The resolution of this problem obviously entails the knowledge of the manifold to reduce, *i.e.*, \mathcal{M} . In the setup considered in this paper, which presupposes that \mathcal{M} is unknown, computing an approximation subspace according to (11) is therefore not possible.

Nevertheless, as discussed in the previous subsection, the presence of prior information and partial measurements on the unknown manifold \mathcal{M} reduces the uncertainty about its localization in \mathcal{H} . More specifically, any manifold compatible with the prior information and the received observations must belong to the feasible set Ξ_{feas} . A sensible approach, followed hereafter, then consists of including this information within the worst-case criterion

^dIn order to ease the discussion, we will always assume hereafter that all the suprema and infima exist and can be attained. We refer the reader to [20] for a detailed discussion on the conditions ensuring the existence of the extremizers of the Kolmogorov i -width.

used to evaluate the approximation subspace. More precisely, we suggest to compute the approximation subspace as the solution of

$$S_i^{\text{post}} \in \arg \min_{S: \dim(S)=i} \left(\sup_{\tilde{\mathcal{M}} \in \Xi_{\text{feas}}} \sup_{\mathbf{h} \in \tilde{\mathcal{M}}} \text{dist}(\mathbf{h}, S) \right). \quad (12)$$

Problem (12) is tantamount to finding the approximation subspace minimizing the maximum approximation error over all the feasible manifolds. We note that since the posterior manifold $\mathcal{M}_{\text{post}}$ defined in (4) is the largest element of Ξ_{feas} (see Lemma 1), (12) can also be rewritten as

$$S_i^{\text{post}} \in \arg \min_{S: \dim(S)=i} \left(\sup_{\mathbf{h} \in \mathcal{M}_{\text{post}}} \text{dist}(\mathbf{h}, S) \right). \quad (13)$$

By definition of S_i^{post} and since $\mathcal{M} \subseteq \mathcal{M}_{\text{post}} \subseteq \mathcal{M}_{\text{prior}}$, we have

$$\kappa_i(\mathcal{M}) \leq \sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S_i^{\text{post}}) \leq \kappa_i(\mathcal{M}_{\text{post}}) \leq \kappa_i(\mathcal{M}_{\text{prior}}). \quad (14)$$

The first inequality follows from the definition the Kolmogorov i -width of \mathcal{M} , the second from $\mathcal{M} \subseteq \mathcal{M}_{\text{post}}$ and the last one is a consequence of $\mathcal{M}_{\text{post}} \subseteq \mathcal{M}_{\text{prior}}$.

In the light of (14), we see that in the partially-informed setup considered in this paper, the optimal reduction performance is lower bounded by $\kappa_i(\mathcal{M})$ and upper bounded by $\kappa_i(\mathcal{M}_{\text{post}})$. The gap between $\kappa_i(\mathcal{M})$ and $\kappa_i(\mathcal{M}_{\text{post}})$ “materializes” the loss of reducibility which can occur by working in a partially-informed setting rather than a perfectly-informed one. We also note that $\kappa_i(\mathcal{M}_{\text{prior}})$ characterizes the best achievable worst-case performance when the prior constraint (2) is the only information available to the practitioner (*i.e.*, there are no observations). More precisely we have

$$\inf_{S: \dim(S)=i} \left(\sup_{\tilde{\mathcal{M}} \subseteq \mathcal{M}_{\text{prior}}} \sup_{\mathbf{h} \in \tilde{\mathcal{M}}} \text{dist}(\mathbf{h}, S) \right) = \inf_{S: \dim(S)=i} \left(\sup_{\mathbf{h} \in \mathcal{M}_{\text{prior}}} \text{dist}(\mathbf{h}, S) \right) = \kappa_i(\mathcal{M}_{\text{prior}}).$$

As expected, from a worst-case perspective, there is thus always a gain in exploiting the received observations on top of the prior information; the gain brought by the former is characterized by the gap between $\kappa_i(\mathcal{M}_{\text{post}})$ and $\kappa_i(\mathcal{M}_{\text{prior}})$. In Section 6, we will discuss more specifically the connections between $\kappa_i(\mathcal{M})$, $\kappa_i(\mathcal{M}_{\text{post}})$ and $\kappa_i(\mathcal{M}_{\text{prior}})$ in a simplified setup.

Algorithm 1 Greedy algorithm

inputs: $\{\mathbf{h}_j\}_{j=1}^{N_{\text{snap}}}$
init: $\hat{S} = \{\mathbf{0}\}$
while Stopping criterion is not satisfied **do**
 Compute $\mathbf{h} = \arg \max_{\mathbf{h}' \in \{\mathbf{h}_j\}_{j=1}^{N_{\text{snap}}}} \|\mathbf{h}' - P_{\hat{S}}(\mathbf{h}')\|$
 Set $\hat{S} = \hat{S} \oplus \text{span}(\mathbf{h})$
end while
output: \hat{S}

4 Practical Implementation: Greedy Procedure and Sampling Schemes

Solving (13) is typically an intractable problem. Suboptimal approaches have therefore to be considered to find good approximated solutions. Interestingly, since (13) shares exactly the same structure as (11) (with the difference that the supremum is taken over $\mathcal{M}_{\text{post}}$ rather than \mathcal{M}), one can benefit from the numerous suboptimal techniques which have been proposed in the standard setup to tackle our partially-informed problem. In the sequel, we will focus more specifically on the standard “greedy” procedure described in Algorithm 1, see *e.g.*, [21, Section 7.1.1]. The stopping criterion mentioned in the procedure may be for example the dimension of the approximation subspace or some accuracy requirements.

The algorithm presupposes that a set of elements of the manifold to reduce (commonly referred to as “snapshots” in the literature), say $\{\mathbf{h}_j\}_{j=1}^{N_{\text{snap}}}$, is available. At each iteration, the procedure increases the dimension of the approximation subspace, by including one of the snapshots leading to the largest projection error. The performance of the greedy algorithm thus depends on the choice of $\{\mathbf{h}_j\}_{j=1}^{N_{\text{snap}}}$. A sensible choice, that we will follow hereafter, consists of drawing the snapshots randomly from some distribution supported on $\mathcal{M}_{\text{post}}$. We note that

$$\begin{aligned}\mathcal{M}_{\text{post}} &= \mathcal{M}_{\text{prior}} \cap (\cup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}), \\ &= \cup_{\mathbf{h} \in \mathcal{M}} (\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}),\end{aligned}$$

and therefore a sampling of $\mathcal{M}_{\text{post}}$ can be achieved by drawing snapshots from some distribution supported on $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}, \forall \mathbf{h} \in \mathcal{M}$.

The structure of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ depends obviously on the choice of $\mathcal{M}_{\text{prior}}$. From a practical point of view, some choices of $\mathcal{M}_{\text{prior}}$ may thus enable more

favorable implementations than others. In Section 4.1, we introduce some specific choices for $\mathcal{M}_{\text{prior}}$ which lead to a simple mathematical characterization of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$. We emphasize that, in many situations, this type of priors can be obtained quite easily by applying standard reduced-order model techniques. In Section 4.2, we present some material and notations needed to properly characterize $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$. In Section 4.3, we emphasize that there exist simple schemes to sample $\mathcal{M}_{\text{post}}$ for the specific choices of $\mathcal{M}_{\text{prior}}$ introduced in Section 4.1. Our approach relies on the derivations of Binev *et al.* showing that $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ corresponds to a high-dimensional ellipsoid with orthogonal principal axes for some choice of $\mathcal{M}_{\text{prior}}$, see [18, section 2.4]. In Section 4.4, we establish some connections between the proposed procedure and some other approaches based on point estimates of the elements of \mathcal{M} . Finally, in Section 4.5, we analyze the complexity of the proposed methods and discuss some implementation issues.

4.1 Some Specific Choices for $\mathcal{M}_{\text{prior}}$

In this section, we advocate that the following choice of prior manifold can be obtained from standard model-order reduction techniques:

$$\mathcal{M}_{\text{prior}} = \bigcap_{j=1}^L \{\mathbf{h} : \text{dist}(\mathbf{h}, V_j) \leq \hat{\epsilon}_j\}, \quad (15)$$

where V_j is some n_j -dimensional subspace, $\hat{\epsilon}_j$ is some positive scalar and L is an integer. In Section 4.3, we will emphasize that such a choice for $\mathcal{M}_{\text{prior}}$ allows for an easy implementation of the sampling of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$.

In order to show that (15) can (for example) be obtained from standard model-order reduction techniques, let us first consider the case where the uncertainty one has on the set of possible solutions of the PPDE (*i.e.*, \mathcal{M}) is due to an imperfect knowledge of the set of feasible parameters Θ (The general case will be discussed at the end of this section). Although Θ may not be precisely known, an information the practitioner usually has at its disposal is that Θ is contained in some larger set Θ_{relax} , *i.e.*, $\Theta \subseteq \Theta_{\text{relax}}$. This relaxed set defines another manifold

$$\mathcal{M}_{\text{relax}} \triangleq \{\mathbf{h}(\theta) \in \mathcal{H} : \theta \in \Theta_{\text{relax}}\}, \quad (16)$$

which obeys $\mathcal{M} \subseteq \mathcal{M}_{\text{relax}}$ (because $\Theta \subseteq \Theta_{\text{relax}}$). Now, constructing a reduced-order model for $\mathcal{M}_{\text{relax}}$ defined as in (16) via *e.g.*, reduced-basis techniques [5] leads to a sequence of subspaces

$$V_1 \subset \dots \subset V_L, \quad (17)$$

and scalars

$$\hat{\epsilon}_1 \geq \dots \geq \hat{\epsilon}_L, \quad (18)$$

such that^e

$$\mathcal{M}_{\text{relax}} \subseteq \{\mathbf{h} : \text{dist}(\mathbf{h}, V_j) \leq \hat{\epsilon}_j\} \text{ for all } j \leq L. \quad (19)$$

Since $\mathcal{M} \subseteq \mathcal{M}_{\text{relax}}$, one can thus construct a prior of the form (15) satisfying (2), from any combination of the subspaces (17) and scalars (18).

We note that this procedure applies even if $\mathcal{M}_{\text{relax}}$ is not defined via a relaxed set of parameters Θ_{relax} as in (16). In a general setting, the only constraints to satisfy in order to fulfil the condition $\mathcal{M} \subseteq \bigcap_{j=1}^L \{\mathbf{h} : \text{dist}(\mathbf{h}, V_j) \leq \hat{\epsilon}_j\}$ are as follows: *i*) identify a manifold $\mathcal{M}_{\text{relax}}$ such that $\mathcal{M} \subseteq \mathcal{M}_{\text{relax}}$; *ii*) apply a model-order reduction technique on $\mathcal{M}_{\text{relax}}$ which certifies that the sequence of subspaces V_j and scalars $\hat{\epsilon}_j$ obey (19).

4.2 Definition of Suitable Representation Bases

In this section, we introduce some representation bases which will be useful for the characterization of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ in the next section. We let W and V denote respectively m - and n -dimensional subspaces of \mathcal{H} . In the next section, W and V will play the role of the observation subspace defined in (3) and one of the ‘‘prior’’ subspaces V_j appearing in the definition of $\mathcal{M}_{\text{prior}}$ in (15).

Let $\mathbf{G} \in \mathbb{R}^{m \times n}$ be the matrix representation of the projector P_W from V to W in some (arbitrary) ONBs $\{\mathbf{w}_j\}_{j=1}^m$ and $\{\mathbf{v}_j\}_{j=1}^n$, *i.e.*,

$$g_{ij} \triangleq \langle \mathbf{w}_i, \mathbf{v}_j \rangle.$$

We define new ONBs, $\{\mathbf{w}_j^*\}_{j=1}^m$ for W and $\{\mathbf{v}_j^*\}_{j=1}^n$ for V , as

$$\begin{aligned} \mathbf{w}_j^* &\triangleq \sum_{i=1}^m \mathbf{w}_i x_{ij}, \\ \mathbf{v}_j^* &\triangleq \sum_{i=1}^n \mathbf{v}_i z_{ij}, \end{aligned}$$

^eFor the sake of precision, let us mention that, as far as reduced-basis techniques are concerned, the inclusion in (19) is guaranteed only if the reduced-basis procedures exploit the *whole* set Θ_{relax} in the computation of V_j and $\hat{\epsilon}_j$. In practice, however, these algorithms only use a fine discretization of Θ_{relax} to speed up the computations. In such a case, ensuring that (19) holds requires some additional care. In the sequel, we do not elaborate on this technical issue and assume that V_j and $\hat{\epsilon}_j$ are such that the inclusion (19) holds.

where x_{ij} and z_{ij} are the coefficients of the matrices appearing in the singular value decomposition of $\mathbf{G} = \mathbf{X}\mathbf{S}\mathbf{Z}^T$; $\mathbf{X} \in \mathbb{R}^{m \times m}$ and $\mathbf{Z} \in \mathbb{R}^{n \times n}$ are orthogonal matrices; $\mathbf{S} \in \mathbb{R}^{m \times n}$ is a rectangular, diagonal matrix.

Clearly, from the definition of $\{\mathbf{w}_i^*\}_{i=1}^m$ and $\{\mathbf{v}_i^*\}_{i=1}^n$, we have

$$\langle \mathbf{w}_i^*, \mathbf{v}_j^* \rangle = s_{ij}. \quad (20)$$

In the sequel, we will use the shorthand notation λ_j to refer to the j th diagonal element of \mathbf{S} , that is $\{\lambda_j\}_{j=1}^{\min(m,n)}$ represents the set of singular values of \mathbf{G} .^f The singular values are assumed to be sorted in a decreasing order of magnitude, *i.e.*,

$$1 \geq \lambda_1 \geq \dots \geq \lambda_{\min(m,n)} \geq 0.$$

The first inequality follows from the fact that \mathbf{G} is the matrix representation of a projection operator: we thus necessarily have that $\lambda_j \leq 1$. Moreover, if $\lambda_j = 1$, we must have $\mathbf{w}_j^* = \mathbf{v}_j^*$.

We define the following short-hand notations that will be useful in the rest of the paper:

$$\begin{aligned} p &\triangleq \text{card}(\{j : \lambda_j = 1\}), \\ q &\triangleq \text{card}(\{j : \lambda_j > 0\}). \end{aligned}$$

From an operational point of view, p represents the number of dimensions of V which are included in W , that is $p = \dim(W \cap V)$. Moreover, $n - q$ corresponds to the number of dimensions of V which are orthogonal to W , that is $n - q = \dim(W^\perp \cap V)$. In a nutshell, q thus represents the number of measurements (out of m) providing information about the position of the points in V .

Finally let us mention that W^\perp can be decomposed as the following direct sum of two orthogonal subspaces (see Appendix A.1):

$$W^\perp = P_{W^\perp}(V) \oplus W^\perp \cap V^\perp. \quad (21)$$

Moreover an ONB for $P_{W^\perp}(V)$ can be expressed in terms of the elements of $\{\mathbf{w}_j^*\}_{j=1}^m$ and $\{\mathbf{v}_j^*\}_{j=1}^n$. More specifically, letting

$$\phi_j \triangleq (1 - \lambda_j^2)^{-\frac{1}{2}} (\mathbf{v}_j^* - \lambda_j \mathbf{w}_j^*),$$

we have that $\{\phi_j\}_{j=p+1}^q \cup \{\mathbf{v}_j^*\}_{j=q+1}^n$ forms an ONB of $P_{W^\perp}(V)$, see Appendix A.1.

^fIn order to provide some geometrical interpretation to the reader, let us mention that the singular value λ_j can also be understood as the cosine of the ‘‘angle’’ between the singular vectors \mathbf{w}_j^* and \mathbf{v}_j^* .

Algorithm 2 Random Sampling of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ for $L = 1$

inputs: $\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m$, $\hat{\epsilon}$, V

init: Compute $\{\mathbf{v}_j^*\}_{j=1}^n$, $\{\mathbf{w}_j^*\}_{j=1}^m$, $\{\lambda_j\}_{j=1}^{\min(m,n)}$ as in Section 4.2

Evaluate $\langle \mathbf{w}_j^*, \mathbf{h} \rangle = \sum_{i=1}^m \langle \mathbf{w}_i, \mathbf{h} \rangle x_{ij}$ for $j = 1, \dots, m$

while $i < N_{\text{samples}}$ **do**

1) Draw π according to some distribution supported on $[0, 1]$

2) Draw γ uniformly on $\left[0, \hat{\epsilon}^2 - \sum_{j=p+1}^m \langle \mathbf{w}_j^*, \mathbf{h} \rangle^2\right]$

3) Draw $\{b_j\}_{p+1}^q$ uniformly on the $(q-p)$ -dimensional unit ball;
scale the result so that $\sum_{j=p+1}^q b_j^2 = \gamma\pi$

4) Draw $\{d_j\}_{q+1}^n$ uniformly on \mathbb{R}^{n-q}

5) Draw \mathbf{z} uniformly on $V^\perp \cap W^\perp$; scale the result so that $\|\mathbf{z}\|^2 = \gamma(1 - \pi^2)$

6) Set $\mathbf{h}_i = \mathbf{h}_{\text{center}} - \sum_{j=p+1}^q b_j \lambda_j^{-1} \phi_j + \sum_{j=q+1}^n d_j \mathbf{v}_j^* + \mathbf{z}$
 $i = i + 1$

end while

output: $\{\mathbf{h}_i \in \mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}\}_{i=1}^{N_{\text{samples}}}$

4.3 Sampling Strategies

We now expose our strategies to sample $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ with $\mathcal{M}_{\text{prior}}$ defined as in (15). First note that each set $\{\mathbf{h} : \text{dist}(\mathbf{h}, V_j) \leq \hat{\epsilon}_j\}$ can be seen as a degenerate ellipsoid (the degenerate directions are defined by the subspace V_j); the definition in (15) thus corresponds to the intersection of L degenerate ellipsoids. In what follows, we will see that this particular geometrical structure allows for an easy sampling of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$.

We first consider (15) in the case where $L = 1$, that is $\mathcal{M}_{\text{prior}}$ is defined as

$$\mathcal{M}_{\text{prior}} = \{\mathbf{h} : \text{dist}(\mathbf{h}, V) \leq \hat{\epsilon}\}, \quad (22)$$

for some n -dimensional subspace V and scalar $\hat{\epsilon} > 0$. In this setup, $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ takes the form of a high-dimensional ellipsoid with orthogonal principal axes. More specifically, Binev *et al.* [18, Section 2.4] (see also Appendix A.2) showed that $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ can be characterized as follows:

$$\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}} = \mathbf{h}_{\text{center}} \oplus \mathcal{E}_{\mathbf{h}}, \quad (23)$$

where

$$\mathbf{h}_{\text{center}} = \sum_{j=1}^q \langle \mathbf{w}_j^*, \mathbf{h} \rangle \lambda_j^{-1} \mathbf{v}_j^* + \sum_{j=q+1}^m \langle \mathbf{w}_j^*, \mathbf{h} \rangle \mathbf{w}_j^*, \quad (24)$$

and $\mathcal{E}_{\mathbf{h}}$ is an ellipsoid defined as

$$\mathcal{E}_{\mathbf{h}} = \left\{ \begin{array}{l} \mathbf{h}' = -\sum_{j=p+1}^q b_j \lambda_j^{-1} \phi_j + \sum_{j=q+1}^n d_j \mathbf{v}_j^* + \mathbf{z} \\ \text{with } \left\{ \begin{array}{l} \mathbf{z} \in V^\perp \cap W^\perp \\ \sum_{j=p+1}^q b_j^2 + \|\mathbf{z}\|^2 \leq \hat{\epsilon}^2 - \sum_{j=q+1}^m \langle \mathbf{w}_j^*, \mathbf{h} \rangle^2 \end{array} \right. \end{array} \right\}. \quad (25)$$

In (24)-(25), we have used the definitions of \mathbf{v}_j^* , \mathbf{w}_j^* and ϕ_j introduced in Section 4.2 with the following conventions: V is the subspace characterizing $\mathcal{M}_{\text{prior}}$ in (22), W is the observation subspace introduced in (3).

From (23)-(25) it is clear that $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ is an ellipsoid centered in $\mathbf{h}_{\text{center}}$. The set $\mathcal{E}_{\mathbf{h}}$ characterizes the deviation of the elements of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ from its center. From our final remark in Section 4.2 (namely, W^\perp can be decomposed as in (21) and $\{\phi_j\}_{j=p+1}^q \cup \{\mathbf{v}_j^*\}_{j=q+1}^n$ forms an ONB of $P_{W^\perp}(V)$), we can deduce that: *i)* $\mathcal{E}_{\mathbf{h}} \subseteq W^\perp$; *ii)* the elements $\{\phi_j\}_{j=p+1}^q \cup \{\mathbf{v}_j^*\}_{j=q+1}^n$ correspond to principal axes of the ellipsoid. Hence, it is clear that the maximum deviation of the ellipsoid $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ from its center $\mathbf{h}_{\text{center}}$ occurs in the direction ϕ_q (resp. $\{\mathbf{v}_j^*\}_{j=q+1}^n$) and is equal to $\lambda_q^{-1}(\hat{\epsilon}^2 - \sum_{j=q+1}^m \langle \mathbf{w}_j^*, \mathbf{h} \rangle^2)$ (resp. $+\infty$) when $q = n$ (resp. $q < n$).

Exploiting (23)-(25), one can randomly sample $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ by using the procedure described in Algorithm 2. The actual distribution according to which the samples are drawn depends on the choice of the distribution on π in the first step of the procedure. For example, in the finite dimensional setting, if

$$\pi = \frac{\sum_{j=1}^{q-p} \xi_j^2}{\sum_{j=1}^{q-p+\dim(W^\perp \cap V^\perp)} \xi_j^2}, \quad (26)$$

where the ξ_j 's are independent realizations of a zero-mean Gaussian distribution, the procedure described in Algorithm 2 draws snapshots from a uniform distribution supported on $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$. Other choices for the distribution on π enable to put more emphasis on some parts of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$, *i.e.*, to draw samples in some regions of the ellipsoid with higher probability (we discuss

Algorithm 3 Random Sampling of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ for $L > 1$

inputs: $\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m$, $\{\hat{\epsilon}_j\}_{j=1}^L$, $\{V_j\}_{j=1}^L$, $j^* \in \{1, \dots, L\}$
while $i < N_{\text{samples}}$ **do**

1) Draw \mathbf{h}_i uniformly at random in the ellipsoid

$$\{\mathbf{h}' : \text{dist}(\mathbf{h}', V_{j^*}) \leq \hat{\epsilon}_{j^*}\} \cap H_{\mathbf{h}},$$

by using the procedure described in the inner loop of Algorithm 2.

2) Acceptance-Rejection:

if $\|P_{V_j^\perp}(\mathbf{h}_i)\| \leq \hat{\epsilon}_j$ for all j **then**

 Include \mathbf{h}_i to the set of snapshots

else

 Reject \mathbf{h}_i

end if

end while

output: $\{\mathbf{h}_i \in \mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}\}_{i=1}^{N_{\text{samples}}}$

this option in Section 4.5). We note however that, as long as the support of the distribution on π is $[0, 1]$, the samples are drawn from a distribution whose support is equal to $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$.

We now focus on the general case where $\mathcal{M}_{\text{prior}}$ is defined by (15) with $L > 1$. In this case $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ does not usually have any “desirable” structure. We note however that the intersection between $H_{\mathbf{h}}$ and each of the L degenerate ellipsoids defining $\mathcal{M}_{\text{prior}}$ (taken separately) forms an ellipsoid. We thus propose the “acceptance-rejection” strategy described in Algorithm 3. We consider as a reference ellipsoid, the ellipsoid defined by the intersection of $H_{\mathbf{h}}$ and the j^* -th degenerate ellipsoid defining $\mathcal{M}_{\text{prior}}$, $j^* \in \{1, \dots, L\}$. We then draw randomly an element from this ellipsoid by using the procedure described in Algorithm 2. This element, say \mathbf{h}_i , is added to the set of snapshots if it verifies all the constraints defining $\mathcal{M}_{\text{prior}}$, *i.e.*,

$$\left\| P_{V_j^\perp}(\mathbf{h}_i) \right\| \leq \hat{\epsilon}_j \quad \text{for all } j \in \{1, \dots, L\},$$

and rejected otherwise.

4.4 Reduction Based on Point Estimates

An alternative approach to build a reduced-order model from partial observations may rely on point estimates of the elements of \mathcal{M} . More specifically, for

Algorithm 4 Approximation Subspace from Point Estimates

inputs: $\left\{ \left\{ \langle \mathbf{w}_j, \mathbf{h} \rangle \right\}_{j=1}^m \right\}_{\mathbf{h} \in \mathcal{M}}, \mathcal{M}_{\text{prior}}$
for $\mathbf{h} \in \mathcal{M}$ **do**
 Evaluate $\hat{\mathbf{h}}$ from $\left\{ \langle \mathbf{w}_j, \mathbf{h} \rangle \right\}_{j=1}^m$ and $\mathcal{M}_{\text{prior}}$
end for
 Apply Algorithm 1 with $\{\hat{\mathbf{h}}\}_{\mathbf{h} \in \mathcal{M}}$ as input to derive \hat{S}^{point}
output: \hat{S}^{point}

each $\mathbf{h} \in \mathcal{M}$, one can compute a point estimate $\hat{\mathbf{h}}$, obeying some optimality criterion, by combining the partial observations $\{\langle \mathbf{w}_i, \mathbf{h} \rangle\}_{i=1}^m$ and some prior information. Then, an approximation subspace for \mathcal{M} can be constructed by considering the manifold of all point estimates, *i.e.*, $\hat{\mathcal{M}} \triangleq \{\hat{\mathbf{h}}\}_{\mathbf{h} \in \mathcal{M}}$, as a good surrogate for \mathcal{M} . This procedure is summarized in Algorithm 4.

We note that the procedure described in Algorithm 4 shares strong connections with the methodology exposed in [11]. In the latter paper, the authors assume that some samples of the solution manifold \mathcal{M} are available but some others are only partially observed. They thus propose to complement the set of known samples with point estimates of those only partially observed to compute an approximation subspace for \mathcal{M} . We see that this setup boils down to the one considered here when *all* the elements of \mathcal{M} are partially observed.

We now make a connection between the procedure described in Algorithm 4 and the material presented in the previous sections. In particular, we emphasize why such point-estimate procedures may fail in providing reliable results in some situations.

There are many choices to compute an estimate $\hat{\mathbf{h}}$ from partial measurements $\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m$ and prior information $\mathcal{M}_{\text{prior}}$. Let us consider the following particular option:

$$\hat{\mathbf{h}} \in \arg \min_{\mathbf{h}'} \left(\sup_{\mathbf{h}'' \in \mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}} \|\mathbf{h}'' - \mathbf{h}'\| \right). \quad (27)$$

From the worst-case perspective pursued in this paper, (27) seems indeed to be a sensible choice since $\hat{\mathbf{h}}$ minimizes the worst-case error over all the elements of \mathcal{H} compatible with the received observations and the prior constraints (namely $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$).[§]

The solution of (27) has recently received some attention for some specific

[§]Note that no particular structure for $\mathcal{M}_{\text{prior}}$ is assumed in (27).

choices of $\mathcal{M}_{\text{prior}}$ [18]. When $\mathcal{M}_{\text{prior}}$ is defined as the intersection of degenerate ellipsoids as in (15), the authors of [18] showed that $\hat{\mathbf{h}}$ corresponds to some specific point of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ (namely the center of the Chebyshev ball of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$).^h In particular, when $L = 1$, $\hat{\mathbf{h}}$ takes the following simple form

$$\hat{\mathbf{h}} = \sum_{j=1}^q \langle \mathbf{w}_j^*, \mathbf{h} \rangle \lambda_j^{-1} \mathbf{v}_j^* + \sum_{j=q+1}^m \langle \mathbf{w}_j^*, \mathbf{h} \rangle \mathbf{w}_j^*, \quad (28)$$

i.e., exactly corresponds to the center of the ellipsoid $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ defined in (23)-(25).ⁱ

From the perspective of our sampling strategies described in Section 4.3, building an approximation subspace from the point estimates $\{\hat{\mathbf{h}}\}_{\mathbf{h} \in \mathcal{M}}$ is then tantamount to sampling *one* (specific) point of each ellipsoid $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$. This is in contrast with the sampling strategies described in Algorithms 2 and 3 where N_{samples} points of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ are drawn at random. We may thus expect the point-estimate procedure to lead to performance close to the optimal worst-case solution (13) when all the points of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ are concentrated around $\hat{\mathbf{h}}$. When $L = 1$, this will for example be the case when $\hat{\epsilon}$ is small and $\lambda_i \simeq 1$ for all $i = 1, \dots, n \leq m$. On the other hand, when $\lambda_i \simeq 0$ for some i , it is easy to see from (25) that the ellipsoid $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ will be very elongated along some directions. In such a case, the center of the ellipsoid $\hat{\mathbf{h}}$ may be a poor representative of the elements of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ and, as a consequence, the approximation subspace computed from $\{\hat{\mathbf{h}}\}_{\mathbf{h} \in \mathcal{M}}$ may significantly differ from the optimal solution (13). We will illustrate this behavior in our numerical simulations in the next section. We provide below an illustrative toy example in which the proposed methodology may succeed in finding a good approximation subspace whereas the methodology based on point estimate (28) is doomed to fail.

Example 1. *We consider the simple case where \mathcal{M} is a k -dimensional linear subspace of \mathcal{H} and the prior manifold $\mathcal{M}_{\text{prior}}$ is defined as in (22) for some n -dimensional subspace V and width $\hat{\epsilon}$. We assume that by some misfortune the observation subspace W is orthogonal to \mathcal{M} . In this case, we have $\forall \mathbf{h} \in \mathcal{M}$,*

$$\langle \mathbf{w}_j^*, \mathbf{h} \rangle = 0 \quad \forall j = 1, \dots, m, \quad (29)$$

^hIn fact, as emphasized in [18, Remark 2.4], $\hat{\mathbf{h}} \in \mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ as soon as $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ is a bounded, closed, convex set.

ⁱWhen $q < n$, the ellipsoid $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ is degenerate along the directions $\{\mathbf{v}_j\}_{j=q+1}^n$. In such a case, there is some ambiguity in the definition of the center of the ellipsoid along these directions. The expression given in (28) then corresponds to the “center” with the minimum norm.

and the “worst-case optimal” point estimate defined in (28) is equal to zero. The manifold $\{\hat{\mathbf{h}}\}_{\mathbf{h} \in \mathcal{M}}$ then reduces to the singleton $\{\mathbf{0}\}$ and a good approximation subspace for \mathcal{M} may obviously not be inferred from the latter.

On the other hand, the posterior manifold $\mathcal{M}_{\text{post}}$ also takes a simple form when W is orthogonal to \mathcal{M} . More specifically, we have^j

$$\mathcal{M}_{\text{post}} = \mathcal{E}_0, \quad (30)$$

where \mathcal{E}_0 corresponds to the ellipsoid defined in (25) with $\mathbf{h} = \mathbf{0}$. By definition, S^{post} will thus be a good approximation subspace for \mathcal{M} as soon as \mathcal{M} is included in the span of the directions corresponding to the largest variations of \mathcal{E}_0 (i.e., $\{\mathbf{v}_j\}_{j=q+1}^n$ and the ϕ_j ’s associated to the smallest singular values λ_j). We give an illustration of such a scenario (in a slightly more complex setup) in Section 5. \square

4.5 Complexity and Implementation Issues

In this section, we discuss the computational complexity of Algorithms 2 and 3, and elaborate on some related issues.

Let us first note that the overall complexity of Algorithms 2 and 3 is intimately related to the costs of evaluating the sum and the inner product between two elements of the Hilbert space \mathcal{H} . In this section, we will assume that these operations can be carried out with a complexity scaling as $\mathcal{O}(\dim(\mathcal{H}))$. This is for example the case in the finite dimensional setting, which is the one of most interest from an operational point of view. With this convention, we have that the complexity of Algorithms 2 and 3 scales as $\mathcal{O}(mn \dim(\mathcal{H}) + N_{\text{samples}}(m+n) \dim(\mathcal{H}))$ and thus evolves favorably with the problem’s dimensions.

Regarding Algorithm 2, this order of magnitude can be obtained by dividing the analysis of the complexity into the costs of the initial step and the main loop. The most demanding operation in the initial step is the computation of $\{\mathbf{v}_j^*\}_{j=1}^n$, $\{\mathbf{w}_j^*\}_{j=1}^m$ and $\{\lambda_j\}_{j=1}^{\min(m,n)}$. This task requires to build the Gram matrix associated to the projection operator between the subspaces V and W and to evaluate its singular value decomposition (see Section 4.2). The first operation involves the computation of mn inner products in \mathcal{H} and has therefore a complexity scaling as $\mathcal{O}(mn \dim(\mathcal{H}))$. Moreover, the evaluation of the singular value decomposition of an $m \times n$ matrix requires at most $\mathcal{O}(\min(m^2n, mn^2))$ operations, see e.g., [22, Lecture

^jThis can easily be seen by setting $\mathbf{h}_{\text{center}} = \mathbf{0}$ in (23)-(25) and noticing that $\mathcal{E}_{\mathbf{h}}$ is equal to $\mathcal{E}_0 \forall \mathbf{h} \in \mathcal{M}$.

31]. Since $\min(m^2n, mn^2) \leq mn \dim(\mathcal{H})$, the complexity of the initial step is thus of order $\mathcal{O}(mn \dim(\mathcal{H}))$.

The most demanding operations in the main loop of Algorithm 2 are the evaluations of steps 5 and 6. These tasks can be performed with a complexity scaling at most as $\mathcal{O}((m+n) \dim(\mathcal{H}))$. Indeed, on the one hand, the uniform sampling of \mathbf{z} in $V^\perp \cap W^\perp$ can be done efficiently as follows: *i)* sample uniformly an element of \mathcal{H} (complexity $\mathcal{O}(\dim(\mathcal{H}))$); *ii)* set $\mathbf{z} = \mathbf{h} - P_{(V^\perp \cap W^\perp)^\perp}(\mathbf{h})$ and evaluate $P_{(V^\perp \cap W^\perp)^\perp}(\mathbf{h})$ with at most $\mathcal{O}((m+n) \dim(\mathcal{H}))$ operations by noticing that $\{\mathbf{v}_j^*\}_{j=1}^n \cup \{\phi_j\}_{j=p+1}^q \cup \{\mathbf{w}_j^*\}_{j=q+1}^m$ is an ONB of $(V^\perp \cap W^\perp)^\perp$, see Appendix A.1. On the other hand, the construction of the snapshots \mathbf{h}_i in step 6 requires the summation of at most $m+n+1$ elements of \mathcal{H} . Gathering these different elements together, we finally obtain that sampling N_{samples} elements of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ with Algorithm 2 requires a complexity of order $\mathcal{O}(mn \dim(\mathcal{H}) + N_{\text{samples}}(m+n) \dim(\mathcal{H}))$ as stated above.

The complexity of Algorithm 3 is essentially of the same order as the one of Algorithm 2 since the latter constitutes the main building block of the former. Nevertheless, we note that the running time of Algorithm 3 can be significantly larger than the one of Algorithm 2 if the rejection ratio (step 2 in the main loop of Algorithm 3) is important. The choice of the reference ellipsoid j^* appearing in Algorithm 3 should therefore be made with care in order to decrease as much as possible the rejection ratio.

Finally, let us note that our study of the complexity is based on the sampling of N_{samples} elements of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ for *one* $\mathbf{h} \in \mathcal{M}$. In practice, the operations stated in Algorithms 2 and 3 must be repeated for all $\mathbf{h} \in \mathcal{M}$. As a matter of fact, if $\mathbf{h} \in \mathcal{M}$ contains an infinite number of elements, the framework exposed previously cannot, strictly speaking, be applied (it would in particular require to collect a set of observations $\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m$ for each element in \mathcal{M}). In such a case, one can nevertheless apply the proposed procedure on a finely-sampled version of the target manifold \mathcal{M} . All the results discussed previously then carry over by considering the sampled version of \mathcal{M} as the new target manifold. One may expect the approximation subspace computed from this sampled manifold to lead to a good approximation subspace for the true manifold \mathcal{M} as long as the latter is sampled “finely enough”. The question of the proper sampling of \mathcal{M} is however out of the scope of the present paper and is not further discussed here. We nevertheless mention that, for a given precision, the required number of samples is expected to (typically) scale exponentially with the dimension of

the parameter space Θ .

To conclude this section, we discuss the choice of the distribution of π appearing in step 1 of Algorithm 2. Since any distribution on π with support $[0, 1]$ defines a distribution on the snapshots supported on $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$, we see that any choice of the distribution on π should asymptotically (in the number of samples) leads to the same performance. A proper choice of this distribution may however have a significant impact on the achievable performance when the number of samples is finite. We advocate below that sampling $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ with higher probability in the directions of greatest uncertainty may be a good rule of thumbs.

In order to provide some elements supporting this fact, let us consider the case where $\mathcal{M}_{\text{prior}}$ is defined as in (22) and $n = m$. In such a case, the posterior manifold $\mathcal{M}_{\text{post}}$ can be simply written as

$$\mathcal{M}_{\text{post}} = \hat{\mathcal{M}} \oplus \mathcal{E}_0, \quad (31)$$

where $\hat{\mathcal{M}} = \{\mathbf{h}_{\text{center}}\}_{\mathbf{h} \in \mathcal{M}}$ with $\mathbf{h}_{\text{center}}$ defined in (24) and \mathcal{E}_0 is the ellipsoid defined in (25) with $\mathbf{h} = \mathbf{0}$.^k

Considering this simple expression and assuming that some directions of large variations of the target manifold \mathcal{M} are not captured by the point-estimate manifold $\hat{\mathcal{M}}$, we see that there is still a hope to identify the latter if they correspond to large variations in \mathcal{E}_0 . In such a case, one may expect the proposed worst-case optimal subspace S^{post} to be a better approximation of the target manifold \mathcal{M} than \hat{S}^{point} , the subspace evaluated from the sole point-estimate manifold $\hat{\mathcal{M}}$. According to this intuition, when only a limited number of snapshots can be drawn from $\mathcal{M}_{\text{prior}} \cap H$ for each $\mathbf{h} \in \mathcal{M}$, sampling $\mathcal{M}_{\text{prior}} \cap H$ with higher probability in the directions of large uncertainty seems to be a sensible choice.

5 Simulation Results

In this section, we illustrate the performance of the proposed reduction procedures on the standard ‘‘thermal-block’’ problem [23]: the goal is to evaluate the distribution of the temperature on a plate subject to some boundary conditions, for some specific configurations of the plate’s heat conductivity and some external heating source. More specifically, the problem is defined by the following set of differential/boundary equations ($\mathbf{x} \in \mathbb{R}^2$ plays the role

^k $\hat{\mathcal{M}}$ is in fact the manifold of the point estimates computed from (27) with prior (22) as discussed in Section 4.4.

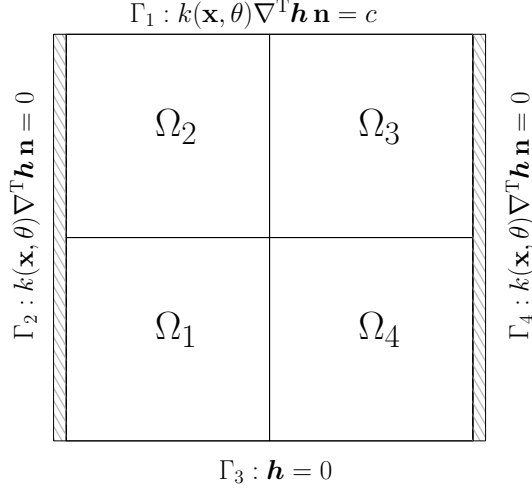


Figure 1: Schematic representation of the physical system.

of a “spatial” variable, \mathbf{n} is a unitary vector normal to the boundary and ∇ is the gradient operator):

$$\text{PDE}(\mathbf{h}, \theta) = \begin{cases} \nabla^T(k(\mathbf{x}, \theta)\nabla\mathbf{h}) = s(\mathbf{x}), & \mathbf{x} \in \Omega, \\ k(\mathbf{x}, \theta)\nabla^T\mathbf{h}\mathbf{n} = c, & \mathbf{x} \in \Gamma_1, \\ k(\mathbf{x}, \theta)\nabla^T\mathbf{h}\mathbf{n} = 0, & \mathbf{x} \in \Gamma_2 \cup \Gamma_4, \\ \mathbf{h} = 0, & \mathbf{x} \in \Gamma_3, \end{cases} \quad (32)$$

where $\Omega = [0, 1] \times [0, 1]$ and the boundaries Γ_i are defined in Fig. 1. We assume that the heat conductivity coefficient $k(\mathbf{x}, \theta)$ depends on the parameter $\theta = [\theta_1 \theta_2 \theta_3 \theta_4] \in \Theta$ as follows:

$$k(\mathbf{x}, \theta) = \sum_{i=1}^4 \theta_i \mathbb{I}_{\Omega_i}(\mathbf{x}), \quad (33)$$

where $\mathbb{I}_{\Omega_i}(\mathbf{x})$ is the indicator function of Ω_i and the subdomains $\Omega_i \subset \Omega$ are defined in Fig. 1. The definitions of the external heating source $s(\mathbf{x})$ and the boundary parameter c depend on the experiment and are specified below.

We consider the weak formulation of (32) and approximate its solution via a finite-element method [24] (see Appendix B). The resolution mesh is chosen fine enough so that the error between the solution of the weak formulation of (32) and the solution of the finite-element method can be neglected. The discretized system has a dimension equal to $\dim(\mathcal{H}) = 2113$.

The solution of the discretized system is computed via the Matlab[®] toolbox “redbKIT” available at <http://redbkit.github.io/redbKIT>.

In our simulations, we apply the subspace identification procedures described in Algorithms 2, 3 and 4 to the following two setups:

Setup 1: We first assume that the target manifold \mathcal{M} is defined by the solutions of (the discretized weak formulation of) (32) for

$$\Theta = \left\{ \theta \in \mathbb{R}^4 : \begin{array}{l} \theta_i = \theta_{\min} + \theta_{\text{step}}t, t \in \{0, \dots, T\} \\ \theta_1 = \theta_2, \theta_3 = \theta_4 \end{array} \right\},$$

where $\theta_{\min} = 0.1$, $\theta_{\text{step}} = 0.1$, $T = 20$. In this setup, we assume that $c = 1$ and $s(\mathbf{x}) = 0 \forall \mathbf{x}$. Moreover, we suppose that, for some reasons, the practitioner does not have a perfect knowledge of Θ but only knows that $\Theta \subseteq \Theta_{\text{relax}}$ where

$$\Theta_{\text{relax}} = \{\theta \in \mathbb{R}^4 : \theta_i = \theta_{\min} + \theta_{\text{step}}t, t \in \{0, \dots, T\}, \forall i \in \{1, \dots, 4\}\}.$$

We denote by $\mathcal{M}_{\text{relax}}$ the set of solutions of the discretized weak formulation of (32) for $\theta \in \Theta_{\text{relax}}$. Clearly, we have $\mathcal{M} \subseteq \mathcal{M}_{\text{relax}}$ since $\Theta \subseteq \Theta_{\text{relax}}$. Based on this knowledge we derive a prior manifold $\mathcal{M}_{\text{prior}}$ of the form (15) by following the procedure described in Section 4.1. More specifically, we apply Algorithm 1 on the elements of $\mathcal{M}_{\text{relax}}$. Letting

$$\hat{S}_1 \subset \dots \subset \hat{S}_n$$

be the approximation subspaces produced during the first $n \geq L$ iterations of Algorithm 1, the subspaces $\{V_i\}_{i=1}^L$ and scalars $\{\hat{\epsilon}_i\}_{i=1}^L$ appearing in (15) are then specified as follows:¹

$$\begin{aligned} V_1 &= \hat{S}_1, \\ &\vdots \\ V_{L-1} &= \hat{S}_{L-1}, \\ V_L &= \hat{S}_n, \end{aligned}$$

and, $\forall j \in \{1, \dots, L\}$,

$$\hat{\epsilon}_j = \sup_{\mathbf{h} \in \mathcal{M}_{\text{relax}}} \text{dist}(\mathbf{h}, V_j).$$

We note that (19) is verified by definition of $\hat{\epsilon}_j$, so that our working hypothesis (2) is satisfied. In our simulation, we consider the case where $L = 1$ and $L = 21$. The ONB defining the observation subspace W is chosen uniformly at random.

¹If $L = 1$ as in (22), we simply set $V = \hat{S}_n$.

Setup 2: In this setup, we consider a scenario where the main directions of \mathcal{M} are poorly aligned with the observation subspace W , resulting in (potentially) adverse operating conditions for the point-estimate procedure described in Section 4.4.

In order to precisely describe the experimental setup considered here, we introduce two ONBs, $\{\tilde{\mathbf{v}}_j\}_{j=1}^{N_{\max}}$ and $\{\tilde{\mathbf{w}}_j\}_{j=1}^{N_{\max}}$, such that

$$\langle \tilde{\mathbf{w}}_i, \tilde{\mathbf{v}}_j \rangle = 0 \quad \text{for } i \neq j, \quad (34)$$

and

$$\langle \tilde{\mathbf{w}}_j, \tilde{\mathbf{v}}_j \rangle = \begin{cases} \delta & j = 1 \dots k, \\ 1 & \text{otherwise,} \end{cases} \quad (35)$$

for some scalar $\delta \in (0, 1)$ and some integers $k \leq N_{\max}$. We also choose these ONBs so that any $\tilde{\mathbf{v}}_j$ and $\tilde{\mathbf{w}}_j$ verify the homogeneous boundary conditions of (32). We then let $\mathbf{u}_j = (1 - \delta^2)^{-\frac{1}{2}}(\tilde{\mathbf{v}}_j - \delta\tilde{\mathbf{w}}_j)$ for $j = 1 \dots k$.

We construct \mathcal{M} as follows: we set $c = 0$, the parameters θ defining the heat conductivity coefficient $k(\mathbf{x}, \theta)$ are fixed to some constant values and we choose the source term $s(\mathbf{x})$ so that:^m

$$\mathcal{M} = \mathcal{E}_{\text{main}} \oplus \mathcal{E}_{\text{perturb}}, \quad (36)$$

where

$$\begin{aligned} \mathcal{E}_{\text{main}} &= \left\{ \mathbf{h} = \sum_{j=1}^k \alpha_j \mathbf{u}_j : \sum_{j=1}^k \alpha_j^2 \leq \epsilon_{\text{main}}^2 \right\}, \\ \mathcal{E}_{\text{perturb}} &= \left\{ \mathbf{h} = \delta \sum_{j=1}^k \beta_j \tilde{\mathbf{w}}_j + \sum_{j=k+1}^{N_{\max}} \beta_j \tilde{\mathbf{w}}_j : \sum_{j=1}^{N_{\max}} \gamma_j^2 \beta_j^2 \leq \epsilon_{\text{perturb}}^2 \right\}, \end{aligned}$$

for some $\epsilon_{\text{main}} \geq 0$, $\epsilon_{\text{perturb}} \geq 0$ and $\gamma_j \geq 1$. The solution manifold \mathcal{M} is thus the direct sum of two ellipsoids, $\mathcal{E}_{\text{main}}$ and $\mathcal{E}_{\text{perturb}}$. We will see below that this particular choice for \mathcal{M} together with some specific choice of the observation and prior subspaces lead to a difficult problem for the point-estimate approach.

In particular, we consider a prior manifold $\mathcal{M}_{\text{prior}}$ of the form (15) by making the following choices:ⁿ

$$\begin{aligned} V_1 &= \text{span}(\{\tilde{\mathbf{v}}_1\}), \\ &\vdots \\ V_{L-1} &= \text{span}\left(\{\tilde{\mathbf{v}}_j\}_{j=1}^{L-1}\right), \\ V_L &= \text{span}\left(\{\tilde{\mathbf{v}}_j\}_{j=1}^n\right), \end{aligned}$$

^mWe refer the reader to Appendix B for more explanations about the reasons why the manifold (36) can be generated by properly defining the source term $s(\mathbf{x})$.

ⁿIf $L = 1$ as in (22), we simply set $V = \text{span}\left(\{\tilde{\mathbf{v}}_j\}_{j=1}^n\right)$.

for some $n \leq N_{\max}$ and, $\forall j \in \{1, \dots, L\}$,

$$\hat{\epsilon}_j = \sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, V_j).$$

The observation subspace is defined as $W = \text{span} \left(\{\tilde{\mathbf{w}}_j\}_{j=1}^m \right)$ for some $m \leq N_{\max}$. We note that with this particular choice for V and W , we have from (34)-(35) that the singular vectors $\{\mathbf{v}_j^*\}_{j=1}^n$, $\{\mathbf{w}_j^*\}_{j=1}^m$ simply correspond to $\{\tilde{\mathbf{v}}_j\}_{j=1}^n$, $\{\tilde{\mathbf{w}}_j^*\}_{j=1}^m$ whereas the associated singular values are defined by (35).

In this particular setup, the main variations of \mathcal{M} occur in the direction of $\text{span} \left(\{\mathbf{u}_j\}_{j=1}^k \right)$ as soon as $\epsilon_{\text{perturb}} \ll \epsilon_{\text{main}}$. Moreover, if we choose $\delta \simeq 0$, we have from (34)-(35) that $\langle \mathbf{u}_i, \tilde{\mathbf{w}}_j \rangle \simeq 0$, so that only a small portion of the variations of \mathcal{M} in the directions of $\text{span} \left(\{\mathbf{u}_j\}_{j=1}^k \right)$ can be captured by the observation operator. In particular, if we consider the case where $\delta = 0$ and $\epsilon_{\text{perturb}} = 0$, we obtain the same setup as in Example 1: W is orthogonal to $\mathcal{M} = \mathcal{E}_{\text{main}} \subseteq \text{span} \left(\{\mathbf{u}_j\}_{j=1}^k \right)$ and the point-estimate manifold $\hat{\mathcal{M}}$ reduces to the singleton $\{\mathbf{0}\}$; on the other hand, we have $\mathbf{u}_j = \mathbf{v}_j^*$ and $\langle \mathbf{w}_j^*, \mathbf{v}_j^* \rangle = 0$ so that the directions $\{\mathbf{u}_j\}_{j=1}^k$ correspond to directions of large uncertainties in $\mathcal{M}_{\text{post}}$.^o

Hereafter, we consider the following scenario: we set $k = 5$, $N_{\max} = 50$, $\delta = 10^{-4}$, $\epsilon_{\text{main}} = 1$ and $\epsilon_{\text{perturb}} = 10^{-3}$; the weights γ_j in the definition of $\mathcal{E}_{\text{perturb}}$ are defined as follows:

$$\gamma_j = \begin{cases} 0.85^{-N_{\max}} & j = 1 \dots k, \\ 0.85^{-(j-k)} & j > k. \end{cases} \quad (37)$$

As we will see below, this choice of γ_j (together with the definition of W and V) is such that the point-estimate manifold $\hat{\mathcal{M}}$ has small variations in the directions of $\text{span} \left(\{\mathbf{u}_j\}_{j=1}^k \right)$.

In our simulation, we chose the ONB $\{\tilde{\mathbf{v}}_j\}_{j=1}^{N_{\max}}$ arbitrarily and constructed $\{\tilde{\mathbf{w}}_j^*\}_{j=1}^{N_{\max}}$ so that (34)-(35) is satisfied. Regarding the definition of $\mathcal{M}_{\text{prior}}$, we considered both the cases where $L = 1$ and $L = 11$.

We now consider the application of the procedures presented in Sections 4.3 and 4.4 to the two setups described above. For the construction of $\mathcal{M}_{\text{post}}$,

^oIn particular, as long as prior (22) is considered, $\{\mathbf{u}_j\}_{j=1}^k$ corresponds to some degenerate directions of the ellipsoid $\mathcal{E}_{\mathbf{h}}$ in (25).

we draw $N_{\text{samples}} = 5$ elements randomly from $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ for each $\mathbf{h} \in \mathcal{M}$ by using the methodologies described in Algorithms 2 and 3. In step 1 of Algorithm 2, we define the distribution on π as follows:

$$\pi = \begin{cases} \frac{\sum_{j=1}^{q-p} \xi_j^2}{\sum_{j=1}^{q-p+\dim(W^\perp \cap V^\perp)} \xi_j^2} & \text{with probability 0.1,} \\ \frac{10^4 \sum_{j=1}^{q-p} \xi_j^2}{10^4 \sum_{j=1}^{q-p} \xi_j^2 + \sum_{j=q-p+1}^{q-p+\dim(W^\perp \cap V^\perp)} \xi_j^2} & \text{with probability 0.9,} \end{cases} \quad (38)$$

where the ξ_j 's are independent realizations of a zero-mean Gaussian distribution. The first row in (38) leads to a uniform sampling of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ whereas the second one favors the directions of greatest uncertainty. In step 4 of Algorithm 2, we approximate the uniform sampling of $\{d_j\}_{q+1}^n$ over \mathbb{R}^{n-q} by a uniform drawing over $[-10, 10]^{n-q}$.

Fig. 2 and 4 represent the maximum projection error obtained by projecting the element of \mathcal{M} onto different approximation subspaces obtained in Setups 1 and 2, respectively. These figures thus illustrate the actual reduction performance obtained by different approximation methods. On the other hand, Fig. 3 and 5 represent the maximum error obtained by projecting the elements of $\mathcal{M}_{\text{post}}$ onto the same approximation subspaces for Setups 1 and 2, respectively. Since $\mathcal{M}_{\text{post}}$ is the largest manifold compatible with the prior constraint (2) and the received observations (see Lemma 1), the curves in Fig. 3 and 5 thus provide the worst performance attainable over the set of feasible manifolds. The parameters m and n defining the observation operator and the prior manifold used in each simulation are mentioned on the top of each figure.

The approximation subspaces considered in all these figures are obtained by applying Algorithm 1 on:

- the true manifold \mathcal{M} (“perf”).
- the posterior manifold $\mathcal{M}_{\text{post}}$ defined from a prior made up of either a single (22) or an intersection (15) of degenerate ellipsoids (respectively “post single” and “post multi”). These curves thus represent the performance achievable by (approximately) solving the worst-case optimal problem (13). We note that different experiments may lead to different performance since the snapshots are drawn randomly from $\mathcal{M}_{\text{post}}$ (see Algorithms 2 and 3). This variability of the results is reported in the figures: for each simulation point, we ran 20 experiments and averaged the results to obtain the main curve; the shaded area surrounding this curve represents the interval containing the performance obtained for all the 20 experiments.

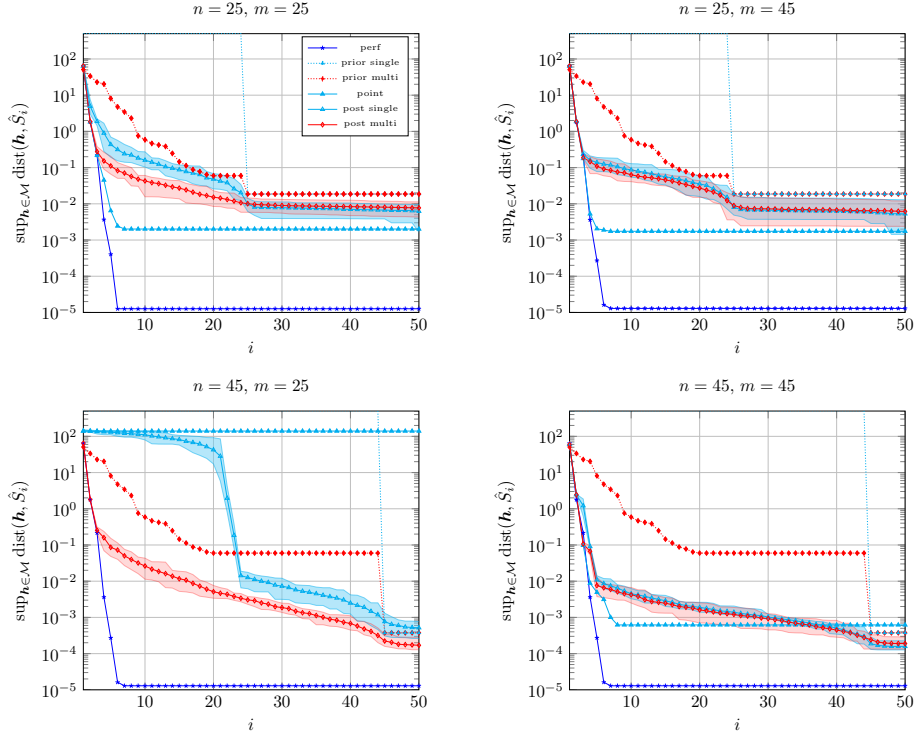


Figure 2: **Setup 1**: Maximum error obtained by projecting the element of \mathcal{M} onto different approximation subspaces. The abscissa represents the dimension i of the approximation subspace. Each figure corresponds to a particular dimension of the prior and observation subspaces, V and W .

- the point-estimate manifold $\hat{\mathcal{M}} \triangleq \{\hat{\mathbf{h}}\}_{\mathbf{h} \in \mathcal{M}}$, where $\hat{\mathbf{h}}$ is the solution of (27) with $\mathcal{M}_{\text{prior}}$ defined as in (22) (“point”). We do not consider the case where $\mathcal{M}_{\text{prior}}$ is defined as in (15) since, in this case, problem (27) does not have any simple analytical solution (the problem is actually NP-hard [18]).

We have also reported the Kolmogorov i -width, $\kappa_i(\mathcal{M}_{\text{prior}})$, of the priors used by the different procedures (“prior single” and “prior multi”) as a point of comparison. Regarding Fig. 3 and 5, we note that the definition of $\mathcal{M}_{\text{post}}$ depends on the choice of $\mathcal{M}_{\text{prior}}$. The curve “post multi” has been computed by using the definition of $\mathcal{M}_{\text{prior}}$ in (15). The other curves (“perf”, “point” and “post single”) have been evaluated by considering the posterior manifold in the case where (22) holds.

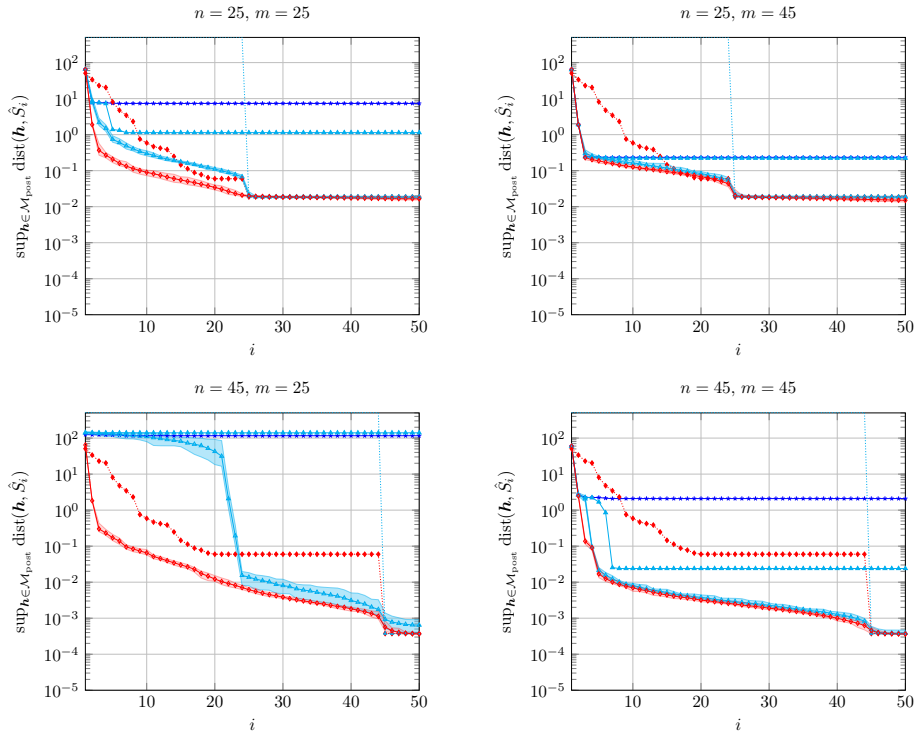


Figure 3: **Setup 1:** Maximum error obtained by projecting the element of $\mathcal{M}_{\text{post}}$ onto different approximation subspaces. The abscissa represents the dimension i of the approximation subspace. Each figure corresponds to a particular dimension of the prior and observation subspaces, V and W .

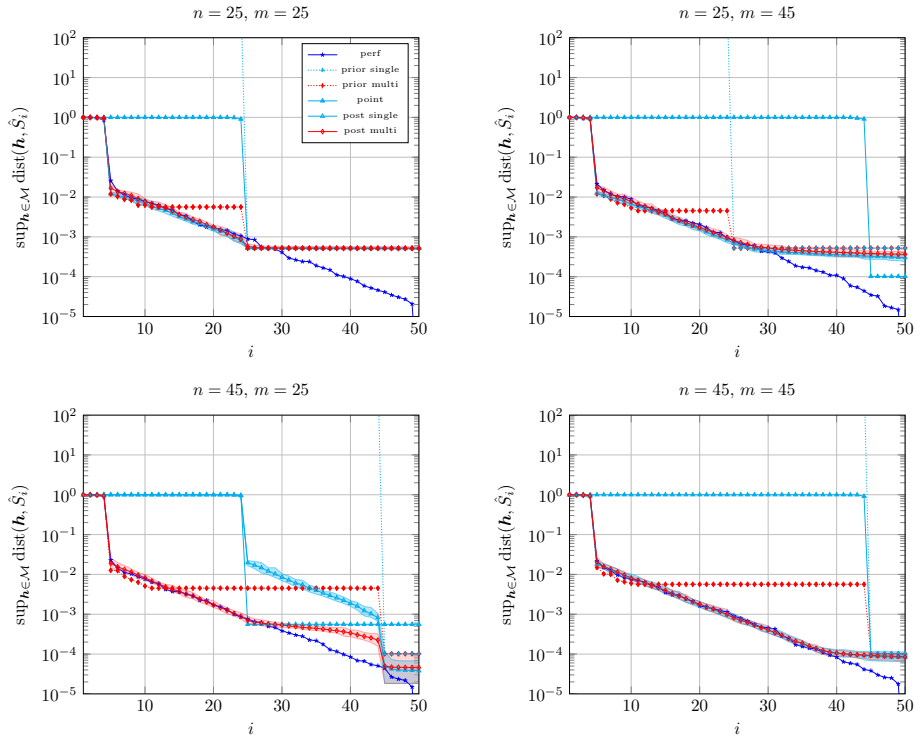


Figure 4: **Setup 2**: Maximum error obtained by projecting the element of \mathcal{M} onto different approximation subspaces. The abscissa represents the dimension i of the approximation subspace. Each figure corresponds to a particular dimension of the prior and observation subspaces, V and W .

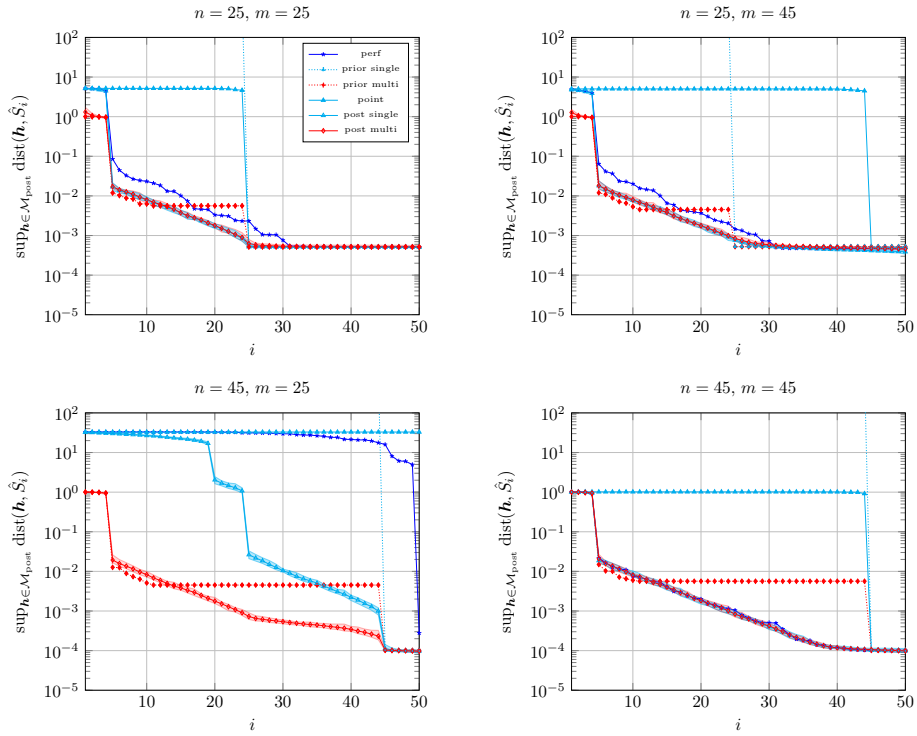


Figure 5: **Setup 2**: Maximum error obtained by projecting the element of $\mathcal{M}_{\text{post}}$ onto different approximation subspaces. The abscissa represents the dimension i of the approximation subspace. Each figure corresponds to a particular dimension of the prior and observation subspaces, V and W .

We now discuss the performance achieved by the different methodologies mentioned in the paper. The dark blue curve (“perf”) in Fig. 2 and 4 corresponds to the performance which can be attained if one has access to the true solution manifold \mathcal{M} . The other solid curves illustrate the performance obtained by exploiting partial observations in the reduction process. As far as the procedures introduced in this paper are concerned (“post single” and “post multi”), we observe that the presence of partial observations (almost) always leads to some improvement as compared to the Kolmogorov i -width of the prior used in the reduction procedure. This behavior could theoretically be expected from (14) since one must have $\sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S_i) \leq \kappa_i(\mathcal{M}_{\text{prior}})$ as long as the approximation subspace S_i is computed as the solution of (13). Because we only consider an approximate implementation of (13) (we draw a finite number of snapshots from $\mathcal{M}_{\text{post}}$ and use a greedy strategy to search for a solution of (13)), we can nevertheless observe some (limited) degradations of the performance with respect to $\kappa_i(\mathcal{M}_{\text{prior}})$ in some regions.

On the other hand, the behavior of the procedure based on point estimates heavily depends on the considered scenario. We see in Fig. 2 that, as long as Setup 1 is concerned, the approximation subspace leads to good reduction performance for all choices of m and n except for $m = 25, n = 45$. The latter case corresponds to the scenario where $\mathcal{M}_{\text{prior}}$ has more degrees of freedom (*i.e.*, $n = 45$) than the number of observations (*i.e.*, $m = 25$); hence the point estimates (28) computed from the received measurements $\{\langle \mathbf{w}, \mathbf{h} \rangle\}$ are poor representatives of the true elements $\mathbf{h} \in \mathcal{M}$. On the other hand, when Setup 2 is considered, we see in Fig. 4 that the point-estimate approach may, in some cases, not bring any improvement over the performance achievable by the prior manifold ($m = 25, n = 25$ and $m = 45, n = 45$) or even degrade the performance ($m = 45, n = 25$).

The performance achieved by the point-estimate procedure in Setup 2 can be understood by noticing that the point-estimate manifold $\hat{\mathcal{M}} = \{\hat{\mathbf{h}}\}_{\mathbf{h} \in \mathcal{M}}$ takes the following form:

$$\hat{\mathcal{M}} = \left\{ \hat{\mathbf{h}} = \sum_{j=1}^{\min(n,m)} \beta_j \tilde{\mathbf{v}}_j + \sum_{j=\min(n,m)+1}^m \beta_j \tilde{\mathbf{w}}_j : \sum_{j=1}^m \gamma_j^2 \beta_j^2 \leq \epsilon_{\text{perturb}}^2 \right\}. \quad (39)$$

We first note that $\hat{\mathcal{M}}$ is contained in a subspace of dimension m and the greedy procedure defined in Algorithm 1 thus necessarily stops after m iterations. Moreover, since $\delta \simeq 0$, the main directions of \mathcal{M} , namely $\{\mathbf{u}_j\}_{j=1}^k$, are such that $\{\mathbf{u}_j\}_{j=1}^k \simeq \{\tilde{\mathbf{v}}_j\}_{j=1}^k$. On the other hand, the subspace $\{\tilde{\mathbf{v}}_j\}_{j=1}^k$ corresponds to the directions of smallest amplitudes in $\hat{\mathcal{M}}$ because of (37).

Hence the greedy procedure applied on $\hat{\mathcal{M}}$ will select the directions $\{\tilde{\mathbf{v}}_j\}_{j=1}^k$ after only $m - k$ iterations, explaining the observed results. In particular, increasing the number of observations may degrade the performance (compare the cases $m = 25, n = 25$ and $m = 45, n = 25$). We also note that, unlike in Setup 1, the point-estimate procedure may decrease the approximation error in the case $m = 25, n = 45$ for $i > 25$. This is due to the fact that, in the particular scenario considered in Setup 2, the $n - m = 20$ unobserved directions of V are (by construction) orthogonal to the main directions $\{\mathbf{u}_j\}_{j=1}^k$ of \mathcal{M} .

Regarding the worst-case performance, we see from Fig. 3 and 5 that, quite logically, the best results are obtained by the subspaces optimizing the worst projection error over the elements of $\mathcal{M}_{\text{post}}$. In particular, although the point-estimate approach may outperform the proposed procedure regarding the approximation of \mathcal{M} , it always leads to inferior performance in the worst-case scenario. Interestingly, we also observe that the subspace computed from the true solution manifold \mathcal{M} may, in some cases, lead to poor performance, showing that the information contained in the true and the posterior manifolds is quite different. We finally note that since $\mathcal{M} \subseteq \mathcal{M}_{\text{post}}$, we have

$$\sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, \hat{S}_i) \leq \sup_{\mathbf{h} \in \mathcal{M}_{\text{post}}} \text{dist}(\mathbf{h}, \hat{S}_i),$$

for any approximation subspace \hat{S}_i . Hence, the curves in Fig. 3 (resp. Fig. 5) constitute (approximate^P) upper bounds on those represented in Fig. 2 (resp. Fig. 4). Since these bounds only depend on $\mathcal{M}_{\text{post}}$, they can always be computed in practice, even if the true solution manifold \mathcal{M} is not available.

Regarding the performance achieved by the subspaces computed from the posterior manifold $\mathcal{M}_{\text{post}}$, the choice of the prior manifold seems to play a crucial role in the minimal projection error which can be attained. In particular, the smaller the width $\hat{\epsilon}_L$, the lower the minimal projection error. The number of observations seems to only have an impact on the minimum subspace dimension required to reach a given approximation error. For example, in Fig. 3 the minimal projection error achievable for $n = 25$ is roughly 10^{-2} irrespective of the number of observations, whereas an error as low as $\sim 10^{-4}$ can be obtained by setting $n = 45$. In the latter case we note, as far as the case $L = 21$ is concerned, that an approximation error of roughly $\sim 10^{-2}$ can be obtained from $i \simeq 10$ when $m = 45$ whereas one needs to increase

^PThese bounds are only approximate since, in practice, we only draw a finite number of snapshots from $\mathcal{M}_{\text{post}}$.

the dimension of the approximation subspace up to $i \simeq 25$ to obtain the same performance when $m = 25$. In the next section, we will provide some theoretical insights into these observations.

6 Theoretical analysis

In this section, we provide a theoretical analysis of the reduction performance achievable within our partially-informed framework. We consider the following simplified scenario:

$$\mathcal{M} \subseteq \Sigma \cap \mathcal{M}_{\text{prior}}, \quad (40)$$

where

$$\begin{aligned} \Sigma &\triangleq \{\mathbf{h} : \text{dist}(\mathbf{h}, U) \leq \epsilon\}, \\ \mathcal{M}_{\text{prior}} &\triangleq \{\mathbf{h} : \text{dist}(\mathbf{h}, V) \leq \hat{\epsilon}\}, \end{aligned} \quad (41)$$

for some subspaces $U \subseteq V$, with $\dim(U) = k$, $\dim(V) = n$ and some scalars $\epsilon, \hat{\epsilon} \geq 0$. In words, we assume that the unknown manifold \mathcal{M} is contained in the intersection of two degenerate ellipsoids but only one of them (namely $\mathcal{M}_{\text{prior}}$) is known a priori. This scenario thus corresponds to the particular case where $\mathcal{M}_{\text{prior}}$ is defined as in (22).

Our goal is to relate the reduction performance obtained by the worst-case optimal approximation subspace $\mathcal{S}_i^{\text{post}}$, defined in (13), to the values of $k, n, \epsilon, \hat{\epsilon}$ and the choices of V, W . We show below that the latter is closely related to the singular values $\{\lambda_j\}_{j=1}^{\min(m,n)}$ of the Gram matrix \mathbf{G} defined in Section 4.2.

In order to state our result, we first need to introduce some notations. First, let us remind the following notations, introduced in Section 4.2:

$$\begin{aligned} p &\triangleq \text{card}(\{j : \lambda_j = 1\}), \\ q &\triangleq \text{card}(\{j : \lambda_j > 0\}). \end{aligned}$$

As mentioned previously, the operational meaning of these variables is as follows: p represents the number of dimensions of V which are included in W , that is $p = \dim(W \cap V)$; $n - q$ corresponds to the number of dimensions of V which are orthogonal to W , that is $n - q = \dim(W^\perp \cap V)$. In a nutshell, q thus represents the number of measurements (out of m) providing information about the position of points in V .

We also introduce the following new variable

$$k^* \triangleq \min(n, k + n - q),$$

and the sequences $\bar{\kappa}_i$ and $\bar{\bar{\kappa}}_i$:^q

$$\bar{\kappa}_i \triangleq \begin{cases} \infty & i = 1, \dots, k^* - 1, \\ (\epsilon + \hat{\epsilon}) \lambda_{q-(i-k^*)}^{-1} & i = k^*, \dots, n - 1, \\ \hat{\epsilon} & i = n, \dots, \dim(\mathcal{H}) - 1, \end{cases} \quad (42)$$

$$\bar{\bar{\kappa}}_i \triangleq \begin{cases} \infty & i = 1, \dots, k + \dim(W^\perp) - 1, \\ \epsilon & i = k + \dim(W^\perp), \dots, \dim(\mathcal{H}) - 1. \end{cases} \quad (43)$$

We are now ready to state the following result, whose proof is postponed to Section 7:

Theorem 1. *If \mathcal{M} and $\mathcal{M}_{\text{prior}}$ verify (40)-(41), then the following inequality holds:*

$$\kappa_i(\mathcal{M}_{\text{post}}) \leq \min(\bar{\kappa}_i, \bar{\bar{\kappa}}_i) \quad \text{for all } i.$$

Interestingly, the upper bounds $\bar{\kappa}_i$ and $\bar{\bar{\kappa}}_i$ appearing in Theorem 1 only depends on a set of simple parameters defining the partially-informed reduction problem, namely k , n , ϵ , $\hat{\epsilon}$ and the singular values of the Gram matrix \mathbf{G} , characterizing the interplay between the prior and observation subspaces, V and W . This few number of parameters enables an easier understanding of the performance achievable by the worst-case reduction methodology presented in Sections 3 and 4. Let us recall that the largest projection error induced by the optimal approximation subspace S_i^{post} is bounded by the Kolmogorov i -width $\kappa_i(\mathcal{M}_{\text{post}})$, see (14). The upper bound in Theorem 1 thus also defines an upper limit on the projection error made by reducing the true, unknown, manifold \mathcal{M} in the worst-case optimal subspace S_i^{post} .

In order to ease the discussion of the result stated in Theorem 1, we will suppose in the rest of this section that^r

$$\mathcal{M} = \Sigma, \quad (44)$$

where Σ is defined in (41). If ϵ is small, (44) is tantamount to assuming that \mathcal{M} has an ‘‘intrinsic dimensionality’’ equal to k , *i.e.*, the elements of \mathcal{M} can vary with no constraint in a k -dimensional subspace U but can only deviate from the latter by a small amount ϵ in all the other directions.

^qWe remind the reader that the singular values $\{\lambda_j\}_{j=1}^{\min(m,n)}$ are assumed to be sorted in a decreasing order of magnitude.

^rWe note that satisfying (40) then requires $\epsilon \leq \hat{\epsilon}$.

The result stated in Theorem 1 can then be discussed in light of the following comments. First, the Kolmogorov i -width of \mathcal{M} takes a very simple form under hypothesis (44), *i.e.*,

$$\kappa_i(\mathcal{M}) = \begin{cases} \infty & \text{if } i < k, \\ \epsilon & \text{otherwise.} \end{cases} \quad (45)$$

The latter provides the best performance which can be achieved by any reduction procedure. In particular, we recall that

$$\kappa_i(\mathcal{M}) \leq \sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S_i^{\text{post}}). \quad (46)$$

On the other hand, we have from (14) that

$$\sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S_i^{\text{post}}) \leq \kappa_i(\mathcal{M}_{\text{post}}) \leq \kappa_i(\mathcal{M}_{\text{prior}}). \quad (47)$$

As discussed in Section 3.2, $\kappa_i(\mathcal{M}_{\text{prior}})$ is the best-achievable worst-case performance when only $\mathcal{M}_{\text{prior}}$ (but no partial observations) is taken into account in the construction of the approximation subspace. Under assumption (41), the Kolmogorov i -width of $\mathcal{M}_{\text{prior}}$ takes again a very simple form, *i.e.*,

$$\kappa_i(\mathcal{M}_{\text{prior}}) = \begin{cases} \infty & \text{if } i < n, \\ \hat{\epsilon} & \text{otherwise.} \end{cases} \quad (48)$$

The gap between $\kappa_i(\mathcal{M})$ and $\kappa_i(\mathcal{M}_{\text{prior}})$ represents the potential improvement which can be obtained by the presence of observations. Theorem 1 provides an upper bound on the minimal improvement which can be attained by considering S_i^{post} as an approximation subspace. In the simple setup considered here, this improvement can be discussed at two different levels:

- The transition from an infinite to a finite approximation error: we ideally wish to obtain a finite projection error for an approximation subspace whose dimension is as close as possible to k (we remind the reader that we assume $\mathcal{M} = \Sigma$ in the present discussion);
- The approximation error for large dimensions of the approximation subspace: when the size of the approximation subspace increases, we ideally wish to attain a projection error as small as possible.

Obviously, achieving a finite projection error when reducing the solution manifold \mathcal{M} (resp. prior manifold $\mathcal{M}_{\text{prior}}$) requires to consider approximation subspaces of dimensionalities greater than or equal to k (resp. n). On

the other hand, we see from the definitions of $\bar{\kappa}_i$ and $\bar{\bar{\kappa}}_i$ that the approximation subspace S_i^{post} enables to attain a finite projection error as soon as its dimension is greater than or equal to $\min(k^*, k + \dim(W^\perp))$. By definition, k^* is always greater than or equal to k but smaller than or equal to n .

The number $k + \dim(W^\perp)$ corresponds to the ‘‘intrinsic dimensionality’’ of \mathcal{M} (that is $\dim(U) = k$) plus the number of dimensions of \mathcal{H} which are not measured through our observation operator (that is $\dim(W^\perp)$). The number $k^* \triangleq \min(n, k + n - q)$ has also an easy interpretation. The first term in the minimum corresponds to the number of directions in which we have a priori an infinite uncertainty about the position of \mathcal{M} , that is $\dim(V) = n$. The second term is equal to the intrinsic dimensionality of \mathcal{M} (that is $\dim(U) = k$) plus the number of components of the prior subspace V which cannot be measured via our observation operator (that is $n - q$). The ‘‘non-observability’’ of some directions of V means that if some elements of \mathcal{M} have nonzero components in these particular directions, the collected observations provide no information about their magnitudes. When $k^* = k + n - q$, the terms ‘‘ k ’’ and ‘‘ $n - q$ ’’ thus have different meanings: k is the number of directions along which the elements of \mathcal{M} *do* have a large variation whereas $n - q$ represents the number of directions along which, given the received observations, \mathcal{M} *could* have a large variation.

We note that $k + \dim(W^\perp) \leq k^*$ only if the number of collected observations is large as compared to the dimension of \mathcal{H} . In particular, if $\dim(\mathcal{H}) = \infty$, we have $\dim(W^\perp) = \infty$ so that the transition from infinite to finite approximation error always occurs at k^* , *i.e.*, $\bar{\kappa}_i \leq \bar{\bar{\kappa}}_i = \infty \forall i$. Moreover, we have $k^* < n$ as soon as $k < q$, that is when the number of observable components of V is larger than the intrinsic dimensionality of \mathcal{M} . In particular, in the case $\dim(\mathcal{H}) = \infty$, a finite approximation error occurs at $k^* = k$ when $q = n$, that is all the components of V are observed.

Let us now discuss the projection error which can be attained for approximation subspaces of ‘‘sufficient’’ dimensionalities. From (45)-(46), we note that the best achievable performance is lower bounded by ϵ . This projection error can be attained for $i \geq k + \dim(W^\perp)$ as suggested by Theorem 1 and the behavior of $\bar{\bar{\kappa}}_i$ in (43). Nevertheless, as mentioned previously, $\dim(W^\perp)$ can be very large when the number of collected observations is small as compared to the dimension of the ambient space \mathcal{H} . In particular, when $\dim(\mathcal{H}) = \infty$, the transition of $\bar{\bar{\kappa}}_i$ to ϵ never occurs.

In such a case, as suggested by Theorem 1, the behavior of $\kappa_i(\mathcal{M}_{\text{post}})$ is upper bounded by $\bar{\kappa}_i$. For large values of i , we can see that $\bar{\kappa}_i$ converges to $\hat{\epsilon}$. Hence, as far as the number of observations is small with respect to

the dimension of \mathcal{H} , the best projection error which can be achieved for approximation subspaces of “moderate” dimensions seems to be related to the width of the prior manifold $\mathcal{M}_{\text{prior}}$. On the other hand, the rate at which the projection error tends to $\hat{\epsilon}$ is connected to the conditioning of \mathbf{G} . Indeed, we see from (42) that $\bar{\kappa}_i$ behaves like $(\epsilon + \hat{\epsilon})\lambda_{q-(i-k^*)}^{-1}$ in the range $i = k^*, \dots, n - 1$. If $\epsilon \ll \hat{\epsilon}$, we thus have $\bar{\kappa}_i \simeq \hat{\epsilon}\lambda_{q-(i-k^*)}^{-1} \simeq \hat{\epsilon}$ provided that $\lambda_{q-(i-k^*)} \simeq 1$. In such cases, we may thus hope to attain an approximation error close to $\hat{\epsilon}$ for $i < n$, hence improving over the best performance achievable from the prior model, *i.e.*, (48).

We remind the reader that the Gram matrix \mathbf{G} (and thus its singular values $\{\lambda_j\}_{j=1}^{\min(m,n)}$) characterizes the interplay between the prior and observation subspaces, V and W . In particular, some of the singular values λ_j will be close to one (resp. zero) if some directions of V are almost included in (resp. orthogonal to) W . For example, if $V \subseteq W$, we obtain $p = q = n$ and

$$\lambda_j = 1 \quad j = 1, \dots, n,$$

so that $\bar{\kappa}_i = \hat{\epsilon} + \epsilon \simeq \hat{\epsilon}$ for $i \geq k$ when $\epsilon \ll \hat{\epsilon}$. On the other extreme, assuming that W is orthogonal to V leads to

$$\lambda_j = 0 \quad j = 1, \dots, \min(m, n),$$

so that $p = q = 0$ and $k^* = n$. In such a case, the performance cannot be improved over that obtained by exploiting the prior manifold $\mathcal{M}_{\text{prior}}$ only.

To conclude this section, we revisit the empirical results presented in Section 5 in the light of Theorem 1. We focus our discussion on Fig. 3 which is the more amenable to a connection with our theoretical results. First, let us draw a link between the simulation setup in Fig. 3 and the parameters defining $\bar{\kappa}_i$ and $\bar{\bar{\kappa}}_i$. From the dark blue plain curve in Fig. 2, it can be seen that \mathcal{M} is included in a degenerate ellipsoid with $k = 4$ and width $\epsilon \sim 10^{-5}$.^s As far as the case $L = 1$ was concerned, we considered two different dimensions for the subspace V , that is $n = 25$ or $n = 45$. These two choices led to $\hat{\epsilon} \sim 10^{-2}$ and $\hat{\epsilon} \sim 10^{-4}$, respectively. Hence, we have $\epsilon \ll \hat{\epsilon}$ in both cases. The observation subspace W was drawn uniformly at random, so that the situation $p = 0$ and $q = \min(m, n)$ occurs almost surely in our

^sIn the discussion below, we implicitly assume that the 4-dimensional subspace defining the ellipsoid is included in the prior subspace V , so that our simulation results can be discussed in light of Theorem 1.

simulations; in this case, k^* takes the simple form $k^* = k + n - \min(m, n)$. We considered a number of observations equal to either $m = 25$ or $m = 45$.

For these choices of parameters and for the dimensions of the approximation subspaces considered in our simulation (that is $i = 1, \dots, 50$), the bound $\bar{\kappa}_i$ prevails in Theorem 1. Indeed, $\dim(W^\perp) \geq \dim(\mathcal{H}) - 45 = 2113 - 45 = 2068$ so that $k + \dim(W^\perp) \geq 2072$ and $\bar{\kappa}_i = \infty$ for the range of interest.

Although Theorem 1 only provides an upper bound on $\kappa_i(\mathcal{M}_{\text{post}})$, the empirical performance presented in Fig. 3 seems to be in good accordance with the latter. First, we can notice that the performance achieved by the worst-case approximation subspace S_i^{post} (plain magenta curve) saturates at $\hat{\epsilon}$ for large i . As predicted by Theorem 1, the number of observations does not have an effect on this error floor but rather impacts the rate at which the latter is reached. Theorem 1 also gives good insights into the range of dimensions in which small approximation errors can be attained. Of particular interest for us is the case where $m = 25$, $n = 45$ for which $k^* = k + n - m = k + 20$ ($k^* = k$ in all other cases); $n - m = 20$ corresponds to the number of dimensions of V which are not observed. In practice, we can therefore not expect to have a small (worst-case) projection error if $i \leq 20$. This is what we observe^t in Fig. 3.

7 Proof of Theorem 1

In this section, we provide the main steps of the proof of Theorem 1 (the technical details are postponed to the appendices). The simple observation underlying the proof is as follows: for any $\hat{\mathcal{M}}_{\text{post}}$ such that $\mathcal{M}_{\text{post}} \subseteq \hat{\mathcal{M}}_{\text{post}}$ and $\forall S_i$ with $\dim(S_i) = i$, we have

$$\kappa_i(\mathcal{M}_{\text{post}}) \leq \max_{\mathbf{h} \in \hat{\mathcal{M}}_{\text{post}}} \text{dist}(\mathbf{h}, S_i). \quad (49)$$

The result stated in Theorem 1 then follows from (49) and some specific choices for $\hat{\mathcal{M}}_{\text{post}}$ and S_i . These choices are described in Sections 7.1 and 7.2.

^tThe error is not infinite in this range because of the way we generate the posterior manifold $\mathcal{M}_{\text{post}}$, see Section 5.

7.1 Definition of $\hat{\mathcal{M}}_{\text{post}}$

We give hereafter three possible choices for $\hat{\mathcal{M}}_{\text{post}}$ that will be exploited in our proof in Section 7.2. First, we have by definition

$$\mathcal{M}_{\text{post}} \subseteq \mathcal{M}_{\text{prior}}, \quad (50)$$

so that $\hat{\mathcal{M}}_{\text{post}} = \mathcal{M}_{\text{prior}}$ is a valid choice. Moreover, since we assume that $\mathcal{M} \subseteq \Sigma$, we also have

$$\mathcal{M}_{\text{post}} \subseteq \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}}, \quad (51)$$

$$\mathcal{M}_{\text{post}} \subseteq \mathcal{M}_{\text{prior}} \cap \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}}. \quad (52)$$

The right-hand sides of (51)-(52) thus constitute two other possible choices for $\hat{\mathcal{M}}_{\text{post}}$.

A precise mathematical characterization of $\cup_{\mathbf{h} \in \Sigma} (\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}})$ is given by (see Appendix A)

$$\mathcal{M}_{\text{prior}} \cap \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}} = \hat{U} \oplus \mathcal{E}, \quad (53)$$

where

- $\hat{U} \triangleq U \oplus \text{span} \left(\left\{ \mathbf{v}_j^* \right\}_{j=q+1}^n \right)$ is a linear subspace
- \mathcal{E} is an ellipsoid defined as follows: $\mathbf{h} \in \mathcal{E}$ if and only if \mathbf{h} can be written as

$$\mathbf{h} = \sum_{j=p+1}^q \lambda_j^{-1} \left((1 - \lambda_j^2)^{\frac{1}{2}} a_j - b_j \right) \phi_j + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \psi_j + \sum_{j=1}^m a_j \mathbf{w}_j^*, \quad (54)$$

where $\phi_j \triangleq (1 - \lambda_j^2)^{-\frac{1}{2}} (\mathbf{v}_j^* - \lambda_j \mathbf{w}_j^*)$, $\{\psi_j\}_{j=1}^{\dim(W^\perp \cap V^\perp)}$ is an (arbitrary) ONB of $W^\perp \cap V^\perp$, and the coefficients a_j , b_j , c_j obey the following constraints:

$$\begin{cases} \sum_{j=q+1}^m a_j^2 + \sum_{j=p+1}^q b_j^2 + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j^2 \leq \hat{\epsilon}^2, \\ \sum_{j=1}^m a_j^2 \leq \epsilon^2. \end{cases} \quad (55)$$

The vectors \mathbf{v}_j^* , \mathbf{w}_j^* and ϕ_j which appear in the above characterization are those introduced in Section 4.2 with the following conventions: V is the subspace characterizing $\mathcal{M}_{\text{prior}}$ in (41); W is the observation subspace introduced in (3).

7.2 Computation of upper bounds on $\kappa_i(\mathcal{M}_{\text{post}})$

In this section, we prove the upper bound stated in Theorem 1 by exploiting (49) for different choices of $\hat{\mathcal{M}}_{\text{post}}$ and S_i . The upper bound $\bar{\kappa}_i$ straightforwardly derives from the definition of $\hat{\mathcal{M}}_{\text{post}}$ given in (51); $\bar{\kappa}_i$ results from a combination of the upper bounds obtained from the definition of $\hat{\mathcal{M}}_{\text{post}}$ in (50) and (52). The detailed calculations are provided below.

7.2.1 Case $\hat{\mathcal{M}}_{\text{post}} = \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}}$

: We first note that

$$\cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}} = \{\mathbf{h} : \text{dist}(\mathbf{h}, \cup_{\mathbf{h} \in U} H_{\mathbf{h}}) \leq \epsilon\}, \quad (56)$$

where $\cup_{\mathbf{h} \in U} H_{\mathbf{h}}$ is a subspace of dimension at most equal to $k + \dim(W^\perp)$. Hence, if we let S_i be such that $\cup_{\mathbf{h} \in U} H_{\mathbf{h}} \subseteq S_i$ for $i \geq k + \dim(W^\perp)$, we obtain

$$\kappa_i(\mathcal{M}_{\text{post}}) \leq \epsilon. \quad (57)$$

7.2.2 Case $\hat{\mathcal{M}}_{\text{post}} = \mathcal{M}_{\text{prior}}$

: Since $\kappa_i(\mathcal{M}_{\text{post}}) \leq \kappa_i(\mathcal{M}_{\text{prior}})$, setting $\hat{\mathcal{M}}_{\text{post}} = \mathcal{M}_{\text{prior}}$ in (49) leads to

$$\kappa_i(\mathcal{M}_{\text{post}}) \leq \begin{cases} \infty & i = 1, \dots, n-1, \\ \hat{\epsilon} & i = n, \dots, \dim(\mathcal{H}) - 1. \end{cases} \quad (58)$$

7.2.3 Case $\hat{\mathcal{M}}_{\text{post}} = \mathcal{M}_{\text{prior}} \cap \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}}$

: We note that the dimension of the subspace \hat{U} defined in Section 7.1 is at most equal to $\min(n, k + n - q) = k^*$, where $k^* \leq n$ follows from the inclusion $U \subseteq V$. For $i = k^*, \dots, k^* + q - p - 1$, we set $S_i = \hat{U} \oplus R$, where R is an $(i - k^*)$ -dimensional subspace given by

$$R \triangleq \text{span} \left(\{\phi_j\}_{j=q-(i-k^*)+1}^q \right).$$

We note that

$$\max_{\mathbf{h} \in \hat{\mathcal{M}}_{\text{post}}} \text{dist}(\mathbf{h}, S_i) = \max_{\mathbf{h} \in \hat{\mathcal{M}}_{\text{post}}} \|P_{S_i^\perp}(\mathbf{h})\|.$$

Moreover, because of the characterization of $\mathcal{M}_{\text{prior}} \cap \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}}$ in (53)-(55), we have that

$$\begin{aligned} \|P_{S_i^\perp}(\mathbf{h})\| &= \left\| P_{S_i^\perp} \left(\sum_{j=p+1}^{q-(i-k^*)} \lambda_j^{-1} \left((1 - \lambda_j^2)^{\frac{1}{2}} a_j - b_j \right) \phi_j + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \psi_j + \sum_{j=1}^m a_j \mathbf{w}_j^* \right) \right\|, \\ &\leq \left\| \sum_{j=p+1}^{q-(i-k^*)} \lambda_j^{-1} \left((1 - \lambda_j^2)^{\frac{1}{2}} a_j - b_j \right) \phi_j + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \psi_j + \sum_{j=1}^m a_j \mathbf{w}_j^* \right\|, \\ &= \left(\sum_{j=p+1}^{q-(i-k^*)} \lambda_j^{-2} \left((1 - \lambda_j^2)^{\frac{1}{2}} a_j - b_j \right)^2 + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j^2 + \sum_{j=1}^m a_j^2 \right)^{\frac{1}{2}}, \end{aligned} \quad (59)$$

holds for any $\mathbf{h} \in \mathcal{M}_{\text{prior}} \cap \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}}$, where the last equality follows from the orthogonality of $\{\mathbf{w}_j^*\}_{j=1}^m$, $\{\phi_j\}_{j=p+1}^q$ and $\{\psi_j\}_{j=1}^{\dim(W^\perp \cap V^\perp)}$ (see Appendix A.1).

Finally, since the coefficients a_i , b_i and c_i must satisfy the constraints in (55), we have

$$\begin{aligned} \max_{\mathbf{h} \in \hat{\mathcal{M}}_{\text{post}}} \text{dist}(\mathbf{h}, S_i) &\leq \left(\lambda_{q-(i-k^*)}^{-2} \left(\left((1 - \lambda_{q-(i-k^*)}^2)^{\frac{1}{2}} \epsilon + \hat{\epsilon} \right)^2 + \epsilon^2 \right)^{\frac{1}{2}}, \\ &\leq \lambda_{q-(i-k^*)}^{-1} (\hat{\epsilon} + \epsilon). \end{aligned} \quad (60)$$

We follow the same reasoning for $i = k^* + q - p, \dots, k^* + q - 1$. We set

$$S_i = \hat{U} \oplus \text{span} \left(\{\phi_j\}_{j=p+1}^q \right) \oplus R,$$

where R is some arbitrary $(i - (k^* + q - p))$ -dimensional subspace. From the particularization of $\mathcal{M}_{\text{prior}} \cap \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}}$ in (53)-(55), we have for any $\mathbf{h} \in$

$\mathcal{M}_{\text{prior}} \cap \cup_{\mathbf{h} \in \Sigma} H_{\mathbf{h}}$:

$$\begin{aligned} \left\| P_{S_i^\perp}(\mathbf{h}) \right\| &= \left\| P_{S_i^\perp} \left(\sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \boldsymbol{\psi}_j + \sum_{j=1}^m a_j \mathbf{w}_j^* \right) \right\|, \\ &\leq \left\| \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \boldsymbol{\psi}_j + \sum_{j=1}^m a_j \mathbf{w}_j^* \right\|, \\ &= \left(\sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j^2 + \sum_{j=1}^m a_j^2 \right)^{\frac{1}{2}}, \end{aligned}$$

where the last equality results from the orthogonality of $\{\mathbf{w}_i^*\}_{i=1}^m$ and $\{\boldsymbol{\psi}_i\}_{i=1}^{\dim(W^\perp \cap V^\perp)}$.

Since the coefficients a_i and c_i must satisfy the constraints in (55), we finally obtain

$$\max_{\mathbf{h} \in \mathcal{M}_{\text{post}}} \text{dist}(\mathbf{h}, S_i) \leq (\hat{\epsilon}^2 + \epsilon^2)^{\frac{1}{2}} \leq \hat{\epsilon} + \epsilon. \quad (61)$$

We note that since $\lambda_j = 1$ for $j = 1, \dots, p$, the upper bounds stated in (60) and (61) can be jointly rewritten as

$$\max_{\mathbf{h} \in \mathcal{M}_{\text{post}}} \text{dist}(\mathbf{h}, S_i) \leq \lambda_{q-(i-k^*)}^{-1} (\hat{\epsilon} + \epsilon), \quad (62)$$

for $i = k^*, \dots, k^* + q - 1$.

7.2.4 Definition of $\bar{\kappa}_i$ and $\bar{\bar{\kappa}}_i$

: $\bar{\bar{\kappa}}_i$ is a direct consequence of the bound derived in Section 7.2.1. The definition of $\bar{\kappa}_i$ results from a combination of the bounds (58) and (62) exposed in Sections 7.2.2 and 7.2.3.

In order to obtain the tightest bound on $\kappa_i(\mathcal{M}_{\text{post}})$, $\bar{\kappa}_i$ must be equal to the minimum of (58) and (62) for each index i at which both bounds are defined. Since $\epsilon + \hat{\epsilon} \geq \hat{\epsilon}$, the bound in (58) is always smaller than (62) when $i \geq n$. This leads to the last line in (42). On the other hand, when $i < n$ the bound in (58) is infinite and (62) thus takes the lead. This results in the second line in (42).

We note that the bound in (62) is always well-defined in the range $i = k^*, \dots, n - 1$ since

$$\begin{aligned} k^* + q - 1 &= \min(n + q - 1, n + k - 1), \\ &\geq n. \end{aligned}$$

8 Conclusions

In this paper, we tackle the problem of finding a good approximation subspace for a solution manifold \mathcal{M} . Unlike in the standard setup where the solution manifold is assumed to be known, we assume that only partial information is available on the latter. More specifically, we suppose that we have the following information at our disposal: *i*) we know that the target manifold is included in a larger set, dubbed “prior manifold”; *ii*) we have access to a set of partial linear observations for each element of \mathcal{M} . This setup corresponds, for example, to the ubiquitous situation where some parameters of the system to approximate are imperfectly known but some sensing device can provide us with partial measurements of the state of the system.

In this work, we thus address the following questions: how to combine the prior knowledge and the collected measurements to build a good approximation subspace for \mathcal{M} ? In particular, what performance can one expect?

We provide an answer to these questions at both a practical and theoretical level. From a worst-case perspective, we show that the best-achievable performance is characterized by the Kolmogorov width of a well-defined manifold, the so-called “posterior manifold”. Motivated by this finding, we propose a tractable algorithm, combining samples from the posterior manifold and a greedy procedure, to achieve performance close to the optimal solution. The theoretical behavior of the proposed methodology is finally studied in a simplified scenario, where the prior manifold is assumed to be a degenerate ellipsoid. We emphasize that the performance achievable in the partially-informed setup is highly dependent on the behavior of the singular values of the projector between the prior and the observation subspaces.

A Technical Details

A.1 Space Decomposition in the Suitable Bases

In this section we elaborate on some simple facts that will be useful in the rest of the appendix. The material presented hereafter takes the form of two remarks (see Remarks A and B below) and mainly derives from the discussion

carried out in [18, section 2.4]. In particular, we exploit the following lemma from [18]:

Lemma 2. *Let X and Y be two linear subspaces of \mathcal{H} . Then,*

$$Y = P_Y(X) \oplus (Y \cap X^\perp), \quad (63)$$

where $P_Y(X)$ and $(Y \cap X^\perp)$ are orthogonal.

Remark A: \mathcal{H} can be decomposed as the direct sum of four orthogonal subspaces, namely:

$$\mathcal{H} = \underbrace{P_W(V) \oplus (W \cap V^\perp)}_{=W} \oplus \underbrace{P_{W^\perp}(V) \oplus (W^\perp \cap V^\perp)}_{=W^\perp}. \quad (64)$$

This follows from the application of Lemma 2 with $Y \leftarrow W$, $X \leftarrow V$ and $Y \leftarrow W^\perp$, $X \leftarrow V$ respectively.

Moreover, ONBs of the first three spaces in (64) can be derived as a function of the suitable bases $\{\mathbf{w}_j^*\}_{j=1}^m$, $\{\mathbf{v}_j^*\}_{j=1}^n$ defined in Section 4.2. First, $\{\mathbf{w}_j^*\}_{j=1}^q$ is an ONB of $P_W(V)$ by definition.^u Second, since $\{\mathbf{w}_j^*\}_{j=1}^m$ is an ONB for W and, from Lemma 2, $W = P_W(V) \oplus (W \cap V^\perp)$ with $P_W(V)$ orthogonal to $W \cap V^\perp$, we have that $\{\mathbf{w}_j^*\}_{j=q+1}^m$ is an ONB of $W \cap V^\perp$. Third,

$$\left\{ (1 - \lambda_j^2)^{-\frac{1}{2}} (\mathbf{v}_j^* - \lambda_j \mathbf{w}_j^*) \right\}_{j=p+1}^q \cup \{\mathbf{v}_j^*\}_{j=q+1}^n, \quad (65)$$

is an ONB of $P_{W^\perp}(V)$. This can be found by projecting $\{\mathbf{v}_j^*\}_{j=1}^n$ onto W^\perp and observing that the resulting elements are mutually orthogonal. More specifically, exploiting property (20) and using the fact that $\mathbf{w}_j^* = \mathbf{v}_j^*$ when $\lambda_j = 1$, we obtain:

$$\begin{aligned} P_{W^\perp}(\mathbf{v}_j^*) &= \mathbf{v}_j^* - P_W(\mathbf{v}_j^*), \\ &= \begin{cases} \mathbf{0} & \text{if } j = 1, \dots, p, \\ \mathbf{v}_j^* - \lambda_j \mathbf{w}_j^* & \text{if } j = p+1, \dots, q, \\ \mathbf{v}_j^* & \text{if } j = q+1, \dots, n. \end{cases} \end{aligned} \quad (66)$$

Let us note that there is usually no ONB of $W^\perp \cap V^\perp$ which can be expressed as a simple function of $\{\mathbf{w}_j^*\}_{j=1}^m$ and $\{\mathbf{v}_j^*\}_{j=1}^n$. In the sequel, we will thus

^uWe remind the reader that q is defined as $q \triangleq \text{card}(\{j : \lambda_j > 0\})$.

use the notation $\{\boldsymbol{\psi}_j\}_{j=1}^{\dim(W^\perp \cap V^\perp)}$ to denote an arbitrary ONB of $W^\perp \cap V^\perp$.

Remark B: Using Lemma 2, the subspace V^\perp can be decomposed as the following direct sum:

$$V^\perp = P_{V^\perp}(W) \oplus (W^\perp \cap V^\perp), \quad (67)$$

where $P_{V^\perp}(W)$ and $(W^\perp \cap V^\perp)$ are orthogonal subspaces. We also have that

$$\left\{ (1 - \lambda_j^2)^{-\frac{1}{2}} (\mathbf{w}_j^* - \lambda_j \mathbf{v}_j^*) \right\}_{j=p+1}^q \cup \{ \mathbf{w}_j^* \}_{j=q+1}^m, \quad (68)$$

constitutes an ONB of $P_{V^\perp}(W)$. This can be seen by projecting the ONB $\{ \mathbf{w}_j^* \}_{j=1}^m$ onto V^\perp and noticing that the resulting vectors are mutually orthogonal. In particular, exploiting property (20) and using the fact that $\mathbf{w}_j^* = \mathbf{v}_j^*$ when $\lambda_j = 1$, we find:

$$\begin{aligned} P_{V^\perp}(\mathbf{w}_j^*) &= \mathbf{w}_j^* - P_V(\mathbf{w}_j^*), \\ &= \begin{cases} \mathbf{0} & \text{for } j = 1, \dots, p, \\ \mathbf{w}_j^* - \lambda_j \mathbf{v}_j^* & \text{for } j = p+1, \dots, q, \\ \mathbf{w}_j^* & \text{for } j = q+1, \dots, m. \end{cases} \end{aligned} \quad (69)$$

A.2 Characterization of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$

In this section, we provide a mathematical characterization of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ in the case where $\mathcal{M}_{\text{prior}}$ is defined as in (22).

Let $\mathbf{h} \in \mathcal{H}$. By definition, for any $\mathbf{h}' \in \mathcal{H}$, we have

$$\begin{aligned} \mathbf{h}' \in H_{\mathbf{h}} &\iff \langle \mathbf{w}_i^*, \mathbf{h}' \rangle = \langle \mathbf{w}_i^*, \mathbf{h} \rangle \text{ for all } i \leq m, \\ \mathbf{h}' \in \mathcal{M}_{\text{prior}} &\iff \|P_{V^\perp}(\mathbf{h}')\|^2 \leq \hat{\epsilon}^2. \end{aligned} \quad (70)$$

Because of our considerations in Remark A in Section A.1, any $\mathbf{h}' \in \mathcal{H}$ can be expressed as (in the expression below we simply express $\mathbf{h}' \in \mathcal{H}$ in the ONBs of four subspaces appearing in (64))

$$\mathbf{h}' = \sum_{j=1}^m \alpha_j \mathbf{w}_j^* + \sum_{j=p+1}^q \beta_j (1 - \lambda_j^2)^{-\frac{1}{2}} (\mathbf{v}_j^* - \lambda_j \mathbf{w}_j^*) + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \boldsymbol{\psi}_j + \sum_{j=q+1}^n d_j \mathbf{v}_j^*, \quad (71)$$

for some $\alpha_j, \beta_j, c_j, d_j \in \mathbb{R}$. Hereafter, we express the conditions in (70) in terms of these parameters.

On the one hand, the first condition in (70) simply reads

$$\alpha_j = \langle \mathbf{w}_j^*, \mathbf{h} \rangle \text{ for all } j \leq m. \quad (72)$$

On the other hand, the second condition in (70) can be rewritten as follows. Starting from (71) and using the fact that

$$P_{V^\perp}(\mathbf{h}') = \mathbf{h}' - \sum_{i=1}^n \langle \mathbf{v}_i^*, \mathbf{h}' \rangle \mathbf{v}_i^*,$$

we find:

$$\begin{aligned} P_{V^\perp}(\mathbf{h}') &= \sum_{j=p+1}^q \alpha_j (\mathbf{w}_j^* - \lambda_j \mathbf{v}_j^*) + \sum_{j=q+1}^m \alpha_j \mathbf{w}_j^* - \sum_{j=p+1}^q \beta_j \lambda_j (1 - \lambda_j^2)^{-\frac{1}{2}} (\mathbf{w}_j^* - \lambda_j \mathbf{v}_j^*) \\ &\quad + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \boldsymbol{\psi}_j, \end{aligned} \quad (73)$$

$$= \sum_{j=p+1}^q \left(\alpha_j - \beta_j \lambda_j (1 - \lambda_j^2)^{-\frac{1}{2}} \right) (\mathbf{w}_j^* - \lambda_j \mathbf{v}_j^*) + \sum_{j=q+1}^m \alpha_j \mathbf{w}_j^* + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \boldsymbol{\psi}_j. \quad (74)$$

Now, because all the terms in the last expression are orthogonal^a, the second constraint in (70) takes the form:

$$\begin{aligned} \hat{\epsilon}^2 &\geq \sum_{j=p+1}^q \left(\alpha_j - \beta_j \lambda_j (1 - \lambda_j^2)^{-\frac{1}{2}} \right)^2 (1 - \lambda_j^2) + \sum_{j=q+1}^m \alpha_j^2 + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j^2, \\ &= \sum_{j=p+1}^q \lambda_j^2 \left(\alpha_j \lambda_j^{-1} (1 - \lambda_j^2)^{\frac{1}{2}} - \beta_j \right)^2 + \sum_{j=q+1}^m \alpha_j^2 + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j^2. \end{aligned} \quad (75)$$

As a conclusion, (71) together with (72) and (75) fully specify the elements of $\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$. For future use, we re-express this system of equations by making the following change of variable: $b_j = \lambda_j \left(\alpha_j \lambda_j^{-1} (1 - \lambda_j^2)^{\frac{1}{2}} - \beta_j \right)$

^aThis can easily be seen from Remark B in section A.1.

for $j = p+1, \dots, q$. We then obtain that any $\mathbf{h}' \in \mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}}$ can be written as^v

$$\begin{aligned} \mathbf{h}' &= \sum_{j=1}^q \alpha_j \lambda_j^{-1} \mathbf{v}_j^* + \sum_{j=q+1}^m \alpha_j \mathbf{w}_j^* - \sum_{j=p+1}^q b_j \lambda_j^{-1} (1 - \lambda_j^2)^{-\frac{1}{2}} (\mathbf{v}_j^* - \lambda_j \mathbf{w}_j^*) \\ &+ \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \boldsymbol{\psi}_j + \sum_{j=q+1}^n d_j \mathbf{v}_j^*, \end{aligned} \quad (76)$$

with

$$\begin{cases} \alpha_j = \langle \mathbf{w}_j^*, \mathbf{h} \rangle \text{ for all } j \leq m, \\ \sum_{j=q+1}^m \alpha_j^2 + \sum_{j=p+1}^q b_j^2 + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j^2 \leq \hat{\epsilon}^2. \end{cases} \quad (77)$$

A.3 Characterization of $\cup_{\mathbf{h} \in \Sigma} (\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}})$

In this section we show that $\cup_{\mathbf{h} \in \Sigma} (\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}})$ can be expressed as stated in (53)-(55). We start from (76)-(77) and particularize the expression of the α_i 's when $\mathbf{h} \in \Sigma$ to obtain the result.

First, notice that any $\mathbf{h} \in \Sigma$ can (by definition) be written as $\mathbf{h} = \mathbf{u} + \mathbf{z}$ where $\mathbf{u} \in U$, $\mathbf{z} \in B_\epsilon$ and $B_\epsilon = \{\mathbf{h} \in \mathcal{H} : \|\mathbf{h}\| \leq \epsilon\}$ is the $\|\cdot\|$ -ball of radius ϵ . Using the fact that $U \subseteq V$ and $\mathcal{H} = W \oplus W^\perp$, \mathbf{u} and \mathbf{z} can be re-expressed as

$$\mathbf{u} = \sum_{j=1}^n \langle \mathbf{v}_j^*, \mathbf{u} \rangle \mathbf{v}_j^*, \quad (78)$$

$$\mathbf{z} = \sum_{j=1}^m a_j \mathbf{w}_j^* + P_{W^\perp}(\mathbf{z}), \quad (79)$$

we obtain from (20) that

$$\begin{aligned} \alpha_j &\triangleq \langle \mathbf{w}_j^*, \mathbf{h} \rangle, \\ &= \langle \mathbf{w}_j^*, \mathbf{u} \rangle + \langle \mathbf{w}_j^*, \mathbf{z} \rangle, \\ &= \begin{cases} \lambda_j \langle \mathbf{v}_j^*, \mathbf{u} \rangle + a_j & \text{if } j = 1, \dots, q, \\ a_j & \text{if } j = q+1, \dots, m. \end{cases} \end{aligned} \quad (80)$$

^vTo obtain the first term in (76), we use the fact that $\sum_{j=p+1}^q \alpha_j \lambda_j^{-1} \mathbf{v}_j^* + \sum_{j=1}^p \alpha_j \mathbf{w}_j^* = \sum_{j=1}^q \alpha_j \lambda_j^{-1} \mathbf{v}_j^*$ since $\mathbf{v}_j^* = \mathbf{w}_j^*$ and $\lambda_j = 1$ for $j \leq p$, see Section 4.2.

Hence, the first term in (76) becomes:

$$\sum_{j=1}^q \alpha_j \lambda_j^{-1} \mathbf{v}_j^* = \sum_{j=1}^q \langle \mathbf{v}_j^*, \mathbf{u} \rangle \mathbf{v}_j^* + \sum_{j=1}^q \lambda_j^{-1} a_j \mathbf{v}_j^*. \quad (81)$$

Moreover, because $\mathbf{z} \in B_\epsilon$, the a_j 's must verify:

$$\sum_{j=1}^m a_j^2 \leq \epsilon^2,$$

since

$$\sum_{j=1}^m a_j^2 = \|P_W(\mathbf{z})\|^2 \leq \|\mathbf{z}\|^2 \leq \epsilon^2.$$

Combining these results and re-arranging the terms in (76), we obtain that $\cup_{\mathbf{h} \in \Sigma} (\mathcal{M}_{\text{prior}} \cap H_{\mathbf{h}})$ is characterized by the following set of equations:

$$\begin{aligned} \mathbf{h}' &= \sum_{j=1}^q \langle \mathbf{v}_j^*, \mathbf{u} \rangle \mathbf{v}_j^* + \sum_{j=q+1}^n d_j \mathbf{v}_j^* - \sum_{j=p+1}^q b_j \lambda_j^{-1} (1 - \lambda_j^2)^{-\frac{1}{2}} (\mathbf{v}_j^* - \lambda_j \mathbf{w}_j^*) \\ &+ \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j \boldsymbol{\psi}_j + \sum_{j=q+1}^m a_j \mathbf{w}_j^* + \sum_{j=1}^q a_j \lambda_j^{-1} \mathbf{v}_j^*, \end{aligned} \quad (82)$$

with

$$\begin{cases} \sum_{j=q+1}^m a_j^2 + \sum_{j=p+1}^q b_j^2 + \sum_{j=1}^{\dim(W^\perp \cap V^\perp)} c_j^2 \leq \epsilon^2, \\ \sum_{j=1}^m a_j^2 \leq \epsilon^2. \end{cases} \quad (83)$$

Because $\mathbf{u} \in U$ and there is no constraint on the d_j 's, the first two terms in (82) define the subspace $P_V(U) \oplus \text{span} \left(\left\{ \mathbf{v}_j^* \right\}_{j=q+1}^n \right) = U \oplus \text{span} \left(\left\{ \mathbf{v}_j^* \right\}_{j=q+1}^n \right)$.

It thus remains to show that the last terms in (82) together with the constraints (83) define \mathcal{E} as in (54)-(55). This simply follows by re-expressing the last term in (82), that is $\sum_{j=1}^q a_j \lambda_j^{-1} \mathbf{v}_j^*$, in terms of its components in W and W^\perp . More specifically, using the suitable bases for $P_W(V)$ and $P_{W^\perp}(V)$ defined in Section A.1, we find

$$\begin{aligned} P_W \left(\sum_{j=1}^q \lambda_j^{-1} a_j \mathbf{v}_j^* \right) &= \sum_{j=1}^q a_j \mathbf{w}_j^*, \\ P_{W^\perp} \left(\sum_{j=1}^q \lambda_j^{-1} a_j \mathbf{v}_j^* \right) &= \sum_{j=p+1}^q a_j \lambda_j^{-1} (\mathbf{v}_j^* - \lambda_j \mathbf{w}_j^*). \end{aligned}$$

The result then follows by plugging these expressions into (82).

B Weak and algebraic formulations of the Thermal-block Problem

Let $H^1(\Omega)$ denotes the Sobolev space of order 1 on Ω and let

$$X = \{\mathbf{h} \in H^1(\Omega) : \mathbf{h} \text{ satisfies the Dirichlet conditions of problem (32)}\}.$$

Using the derivations of [21, Section 2.3.1], we have that the weak formulation of (32) can be written as

$$\text{find } \mathbf{h} \in X \text{ such that } a(\mathbf{h}, \mathbf{h}') = b(\mathbf{h}') \quad \forall \mathbf{h}' \in X \quad (84)$$

where

$$\begin{aligned} a(\mathbf{h}, \mathbf{h}') &= \int_{\Omega} k(\mathbf{x}, \theta) \nabla^{\top} \mathbf{h} \nabla \mathbf{h}' d\mathbf{x}, \\ b(\mathbf{h}') &= \int_{\Omega} s(\mathbf{x}) \mathbf{h}' d\mathbf{x} + \int_{\Gamma_1} c k(\mathbf{x}, \theta)^{-1} \mathbf{h}' d\mathbf{x}. \end{aligned}$$

Let X_{fe} be a finite-dimensional subspace of X (for example generated by a basis of finite elements). A discrete approximation of (84) can then be defined as

$$\text{find } \mathbf{h} \in X_{\text{fe}} \text{ such that } a(\mathbf{h}, \mathbf{h}') = b(\mathbf{h}') \quad \forall \mathbf{h}' \in X_{\text{fe}}. \quad (85)$$

Considering an ONB of X_{fe} , say $\{\varphi_j\}_j^{\text{N}_{\text{fe}}}$, the algebraic formulation of (85) reads

$$\text{find the solution of } \mathbf{A} \mathbf{h} = \mathbf{b}, \quad (86)$$

where the elements of matrix $\mathbf{A} \in \mathbb{R}^{\text{N}_{\text{fe}} \times \text{N}_{\text{fe}}}$ and vector $\mathbf{b} \in \mathbb{R}^{\text{N}_{\text{fe}}}$ are defined as

$$\begin{aligned} A_{i,j} &= a(\varphi_i, \varphi_j), \\ b_j &= b(\varphi_j). \end{aligned}$$

We note in particular that if $c = 0$ and the source term decomposes as

$$s(\mathbf{x}) = \sum_j s_j \varphi_j, \quad (87)$$

we simply have $b_j = s_j$ because of the orthogonality of the φ'_j 's.

Let us now elaborate briefly on the second simulation setup considered in Section 5. We argue that any vector $\mathbf{h} \in \mathbb{R}^{N_{\text{fe}}}$ (or equivalently any $\mathbf{h} \in X_{\text{fe}}$) can be obtained as the solution of (86) (resp. of (85)) by properly choosing the source term $s(\mathbf{x})$. This can be seen as follows. First note that \mathbf{A} is a fixed matrix if the parameter θ is set to a fixed value. Let \mathbf{s} be the vector gathering the s_j 's in (87). Then, by choosing $\mathbf{s} = \mathbf{A}^{-1}\mathbf{h}$, it can be seen that the solution of (86) is \mathbf{h} because we assume that $c = 0$ in our simulation setup. The solution of (85) is obtained by using (87) with $\mathbf{s} = \mathbf{A}^{-1}\mathbf{h}$.

C Summary Appendix

This appendix is intended to ease the reading of the paper by gathering the main elements appearing in the core of the paper. The goal is to allow the reader to access “at a glance” to the most important definitions and relationships.

The paper addresses the problem of finding a good approximation subspace for a solution manifold \mathcal{M} defined as

$$\mathcal{M} = \{\mathbf{h}(\theta) \in \mathcal{H} : \theta \in \Theta\},$$

where $\mathbf{h}(\theta)$ is the solution of some PPDE depending on parameter θ . No particular constraints are imposed on \mathcal{M} . The best possible worst-case approximation error of the elements \mathcal{M} in a subspace of dimension i is given by the Kolmogorov i -width:

$$\kappa_i(\mathcal{M}) = \inf_{S: \dim(S)=i} \left(\sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S) \right).$$

If the above optimization problem admits a minimizer, we denote the latter by S_i^{perf} , *i.e.*,

$$S_i^{\text{perf}} \in \arg \min_{S: \dim(S)=i} \left(\sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S) \right).$$

The construction of the proposed approximation subspace is based on two ingredients that we remind here:

- i) a priori manifold $\mathcal{M}_{\text{prior}}$* : the only constraint we impose on $\mathcal{M}_{\text{prior}}$ is to be such that

$$\mathcal{M} \subseteq \mathcal{M}_{\text{prior}}.$$

In the main body of the paper, we repeatedly use two particular instances of $\mathcal{M}_{\text{prior}}$:

- An intersection of degenerate ellipsoids, *i.e.*,

$$\mathcal{M}_{\text{prior}} = \bigcap_{j=1}^L \{\mathbf{h} : \text{dist}(\mathbf{h}, V_j) \leq \hat{\epsilon}_j\}.$$

- A single degenerate ellipsoid, *i.e.*,

$$\mathcal{M}_{\text{prior}} = \{\mathbf{h} : \text{dist}(\mathbf{h}, V) \leq \hat{\epsilon}\}.$$

ii) *a set of partial observations of \mathcal{M}* : we assume that we collect, $\forall \mathbf{h} \in \mathcal{M}$, a set of noiseless linear measurements:

$$\{\langle \mathbf{w}_j, \mathbf{h} \rangle\}_{j=1}^m,$$

for some orthonormal basis $\{\mathbf{w}_j\}_{j=1}^m$. We denote by W the m -dimensional subspace induced by $\{\mathbf{w}_j\}_{j=1}^m$, *i.e.*,

$$W = \text{span}\left(\{\mathbf{w}_j\}_{j=1}^m\right).$$

Moreover, for a given $\mathbf{h} \in \mathcal{M}$, we denote by $H_{\mathbf{h}}$ the set of element leading to the same measurement sets, *i.e.*,

$$\begin{aligned} H_{\mathbf{h}} &= \{\mathbf{h}' : \langle \mathbf{w}_j, \mathbf{h}' \rangle = \langle \mathbf{w}_j, \mathbf{h} \rangle \text{ for } j = 1, \dots, m\}, \\ &= \{\mathbf{h}' = \mathbf{h} + \mathbf{w}^\perp : \mathbf{w}^\perp \in W^\perp\}. \end{aligned}$$

Our procedure to find a good approximation subspace from the above ingredients is based on the construction of a “posterior” manifold whose definition is given by

$$\mathcal{M}_{\text{post}} = \mathcal{M}_{\text{prior}} \cap \left(\bigcup_{\mathbf{h} \in \mathcal{M}} H_{\mathbf{h}}\right).$$

By definition, $\mathcal{M}_{\text{post}}$ obeys the following relationship

$$\mathcal{M} \subseteq \mathcal{M}_{\text{post}} \subseteq \mathcal{M}_{\text{prior}}.$$

We denote by S_i^{post} the i -dimensional subspace (assuming it exists) minimizing the worst approximation error over the elements of $\mathcal{M}_{\text{post}}$, *i.e.*,

$$S_i^{\text{post}} \in \arg \min_{S: \dim(S)=i} \left(\sup_{\mathbf{h} \in \mathcal{M}_{\text{post}}} \text{dist}(\mathbf{h}, S) \right).$$

This subspace verifies the following property:

$$\kappa_i(\mathcal{M}) \leq \sup_{\mathbf{h} \in \mathcal{M}} \text{dist}(\mathbf{h}, S_i^{\text{post}}) \leq \kappa_i(\mathcal{M}_{\text{post}}) \leq \kappa_i(\mathcal{M}_{\text{prior}}).$$

Finally, in many parts of the paper, we exploit the following decomposition of \mathcal{H} in four orthogonal subspaces:

$$\mathcal{H} = \underbrace{P_W(V) \oplus (W \cap V^\perp)}_{=W} \oplus \underbrace{P_{W^\perp}(V) \oplus (W^\perp \cap V^\perp)}_{=W^\perp}.$$

We also use the following ONBs for each subspace appearing in the above decomposition:

- ONB for $P_W(V)$: $\{\mathbf{w}_j^*\}_{j=1}^q$,
- ONB for $(W \cap V^\perp)$: $\{\mathbf{w}_j^*\}_{j=q+1}^m$,
- ONB for $P_{W^\perp}(V)$: $\left\{ \left(1 - \lambda_j^2\right)^{-\frac{1}{2}} \left(\mathbf{v}_j^* - \lambda_j \mathbf{w}_j^*\right) \right\}_{j=p+1}^q \cup \{\mathbf{v}_j^*\}_{j=q+1}^n$,
- ONB for $(W^\perp \cap V^\perp)$: $\{\boldsymbol{\psi}_j\}_{j=1}^{\dim(W^\perp \cap V^\perp)}$,

where $\{\mathbf{v}_j^*\}_{j=1}^n$ and $\{\mathbf{w}_j^*\}_{j=1}^m$ denotes the “suitable” ONBs for V and W introduced in Section 4.2 and $\{\lambda_j\}_{j=1}^{\min(m,n)}$ are the singular values of the Gram matrix \mathbf{G} defined as $g_{ij} \triangleq \langle \mathbf{w}_i, \mathbf{v}_j \rangle$. The quantities p and q are defined as

$$\begin{aligned} p &\triangleq \text{card}(\{j : \lambda_j = 1\}), \\ q &\triangleq \text{card}(\{j : \lambda_j > 0\}). \end{aligned}$$

D Acknowledgements

The authors wish to thank Stefano Pagani for his kind advice and his help in using the Matlab[®] toolbox “redbKIT”. The authors also thank the “Agence Nationale de la Recherche” (ANR) which partially funded this research through the GERONIMO project.

References

- [1] J. P. Fink and W. C. Rheinboldt, “On the error behavior of the reduced basis technique for nonlinear finite element approximations,” *ZAMM* -

Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik, vol. 63, no. 1, pp. 21–28, 1983.

- [2] Ulrich Hetmaniuk, Radek Tezaur, and Charbel Farhat, “An adaptive scheme for a class of interpolatory model reduction methods for frequency response problems,” *International Journal for Numerical Methods in Engineering*, vol. 93, no. 10, pp. 1109–1124, 2013.
- [3] Philip Holmes, John L. Lumley, and Gal Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, 1996, Cambridge Books Online.
- [4] A. C. Antoulas, “An overview of approximation methods for large-scale dynamical systems,” *Annual Reviews in Control*, vol. 29, no. 2, pp. 181–190, Jan. 2005.
- [5] Alfio Quarteroni, Gianluigi Rozza, and Andrea Manzoni, “Certified reduced basis approximation for parametrized partial differential equations and applications,” *Journal of Mathematics in Industry*, vol. 1, no. 1, pp. 1–49, Dec. 2011.
- [6] R. Everson and L. Sirovich, “Karhunen-Loève procedure for gappy data,” *J. Opt. Soc. Am. A*, vol. 12, no. 8, pp. 1657–1664, Aug. 1995.
- [7] J. M. Beckers and M. Rixen, “EOF calculations and data filling from incomplete oceanographic datasets,” *J. Atmos. Oceanic Technol.*, vol. 20, no. 12, pp. 1839–1856, Dec. 2003.
- [8] Tan Bui-Thanh, Murali Damodaran, and Karen E. Willcox, “Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition,” *AIAA Journal*, vol. 42, no. 8, pp. 1505–1516, Aug. 2004.
- [9] Hasan Gunes, Sirod Sirisup, and George E. Karniadakis, “Gappy data: To krig or not to krig?,” *Journal of Computational Physics*, vol. 212, no. 1, pp. 358–382, Feb. 2006.
- [10] Nathan E. Murray and Lawrence S. Ukeiley, “An application of gappy POD,” *Experiments in Fluids*, vol. 42, no. 1, pp. 79–91, 2007.
- [11] Y. Maday, A. T. Patera, J. D. Penn, and M. Yano, “PBDW state estimation: Noisy observations; configuration-adaptive background spaces; physical interpretations,” in *Proceedings SMAI CANUM 2014, Carry-le-Rouet, France, in ESAIM: Proceedings and Surveys*, 2014.

- [12] P. Héas and C. Herzet, “Reduced-order modeling of hidden dynamics,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 1268–1272.
- [13] T. Cui, Y. M. Marzouk, and K. E. Willcox, “Data-driven model reduction for the Bayesian solution of inverse problems,” *International Journal for Numerical Methods in Engineering*, vol. 102, pp. 966–990, May 2015.
- [14] T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, and A. Spantini, “Likelihood-informed dimension reduction for nonlinear inverse problems,” *Inverse Problems*, vol. 30, no. 11, pp. 114015, Nov. 2014.
- [15] A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, and Y. Marzouk, “Optimal low-rank approximations of Bayesian linear inverse problems,” *ArXiv e-prints*, July 2014.
- [16] Markus Dihlmann and Bernard Haasdonk, “A reduced basis Kalman filter for parametrized partial differential equations,” *ESAIM: Control, Optimisation and Calculus of Variations*, Apr. 2016.
- [17] Yvon Maday, Anthony T. Patera, James D. Penn, and Masayuki Yano, “A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics,” *International Journal for Numerical Methods in Engineering*, vol. 102, no. 5, pp. 933–965, 2015.
- [18] Peter Binev, Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk, “Data assimilation in reduced modeling,” arXiv:1506.04770, June 2015.
- [19] C. Robert, *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*, Springer-Verlag, New-York, 2001.
- [20] A. Pinkus, *n-Widths in Approximation Theory*, vol. 7 of *A Series of Modern Surveys in Mathematics*, Springer Berlin Heidelberg, 1985.
- [21] Alfio Quarteroni, Andrea Manzoni, and Federico Negri, *Reduced Basis Methods for Partial Differential Equations*, vol. 92, Springer International Publishing, 2016.
- [22] Ricardo D. Fierro, *Numerical Linear Algebra*, vol. 40, Society for Industrial and Applied Mathematics, 1998.

- [23] A. T. Patera and G. Rozza, “Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations,” Tech. Rep., Version 1.0, Copyright MIT 2006, to appear in (tentative rubric) MIT Pappalardo Graduate Monographs in Mechanical Engineering, 2006.
- [24] Susanne C. Brenner and L. Ridgway Scott, *The mathematical theory of finite element methods*, Texts in applied mathematics. Springer, New York, Berlin, Paris, 2002.