REDUCED MODELING OF UNKNOWN SYSTEM TRAJECTORIES*

P. HÉAS† AND C. HERZET†

Abstract. This paper deals with model order reduction of parametrical dynamical systems. We consider the specific setup where the distribution of the system's trajectories is unknown but the following two sources of information are available: *(i)* some "rough" prior knowledge on the system's realisations; *(ii)* a set of "incomplete" observations of the system's trajectories. We propose a Bayesian methodological framework to build reduced-order models (ROMs) by exploiting these two sources of information. We emphasise that complementing the prior knowledge with the collected data provably enhances the knowledge of the distribution of the system's trajectories. We then propose an implementation of the proposed methodology based on Monte-Carlo methods. In this context, we show that standard ROM learning techniques, such *e.g.*, Proper Orthogonal Decomposition or Dynamic Mode Decomposition, can be revisited and recast within the probabilistic framework considered in this paper. We illustrate the performance of the proposed approach by numerical results obtained for a standard geophysical model.

1. Introduction. In many fields of Sciences, one is interested in studying the spatio-temporal evolution of a state variable characterised by a differential equation. Numerical discretisation in space and time leads to parameterised high-dimensional systems of equations of the form:

$$\begin{cases} x_t = f_t(x_{t-1}, \theta), \\ x_1 = g(\theta), \end{cases} \qquad (1.1)$$

where $x_t \in \mathbb{R}^n$ is the state variable, $\theta \in \mathcal{T}$ denotes some parameters and $f_t : \mathbb{R}^n \times \mathcal{T} \to \mathbb{R}^n$, $g : \mathcal{T} \to \mathbb{R}^n$. Because (1.1) may correspond to very high-dimensional systems, computing a trajectory $\{x_t\}_{t=1}^T$ may lead to unacceptable computational burdens in some applications.

As a response to this bottleneck, reduced-order models (ROMs) aim at providing "good" approximations of the trajectories of (1.1) (in some particular regimes of interest) via strategies only requiring significantly-reduced computational resources. Among the most familiar reduction techniques, let us mention Galerkin projection using proper orthogonal decomposition (POD) [19] or reduced basis [26], low-rank dynamic mode decomposition (DMD) [6, 17, 21], second-order nonlinear operator approximation [25], balanced truncation [2] or Taylor expansion [14].

All the techniques mentioned above presuppose (explicitly or implicitly) the knowledge of the trajectories that the ROM should accurately approximate. In many contributions, such a knowledge is characterised by the (so-called) solution manifold defined as

$$\mathcal{M} = \{\mathbf{x} = (x_1 \cdots x_T) \in \mathbb{R}^{n \times T} : \mathbf{x} \text{ obeys } (1.1) \text{ for some } \theta \in \mathcal{T}\}, \qquad (1.2)$$

see *e.g.*, [8]. In this paper, we consider a slightly more general setting by assuming that the set of trajectories to reduce are specified by a probability density on $\mathbf{x}$, say $p_\mathbf{X}$.[1] Unlike the standard formulation (1.2), density $p_\mathbf{X}$ then provides information

---

†INRIA Centre Rennes - Bretagne Atlantique, campus universitaire de Beaulieu, 35042 Rennes, France (`patrick.heas@inria.fr, cedric.herzet@inria.fr`)

[1]The latter density can for example be defined via (1.1) by imposing a probability density on $\theta \in \mathcal{T}$.

on both the set of trajectories of interest (which corresponds to $p_{\mathbf{X}} \neq 0$) and their probability of occurence.

Unfortunately, in practice a precise knowledge of $p_{\mathbf{X}}$ is usually not available. In this paper, we thus address the following question: how to build a good ROM for the trajectories specified by $p_{\mathbf{X}}$ when only a rough knowledge of latter density is available but some partial observations of the trajectories are available? More specifically, we will assume that we have the following two sources of information at our disposal in the ROM construction process:

- *a surrogate density* $\tilde{p}_{\mathbf{X}}$: this density gathers all the information the practitioners may have about the system of interest. This density is very general in the sense it can be of any form and it does not need to satisfy any particular constraints. For example, one may know that the trajectories of interest obey (1.1) for some parameters included in the set $\tilde{\mathcal{T}}$. However, the true parameter set $\mathcal{T}$ and the distribution of θ over this set may be unknown. In this case, the surrogate density $\tilde{p}_{\mathbf{X}}$ could for example be defined via (1.1) by using a uniform distribution on θ over $\tilde{\mathcal{T}}$.
- *incomplete observations on the target trajectories*: we assumed that "incomplete" observations of the trajectories are available; these observations, say $\mathbf{y}$, are supposed to obey a *known* conditional model $p_{\mathbf{Y}|\mathbf{X}}$. The term "incomplete" refers to the fact a realisation $\mathbf{x}$ of $p_{\mathbf{X}}$ cannot be unequivocally recovered from its observation $\mathbf{y}$ by inverting the observation model. This situation occurs for instance when only a subset of components of $\mathbf{x}$ are observed or when the observations are corrupted by some noise. As an applicative example, in geophysics, meteorological sensors only provide low-resolution and noisy observations of the atmosphere state.

The main goal of this paper is therefore to propose a methodology taking benefit from these two sources of information to build a "good" ROM for trajectories distributed according to $p_{\mathbf{X}}$.

Before describing the contributions of this paper, we provide an overview of some state-of-the-art methodologies dealing with the problem of ROM construction from incomplete observations. The first contribution dealing with this type of problem is the "Gappy POD" technique proposed by Everson and Sirovich in [13]. The authors propose to construct an approximation subspace for trajectories distributed according to $p_{\mathbf{X}}$ relying on the observed components of $\mathbf{x}$. However, this method releases poor ROM approximations as soon as some directions of the space embedding the trajectories of interest are never observed [15]. This is for example the case when these trajectories are incompletely observed through the *same* observation model.

In order to circumvent this issue, recent works combine an observation model with a surrogate density, in the case of the reduction of a *static* high-dimensional system. On the one hand, several authors propose this observation and prior knowledge combination in a noise-free deterministic setting. In [24], the authors suggest to iteratively enrich the ROM by using point-wise estimates obtained from linear observations and a surrogate model. In [18], the authors propose to refine this approach by including the uncertainty inherent to the point-wise estimates in the reduction process. Stable recovery guarantees are also provided from a worst-case perspective. On the other hand, several works have investigated the context of combining noisy observations with a probabilistic prior. The methodologies naturally rely in this case on poste-

rior probabilities [10, 11, 29]. More precisely, in [11], the author feed a reduced-basis technique with samples of the posterior. In [10, 29], an optimal low-dimensional subspace projection of the posterior distribution is inferred based on its local Gaussian structure.

In this paper, we propose a general data-driven methodology for the reduction of parametric dynamical systems, exploiting incomplete observations. The proposed procedure exploits the two sources of information mentioned previously, namely: *(i)* a surrogate probabilistic characterisation of the trajectories of interest, *(ii)* incomplete observations of these trajectories. The proposed ROM construction relies on the minimisation of the expectation of a bound on the error between the true and reduced trajectories. The expectation relies on a new data-enhanced surrogate density, say $\hat{p}_{\mathbf{X}}$, inferred from the initial surrogate $\tilde{p}_{\mathbf{X}}$ and the partial observations. An approximated solution to this minimisation problem is effectively computed using Monte-Carlo (MC) and Sequential Monte-Carlo (SMC) techniques. The proposed approach relies on the following assumptions:

- stability of ROM inference when using the surrogate $\tilde{p}_{\mathbf{X}}$ in place of $p_{\mathbf{X}}$,
- tightness of the error bound,
- accuracy of the expectation approximation by MC and SMC techniques.

These properties are discussed and empirically assessed in the context of our numerical simulations. The present work complements and generalises the works [10, 11, 18, 24, 29] in two main respects: it proposes a methodology extending these works to the case of dynamical systems; it provides a Bayesian framework generalising any standard ROM construction to the setup where trajectories to be reduced are not fully known.

The rest of this paper is organised as follows. Section 2 first introduces the target problem and presents its surrogate analog. Section 3 then discusses implementation issues and the MC simulation techniques used to obtain a tractable method. Section 4 continues by detailing the particularisation of this methodology to the context of Galerkin projections or low-rank linear approximations. The ability of the method to take into account uncertainty is discussed at the end of this section. A numerical evaluation of the proposed methodology is exposed in Section 5 and conclusions are finally drawn in a last section.

We will use in what follows some notations. Random vectors will be denoted by uppercase letters (as X) and their realisations by lowercase letters (as x). Boldface letters (as $\mathbf{x}$) will indicate matrices, and will be uppercase (as $\mathbf{X}$) for random matrices. $p_{\mathbf{X}}$ will refer to the probability density of $\mathbf{X}$. When there is no ambiguity, the density subscript will be omitted to lighten notations, *i.e.*, $p_{\mathbf{X}}(\mathbf{x}) = p(\mathbf{x})$. The symbol $\| \cdot \|_F$ and $\cdot^{\mathsf{T}}$ will respectively refer to the Frobenius norm and the transpose operator; $\mathbf{i}_k$ will denote the k-dimensional identity matrix. The definition of the Kullback-Leibler distance between two densities $p_{\mathbf{X}}$ and $\tilde{p}_{\mathbf{X}}$ is

$$KL(p_{\mathbf{X}}, \tilde{p}_{\mathbf{X}}) = \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\tilde{p}(\mathbf{x})} d\mathbf{x}$$

2. Target and Surrogate Problems. In this section, we describe the main elements characterising our ROM construction problem. We first define the performance criterion that the ROM should ideally optimise when the target density $p_{\mathbf{X}}$ is known. We then discuss how to modify this target problem when only a surrogate density $\tilde{p}_{\mathbf{X}}$ is known but some incomplete observations $\mathbf{y}$ of the realisations of $p_{\mathbf{X}}$ are available.

The model-order reduction problem can essentially be formulated as follows: for any choice of $\theta \in \mathcal{T}$, find an (easily-computable) approximation $\tilde{\mathbf{x}}$ of $\mathbf{x}$, where $\mathbf{x}$ is specified by (1.1). Most ROM techniques for dynamical models encountered in the literature impose that $\tilde{\mathbf{x}}$ obey a recursion of the form:

$$\begin{cases} \tilde{x}_t = \tilde{f}_t(\tilde{x}_{t-1}, \theta, \mathbf{u}), \\ \tilde{x}_1 = \tilde{g}(\theta, \mathbf{u}), \end{cases} \tag{2.1}$$

where $\tilde{f}_t : \mathbb{R}^n \times \mathcal{T} \times \mathcal{U} \to \mathbb{R}^n$ and $\tilde{g} : \mathcal{T} \to \mathbb{R}^n$ are some functions specifying the ROM via the choice of parameters $\mathbf{u} \in \mathcal{U}$. The nature of $\tilde{f}_t$, $\tilde{g}$ and $\mathbf{u}$ depends on the family of ROMs one considers. We give two examples of choices for $\tilde{f}_t$, $\tilde{g}$ and $\mathbf{u}$ in Sections 4.1 and 4.2. For now, the only ingredient the reader should keep in mind is that, given a family of reduced models, the ROM is fully characterised by the choice of the parameters $\mathbf{u} \in \mathcal{U}$.

In this respect, we will assume hereafter that an ideal choice for $\mathbf{u}$ is given by

$$\mathbf{u}^\star = \arg\min_{\mathbf{u} \in \mathcal{U}} \left\{ \int p(\mathbf{x}) \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 \, d\mathbf{x} \right\}, \tag{2.2}$$

that is, the choice of the ROM parameters should be such that they minimise the mean square approximation error over the target density $p_{\mathbf{X}}$. Here, the notation $\tilde{\mathbf{x}}(\mathbf{u})$ refers to the fact $\tilde{\mathbf{x}}$ is a function of $\mathbf{u}$. Note that it is also a function of parameter $\mathbf{x}$, since $\tilde{\mathbf{x}}$ depends on θ which is itself related to $\mathbf{x}$ through the constraint (1.1). Unfortunately, when $p_{\mathbf{X}}$ is unknown, evaluating $\mathbf{u}^\star$ according to (2.2) is not possible. One possible option to solve this problem may be to substitute $p_{\mathbf{X}}$ in (2.2) by its surrogate density $\tilde{p}_{\mathbf{X}}$, that is

$$\mathbf{u}^\star = \arg\min_{\mathbf{u} \in \mathcal{U}} \left\{ \int \tilde{p}(\mathbf{x}) \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 \, d\mathbf{x} \right\}. \tag{2.3}$$

This formulation does however not take into account the possible presence of observations $\mathbf{y}$ of the realisations $\mathbf{x}$ of $p_{\mathbf{X}}$. In this paper, we thus propose the following alternative surrogate problem:

$$\mathbf{u}^\star = \arg\min_{\mathbf{u} \in \mathcal{U}} \left\{ \int \hat{p}(\mathbf{x}) \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 \, d\mathbf{x} \right\}, \tag{2.4}$$

where $\hat{p}_{\mathbf{X}}$ is defined as

$$\hat{p}(\mathbf{x}) = \int \hat{p}(\mathbf{x}|\mathbf{y}) p(\mathbf{y}) \, d\mathbf{y}, \tag{2.5}$$

with

$$\hat{p}(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})\tilde{p}(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x}')\tilde{p}(\mathbf{x}') \, d\mathbf{x}'}. \tag{2.6}$$

We note that $\hat{p}_{\mathbf{X}}$ obeys the standard relationship between a joint density and its marginal. More specifically, we have from elementary probability theory that $p(\mathbf{x}) = \int p(\mathbf{x}, \mathbf{y}) \, d\mathbf{y} = \int p(\mathbf{x}|\mathbf{y}) p(\mathbf{y}) \, d\mathbf{y}$. Since $p_{\mathbf{X}|\mathbf{Y}}$ depends on $p_{\mathbf{X}}$ and is therefore unknown, we propose to substitute this quantity by the surrogate posterior $\hat{p}_{\mathbf{X}|\mathbf{Y}}$ defined in

(2.6). Similarly, the latter surrogate verifies the standard definition of the posterior probability $p_{\mathbf{X}|\mathbf{Y}}$ with the difference that the target prior $p_{\mathbf{X}}$ has been replaced by $\tilde{p}_{\mathbf{X}}$.

On top of these intuitive arguments motivating the definition of $\hat{p}_{\mathbf{X}}$, the following result provide a theoretical justification to (2.5)-(2.6):

PROPOSITION 2.1. *Let $\hat{p}_{\mathbf{X}}$ defined as in (2.5)-(2.6). Then we have[2]*

$$\mathrm{KL}(p_{\mathbf{X}}, \hat{p}_{\mathbf{X}}) \leq \mathrm{KL}(p_{\mathbf{X}}, \tilde{p}_{\mathbf{X}}) - \mathrm{KL}(p_{\mathbf{Y}}, \hat{p}_{\mathbf{Y}}), \tag{2.7}$$

where

$$\hat{p}(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})\tilde{p}(\mathbf{x})\, d\mathbf{x}. \tag{2.8}$$

Proof: The result is a consequence of the following inequalities:

$$\mathrm{KL}(p(\mathbf{x}), p^{(k)}(\mathbf{x})) \overset{(a)}{=} \int p(\mathbf{x})\Big(\log p(\mathbf{x}) - \log \hat{p}(\mathbf{x})\Big) d\mathbf{x},$$

$$\overset{(b)}{=} \int p(\mathbf{x})\Big(\log p(\mathbf{x}) - \log \tilde{p}(\mathbf{x}) - \log \int p(\mathbf{y}|\mathbf{x})\frac{p(\mathbf{y})}{\hat{p}(\mathbf{y})} d\mathbf{y}\Big) d\mathbf{x},$$

$$\overset{(c)}{\leq} \int p(\mathbf{x})\Big(\log p(\mathbf{x}) - \log \tilde{p}(\mathbf{x}) - \int p(\mathbf{y}|\mathbf{x}) \log \frac{p(\mathbf{y})}{\hat{p}(\mathbf{y})} d\mathbf{y}\Big) d\mathbf{x},$$

$$\overset{(d)}{=} \int p(\mathbf{x})\Big(\log p(\mathbf{x}) - \log \tilde{p}(\mathbf{x})\Big) - \int d\mathbf{y} p(\mathbf{y})\Big(\log p(\mathbf{y}) - \log \hat{p}(\mathbf{y})\Big) d\mathbf{x},$$

where (a) follows from the definition of the Kullback-Leibler distance, (b) from the definition of $\hat{p}_{\mathbf{X}}$ in (2.5)-(2.6), (c) is a consequence of the Jensen's inequality, and (d) follows from $\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\, d\mathbf{x} = p(\mathbf{y})$. $\square$

The operational meaning of Proposition 2.1 is as follows: as far as the Kullback-Leibler distance is considered, the approximation of $p_{\mathbf{X}}$ by $\hat{p}_{\mathbf{X}}$ may only be as good as the approximation of $p_{\mathbf{X}}$ by our surrogate $\tilde{p}_{\mathbf{X}}$. Moreover, when $\mathrm{KL}(p_{\mathbf{Y}}, \hat{p}_{\mathbf{Y}}) > 0$, the proposed approximation $\hat{p}_{\mathbf{X}}$ leads to a strict improvement of the initial surrogate $\tilde{p}_{\mathbf{X}}$, that is $\mathrm{KL}(p_{\mathbf{X}}, \hat{p}_{\mathbf{X}}) < \mathrm{KL}(p_{\mathbf{X}}, \tilde{p}_{\mathbf{X}})$. On the one hand, $\hat{p}_{\mathbf{Y}}$ can be understood as the distribution that the observations should obey if the state variable $\mathbf{x}$ was distributed

---

[2]The proof of the result stated in Proposition 2.1 requires some additional technical assumptions. In order to keep the result stated in this proposition as simple as possible to the practitioner, we mention these assumptions in this footnote. The proposition holds as long as $\hat{p}_{\mathbf{Y}}$ satisfies

$$\mathrm{supp}(p_{\mathbf{Y}|\mathbf{X}} p_{\mathbf{Y}}) \overset{(e)}{\subseteq} \mathrm{supp}(\hat{p}_{\mathbf{Y}}) \overset{(f)}{\subseteq} \mathrm{supp}(p_{\mathbf{Y}}), \quad \forall \mathbf{x}.$$

In particular, if for all $\mathbf{x}$ the density $p_{\mathbf{Y}|\mathbf{X}}$ has an infinite support, then these inclusions are guaranteed by the definition of $p_{\mathbf{Y}}$ and $\hat{p}_{\mathbf{Y}}$. This sufficient condition is satisfied for example in the case where $p_{\mathbf{Y}|\mathbf{X}}$ is a model with Gaussian additive noise. Let us detail the necessity of e) and f). Inclusion e) is needed for the existence of the integral $\int p(\mathbf{y}|\mathbf{x})\frac{p(\mathbf{y})}{\hat{p}(\mathbf{y})} d\mathbf{y}$, which we have assumed to obtain (b). On the other hand, to obtain (c) we have applied the Jensen inequality with the strictly convex function $-log$ on the interval of strictly positive reals. Thus, we need to check that $\forall \mathbf{y}$ we have $\frac{p(\mathbf{y})}{\hat{p}(\mathbf{y})} > 0$ and in particular that $\frac{p(\mathbf{y})}{\hat{p}(\mathbf{y})} \neq 0$, which is guaranteed by inclusion f).

according to $\hat{p}_{\mathbf{X}}$. On the other hand, $p_{\mathbf{Y}}$ corresponds to the actual distribution of the collected observations $\mathbf{y}$. Since $\mathrm{KL}(p_{\mathbf{Y}}, \hat{p}_{\mathbf{Y}}) = 0$ if and only if $p_{\mathbf{Y}}(\mathbf{y}) = \hat{p}_{\mathbf{Y}}(\mathbf{y})$ (almost everywhere), we see that $\hat{p}_{\mathbf{X}}$ leads to a strict improvement over $\tilde{p}_{\mathbf{X}}$ as soon as the empirical distribution of the collected data deviates from the surrogate $\hat{p}_{\mathbf{Y}}$.

We can notice that $\hat{p}_{\mathbf{X}}$ depends on the observations $\mathbf{y}$ via the distribution $p_{\mathbf{Y}}$ in (2.5). The precise knowledge of $p_{\mathbf{Y}}$ is however inaccessible in most experimental setups and in order to build $\hat{p}_{\mathbf{X}}$, the practitioner can only access to a finite set of realisations of the observed random variable. We will detail in the following how $\hat{p}_{\mathbf{X}}$ is approximated from this finite set of observations.

3. Implementation Issues. The target problem exposed in the previous sections is intractable directly. Its resolution will rely on several levels of approximations. They are detailed bellow.

3.1. MC Approximation. In practice, we are often faced to the lack of knowledge of the density $p_{\mathbf{Y}}$. Nevertheless, a practitioner has often access to a finite set of observations

$$\{\mathbf{y}^{(i)} \in \mathbb{R}^{m \times T}; \quad i = 1, \cdots, D\},$$

composed of realisations assumed independent and identically distributed (*i.i.d.*) according to the density $p_{\mathbf{Y}}$. Relying on these observations, we propose to approximate the marginalisation integral (2.5) using a standard MC technique. This leads to the following approximation of the cost function in problem (2.4)

$$\int \hat{p}(\mathbf{x}) \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 d\mathbf{x} \simeq \frac{1}{D} \sum_{i=1}^{D} \int \hat{p}(\mathbf{x}|\mathbf{y}^{(i)}) \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 d\mathbf{x}. \tag{3.1}$$

We remark that, by introducing this MC approximation, observations $\mathbf{y}^{(i)}$ now appear explicitly in the ROM inference problem, on the contrary to the cost function in problem (2.4) which only exhibits a dependence to the unknown density $p_{\mathbf{Y}}$. It is well know that the right hand side of equation (3.1) is an unbiased estimate of the left hand side with an error variance evolving as $\mathcal{O}(D^{-1})$.

3.2. SMC Approximation. In the general case, the density $\hat{p}_{\mathbf{x}|\mathbf{y}^{(i)}}$ appearing in (3.1) is not closed-form and we can often not compute analytically posterior expectations. We pursue an approximation of this density for $i = 1, ..., D$ by an empirical measure of the form

$$\hat{p}(\mathbf{x}|\mathbf{y}^{(i)}) \simeq \frac{1}{N} \sum_{j=1}^{N} w^{(i,j)} \delta_{\boldsymbol{\xi}^{(i,j)}}(\mathbf{x}), \tag{3.2}$$

which relies on a set of N samples $\{\boldsymbol{\xi}^{(i,j)}\}_{j=1}^{N}$ and weights $\{w^{(i,j)}\}_{j=1}^{N}$ with $\boldsymbol{\xi}^{(i,j)} = (\xi_1^{(i,j)} \cdots \xi_T^{(i,j)}) \in \mathbb{R}^{n \times T}$ and $w^{(i,j)} \in \mathbb{R}_+$. This leads to an approximation of the cost function in problem (2.4) by the weighted sum

$$\int \hat{p}(\mathbf{x}|\mathbf{y}^{(i)}) \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 d\mathbf{x} \approx \frac{1}{N} \sum_{j=1}^{N} w^{(i,j)} \|\boldsymbol{\xi}^{(i,j)} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2, \tag{3.3}$$

and combining approximations (3.1) and (3.3), we obtain

$$\int \hat{p}(\mathbf{x}) \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 d\mathbf{x} \simeq \frac{1}{DN} \sum_{i=1}^{D} \sum_{j=1}^{N} w^{(i,j)} \|\boldsymbol{\xi}^{(i,j)} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2. \tag{3.4}$$

In the case of dynamical systems, $\hat{p}_{\mathbf{X}|\mathbf{Y}}$ in (3.3) often exhibits a nested structure which can be sampled in a sequential manner. In particular, the surrogate density $\tilde{p}_{\mathbf{X}}$ often takes the form of a Markov chain defined by a transition kernel and an initial density

$$\begin{cases} \tilde{p}(x_t|x_{t-1}) = \pi_t(x_t, x_{t-1}), \\ \qquad \tilde{p}(x_1) = \eta_1(x_1), \end{cases} \tag{3.5}$$

which will imply the density factorisation

$$\tilde{p}(\mathbf{x}) = \eta_1(x_1) \prod_{t=2}^{T} \pi_t(x_t, x_{t-1}). \tag{3.6}$$

SMC techniques are particularly well suited to this context and constitute tractable methods able to compute efficiently the set of N samples and weights $\{\boldsymbol{\xi}^{(i,j)}, w^{(i,j)}\}_{j=1}^{N}$ involved in (3.2) or (3.3). Among the variety of SMC techniques, the most known methods are sequential importance sampling or bootstrap particle filtering [12]. These algorithms exploit an observation model admitting the factorisation

$$p_{\mathbf{Y}|\mathbf{X}} = \prod_{t=1}^{T} p_{Y_t|X_t}. \tag{3.7}$$

where the random matrix $\mathbf{Y} = (Y_1 \cdots Y_T)$ gathers the observed variable at the T different temporal indexes. For large N, approximation (3.3) by SMC techniques is accurate in the sense it will yield an unbiased (or asymptotically unbiased) estimation of the posterior expectation in the case $\|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2$ is a bounded function of $\mathbf{x}$. Moreover, under this boundedness hypothesis, the variance of the estimation error will decrease at the rate of $\mathcal{O}(N^{-1})$, see *e.g.*, [9]. However, in general, the error norm is not bounded. Although, progress has been recently accomplished in this direction [1], extending these asymptotical unbiased properties and convergence results to the case of unbounded test functions has not been done yet in the context of SMC approximations.

3.3. Practical Identification of a Minimiser. With the simplifications proposed in Sections 3.1 and 3.2, our surrogate optimisation problem takes the form:

$$\mathbf{u}^{\star} = \arg\min_{\mathbf{u} \in \mathcal{U}} \left\{ \sum_{i=1}^{D} \sum_{j=1}^{N} w^{(i,j)} \|\boldsymbol{\xi}^{(i,j)} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 \right\}. \tag{3.8}$$

Unfortunately, (3.8) is typically[3] a non-convex optimisation problem. Hence, designing polynomial-time optimisation procedures ensuring the identification of a global minimiser $\mathbf{u}^{\star}$ of (3.8) for any problem instance is usually out of reach. In other to circumvent this issue, two different approaches are usually suggested in the literature: *(i)* resorting to local optimisation procedures; *(ii)* optimising an upper bound of the cost function in (3.8).

The local optimisation procedures encountered in practice usually derive from iterative gradient descent methods. When the ROM approximation $\tilde{\mathbf{x}}(\mathbf{u})$ satisfies

---

[3] That is for most choices of functions $\tilde{f}_t$ and $\tilde{g}$ encountered in practice. See also Sections 4.1 and 4.2.

a recursion as (2.1), these methods can be efficiently implemented by using *adjoint* procedures, see for example [16, 22]. The drawback of local optimisation procedures is however that they are prone to converge to local optimum of the cost function. In many situations, this behavior may prevent these methods from delivering a solution close to the global minimiser $\mathbf{u}^\star$, leading in turn to poor reduction performance.

In order to circumvent this problem, another approach pursued in the literature consists of optimising an upper-bound of the cost function, that is

$$\mathbf{u}^\star = \arg\min_{\mathbf{u}\in\mathcal{U}}\{J(\mathbf{u})\}, \tag{3.9}$$

where $J(\mathbf{u})$ is such that

$$\sum_{i=1}^{D}\sum_{j=1}^{N} w^{(i,j)}\|\boldsymbol{\xi}^{(i,j)} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 \leq J(\mathbf{u}) \quad \forall \mathbf{u}\in\mathcal{U}. \tag{3.10}$$

Of course, the minimisers of (3.9) usually differ from those of (3.8). Nevertheless, if the behavior of $J(\mathbf{u})$ is "not too far" from the one of $\sum_{i=1}^{D}\sum_{j=1}^{N} w^{(i,j)}\|\boldsymbol{\xi}^{(i,j)} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2$, one may expect the minimisers of (3.9) to be good approximations of the solutions of (3.8). Moreover, the numerical optimisation (3.9) may be far easier than the one of the initial optimisation problem (3.8). We will provide two instances of such scenarios in Section 4.

In the sequel, we will exclusively focus our attention on methodologies based on the optimisation of an upper bound. This is motivated by the fact that, by using this approach, several methodologies known in the reduced-model community can be revisited and extended in the probabilistic framework considered in this paper. Nevertheless, the methodologies based on local optimisation procedure could be applied in a similar way in the framework discussed in this paper.

4. Two Examples. In this section, we illustrate how the procedure presented in Section 3 particularises to two different families of ROMs. In particular, we will emphasise that these particularisations can be seen as generalisations, in a probabilistic framework, of well-known ROM techniques, namely POD and DMD. We also discuss how they can be seen as generalisations of standard ROM constructions based on point estimates.

4.1. Galerkin Projection. The low-rank approximation called *Galerkin* projection of the dynamics (1.1) is obtained by projecting x_t's onto a subspace spanned by the columns of some matrix $\mathbf{u}\in\mathbb{R}^{n\times k}$ with orthonormal columns and where $k < n$, see *e.g.*, [27]. More precisely, it consists in a recursion

$$\begin{cases} z_t = \mathbf{u}^\mathsf{T} f_t(\mathbf{u}z_{t-1},\theta), \\ z_1 = \mathbf{u}^\mathsf{T} g(\theta), \end{cases} \tag{4.1}$$

defining a sequence of k-dimensional variables $\{z_t \in \mathbb{R}^k\}_{t=1}^T$. Because $k < n$, system (4.1) is usually either tractable or efficient methods can be used to simplify computation [3, 5]. Once recursion (4.1) has been evaluated, an approximation of state x_t can simply be obtained as $\tilde{x}_t = \mathbf{u}z_t$. Recursion (4.1) is thus a particularisation of (2.1) with

$$\tilde{f}_t(\tilde{x}_{t-1},\theta,\mathbf{u}) = \mathbf{u}\mathbf{u}^\mathsf{T} f_t(\mathbf{u}\mathbf{u}^\mathsf{T}\tilde{x}_{t-1},\theta),$$
$$\tilde{g}(\theta) = \mathbf{u}\mathbf{u}^\mathsf{T} g(\theta).$$

Setting $\tilde{f}_t$ and $\tilde{g}$ as above implies that (3.8) with the admissible set

$$\mathcal{U} = \{\mathbf{u} \in \mathbb{R}^{n \times k} | \mathbf{u}^\mathsf{T}\mathbf{u} = \mathbf{i}_k\} \tag{4.2}$$

is a non-convex minimisation problem exhibiting a complex sequential structure. Unfortunately, no polynomial-time optimisation methods can ensure the identification of a global minimiser in this context. We resort instead to the optimisation of the following upper bound of the cost function in (3.8)

$$J(\mathbf{u}) = c \sum_i \sum_j w^{(i,j)} \|\boldsymbol{\xi}^{(i,j)} - \mathbf{u}\mathbf{u}^\mathsf{T}\boldsymbol{\xi}^{(i,j)}\|_F^2 \tag{4.3}$$

where $c > 0$. Indeed, using a generalisation of *Céa*'s lemma, we obtain

$$\|\mathbf{x} - \mathbf{u}\mathbf{u}^\mathsf{T}\mathbf{x}\|_F^2 \le \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 \le c\|\mathbf{x} - \mathbf{u}\mathbf{u}^\mathsf{T}\mathbf{x}\|_F^2. \tag{4.4}$$

for any $\mathbf{x} \in \mathbb{R}^{n \times T}$ and with the constant c independent of $\mathbf{u}$. This bound is tight for small c, since it encloses from above and below the error norm. The upper bound in (4.4) is obtained under the assumption that the mapping $\mathbb{R}^{n \times T} \to \mathbb{R}^{n \times T}, \mathbf{x} \to (x_1, -f_1(x_1) + x_2, -f_2(x_2) + x_3, \cdots, -f_{T-1}(x_{T-1}) + x_T)$ is strongly monotone and that it is Lipshitz-continuous for bounded arguments [7, Theorem 5.3.4]. We remark that the optimisation of the bound (4.3) can be seen as a POD-like problem, where standard snapshots are substituted by weighted samples obtained by MC and SMC simulations.

Problem (3.9) with the upper bound (4.3) and the admissible set (4.2) admits a closed-form solution $\mathbf{u}^\star$. Indeed, a well-known result is that the columns of matrices $\mathbf{u}^\star$ are the eigenvectors of matrice $\mathbf{c}\mathbf{c}^\mathsf{T}$ where

$$\mathbf{c} = (\sqrt{w^{(1,1)}}\boldsymbol{\xi}^{(1,1)} \cdots \sqrt{w^{(D,N)}}\boldsymbol{\xi}^{(D,N)}) \in \mathbb{R}^{n \times TDN},$$

associated to its k largest eigenvalues [20]. In practice, eigenvectors of interest can be derived from the eigen-decomposition of the smaller matrix $\mathbf{c}^\mathsf{T}\mathbf{c}$, see [17, Remark 1].

Finally, from the convergence results of MC and SMC techniques exposed in section 3.1 and 3.2, we deduce the following proposition:

PROPOSITION 4.1. *For any positive integer $k \le TDN$ the eigenvectors corresponding to the k largest eigenvalues of matrix $\mathbf{c}\mathbf{c}^\mathsf{T}$ are unbiased estimators of the columns of the matrix solving (3.9) with the upper bound (4.3) and the admissible set (4.2).*

4.2. Low-Rank Linear Approximation. A low-rank linear approximation of the dynamics (1.1) is a particularisation of (2.1) to

$$\begin{aligned} \tilde{f}_t(\tilde{x}_{t-1}, \theta, \mathbf{u}) &= \mathbf{u}\tilde{x}_{t-1}, \\ \tilde{g}(\theta) &= g(\theta), \end{aligned} \tag{4.5}$$

parameterised by some matrix $\mathbf{u} \in \mathbb{R}^{n \times n}$ of rank lower or equal to $k \le n$. Let its singular value decomposition (SVD) be $\mathbf{u} = \mathbf{w_u}\boldsymbol{\sigma_u}\mathbf{v_u}^\mathsf{T}$, with $\mathbf{w_u}, \mathbf{v_u} \in \mathbb{R}^{n \times k}$ and $\boldsymbol{\sigma_u} \in \mathbb{R}^{k \times k}$ so that $\mathbf{w_u}^\mathsf{T}\mathbf{w_u} = \mathbf{v_u}^\mathsf{T}\mathbf{v_u} = \mathbf{i}_k$ and $\boldsymbol{\sigma_u}$ is diagonal. The n-dimensional reduced states $\{\tilde{x}_t\}_{t=1}^T$ are fully determined by the following recursion,

$$\begin{cases} z_t = (\mathbf{v_u}\boldsymbol{\sigma_u})^\mathsf{T}\mathbf{w_u}z_{t-1}, \\ z_2 = (\mathbf{v_u}\boldsymbol{\sigma_u})^\mathsf{T}\mathbf{w_u}^\mathsf{T}g(\theta), \end{cases} \tag{4.6}$$

only involving k-dimensional variables. By multiplying both sides of (4.6) by $\mathbf{w_u}$, we obtain low-rank approximations $\tilde{x}_t = \mathbf{w_u} z_t$ of the n-dimensional states x_t defined in (1.1).

Setting $\tilde{f}_t$ and $\tilde{g}$ as in (4.5) and the admissible set

$$\mathcal{U} = \{\mathbf{u} \in \mathbb{R}^{n \times n} \,|\, \operatorname{rank}(\mathbf{u}) \leq k\} \tag{4.7}$$

induces a non-convex problem (3.8) due to the low-rank constraint and the sequential structure of (4.5). The global minimiser is out of reach in this context. Here again, we choose to resort to the optimisation of the following upper bound of the cost function in (3.8)

$$J(\mathbf{u}) = c \sum_{i=1,j=1}^{D,N} w^{(i,j)} \sum_{t=2}^{T} \|\xi_t^{(i,j)} - \mathbf{u}\xi_{t-1}^{(i,j)}\|_F^2 \tag{4.8}$$

where $c > 0$. Indeed, for any $\mathbf{x} \in \mathbb{R}^{n \times T}$, it is shown in Appendix A that assuming there exists $\lambda \in \mathbb{R}$ such that $\|\mathbf{u}\|_{2,2} \leq \lambda$, the ROM error can be bounded as

$$\sum_{t=2}^{T} \|x_t - \mathbf{u}x_{t-1}\|_2^2 \leq \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2 \leq c \sum_{t=2}^{T} \|x_t - \mathbf{u}x_{t-1}\|_2^2, \tag{4.9}$$

with a constant $c > 0$ function of λ. As in the previous example, this bound is tight as long as c is small since it encloses from above and below the error norm.

We remark that the optimisation of the bound (4.8) over the domain (4.7) can be seen as a low-rank DMD-like problem, where standard snapshots are substituted by weighted samples obtained by MC and SMC simulations. The optimisation of this upper bound admits a closed-form solution as detailed below. Indeed, constituting matrices $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n \times (T-1)DN}$ as

$$\mathbf{a} = (\sqrt{w^{(1,1)}}\boldsymbol{\xi}_{1:T-1}^{(1,1)} \cdots \sqrt{w^{(D,N)}}\boldsymbol{\xi}_{1:T-1}^{(D,N)}),$$
$$\mathbf{b} = (\sqrt{w^{(1,1)}}\boldsymbol{\xi}_{2:T}^{(1,1)} \cdots \sqrt{w^{(D,N)}}\boldsymbol{\xi}_{2:T}^{(D,N)}),$$

with the notations $\boldsymbol{\xi}_{\ell:m}^{(i,j)} = (\xi_\ell^{(i,j)} \cdots \xi_m^{(i,j)})$, the optimisation problem can be rewritten in the synthetic form

$$\mathbf{u}^\star \in \arg\min_{\mathbf{u} \in \mathcal{U}} \|\mathbf{b} - \mathbf{u}\mathbf{a}\|_F^2, \tag{4.10}$$

where $\mathcal{U}$ is given by (4.7). As shown recently in [17, Theorem 3.1], this problem admits a closed-form solution $\mathbf{u}^\star = \mathbf{p}\mathbf{p}^\intercal \mathbf{b}\mathbf{a}^\dagger$, where the columns of $\mathbf{p} \in \mathbb{R}^{n \times k}$ are real orthonormal eigenvectors associated to the largest eigenvalues of matrix $\mathbf{b}\mathbf{b}^\intercal$, and where $\mathbf{a}^\dagger = \mathbf{v_a}\boldsymbol{\sigma}_\mathbf{a}^{-1}\mathbf{w_a}^\intercal$ is the Moore-Penrose pseudo-inverse of $\mathbf{a}$. We note that this solution can be efficiently computed by SVDs, as detailed in [17, Algorithm 1]. Finally, we deduce the following proposition from the convergence results of MC and SMC techniques exposed in section 3.1 and 3.2:

PROPOSITION 4.2. *For any positive integer $k \leq (T-1)DN$, an unbiased estimator of a solution of* (4.10) *is* $\mathbf{u}^\star = \mathbf{p}\mathbf{p}^\intercal \mathbf{b}\mathbf{a}^\dagger$.

4.3. Comparison with ROM Based on Point Estimates. We show that the ROM parameter $\mathbf{u}^\star$ inferred in Section 4.1 and 4.2 differs from the parameter inferred relying on point estimates [28, 24]. The latter approach consists in building the ROM from estimates of the state, say $\hat{\mathbf{x}}^{(i)}$, computed for $i = 1, ..., D$ by combining the received observation $\mathbf{y}^{(i)}$ and the surrogate $\tilde{p}_{\mathbf{X}}$. A common choice to obtain these estimates is to rely on the minimum mean square error (MMSE) estimator

$$\hat{\mathbf{x}}^{(i)} = \int \mathbf{x}\,\hat{p}(\mathbf{x}|\mathbf{y}^{(i)})\,d\mathbf{x}.$$

The parameter of a ROM based on MMSE point estimates is then obtained by solving

$$\arg\min_{\mathbf{u}\in\mathcal{U}}\left\{\sum_{i=1}^{D} \|\hat{\mathbf{x}}^{(i)} - \tilde{\mathbf{x}}(\mathbf{u})\|_F^2\right\}. \tag{4.11}$$

In what follows, we will refer to this particular choice of estimator when invoking ROM based on point estimates.

Analogously to our approach, we may obtain an unbiased (or asymptotically unbiased) approximation of the MMSE estimator using an SMC technique

$$\hat{\mathbf{x}}^{(i)} \simeq \frac{1}{N}\sum_{j=1}^{N} w^{(i,j)}\boldsymbol{\xi}^{(i,j)}. \tag{4.12}$$

Comparing (3.8) with the optimisation problem (4.11) where the $\hat{\mathbf{x}}^{(i)}$'s are approximated with (4.12), we see that our approach can be seen as a generalisation of a point estimate approach where the approximation of $\hat{p}_{\mathbf{X}|\mathbf{Y}}$ relies on N particles rather than on a single one.

Let us further detail the differences between the two approaches. Note that matrix $\mathbf{cc}^\intercal$ or matrices $\mathbf{ba}^\intercal$ and $\mathbf{aa}^\intercal$ introduced previously are MC and SMC approximations of matrices of the form

$$\int p(\mathbf{y})d\mathbf{y} \int \hat{p}(\mathbf{x}|\mathbf{y})\,\mathbf{x}_{1+\ell:T}\mathbf{x}_{1:T-\ell}^\intercal d\mathbf{x}, \quad \ell \in \{0, 1\}, \tag{4.13}$$

where we have used the notations $\mathbf{x}_{\ell:m} = (x_\ell \cdots x_m) \in \mathbb{R}^{n\times(m-\ell+1)}$. In particular, according to (3.1), matrices (4.13) are approximated in our methodology by a MC technique yielding

$$\frac{1}{D}\sum_{i=1}^{D}\int \hat{p}(\mathbf{x}|\mathbf{y}^{(i)})\,\mathbf{x}_{1+\ell:T}\mathbf{x}_{1:T-\ell}^\intercal d\mathbf{x},$$

which can be decomposed as

$$\frac{1}{D(T-\ell)}\sum_{i,t=1+\ell}^{D,T}\left(\int \hat{p}(\mathbf{x}|\mathbf{y}^{(i)})\left(x_{t-\ell} - \hat{x}_{t-\ell}^{(i)}\right)\left(x_t - \hat{x}_t^{(i)}\right)^\intercal d\mathbf{x} + (\hat{x}_{t-\ell}^{(i)})(\hat{x}_t^{(i)})^\intercal\right), \tag{4.14}$$

where we have introduced the MMSE point estimates

$$\hat{x}_{t-\ell}^{(i)} = \int \hat{p}(\mathbf{x}|\mathbf{y}^{(i)})\,x_{t-\ell}\,d\mathbf{x}, \quad \ell \in \{0, 1\}.$$

11

Note that the first term inside the brackets is the cross-covariance relative to $\hat{p}(\mathbf{x}|\mathbf{y}^{(i)})$ of vectors $X_{t-\ell}$ and X_t, while the second term is the second-order moment of this density. ROM based on point estimates rely only on the second-order moment and ignore cross-covariance terms. They approximate matrices (4.13) by

$$\frac{1}{D(T-\ell)} \sum_{i,t=1+\ell}^{D,T} (\hat{x}_{t-\ell}^{(i)})(\hat{x}_t^{(i)})^\intercal. \tag{4.15}$$

Of course, these matrix are identical to (4.14) in the case of vanishing cross-covariances. However, in the general case (4.15) may constitute poor approximations of (4.13). Making a correspondence between cross-covariances and uncertainty, this suggests that the proposed method integrates uncertainty relative to point estimates in the ROM inference process.

5. Numerical Evaluation. We propose to assess the proposed methodology with a standard physical model known as Rayleigh-Bénard convection. After introducing the parametric partial differential equation inducing the high-dimensional system, we propose different variations of the ROM building problem. They differ from each other by their probabilistic models $p_{\mathbf{X}}$, $\hat{p}_{\mathbf{X}}$ and $p_{\mathbf{Y}|\mathbf{X}}$. Based on this setup, we finally evaluate the performance of four different sampling strategies for building a ROM in our uncertain context.

5.1. The Physical Setup. We consider a Rayleigh-Bénard convective system [4]. An incompressible fluid is contained in a bi-dimensional cell and is subject to periodic boundary conditions. The states of interest are the trajectories of the temperature and of the velocity in the cell.

We introduce the following notations to state the evolution equations: the differential operators $\nabla = (\partial_{s_1}, \partial_{s_2})^\intercal$, $\nabla^\perp = (\partial_{s_2}, -\partial_{s_1})^\intercal$ and $\Delta = \partial_{s_1}^2 + \partial_{s_2}^2$ denote the gradient, the curl and the Laplacian with respect to the two spatial dimensions (s_1, s_2); the operator Δ^{-1} is the formal representation of the inverse of Δ. Convection is driven by the two following coupled partial differential equations: at any point of the unit cell $\mathbf{s} = (s_1, s_2) \in [0,1]^2$ and for any time $t \geq 1$, we have

$$\begin{cases} \partial_t b(\mathbf{s}, t) + \mathbf{v}(\mathbf{s}, t) \cdot \nabla b(\mathbf{s}, t) - \rho \Delta b(\mathbf{s}, t) - \rho \nu \partial_{s_1} \tau(\mathbf{s}, t) = 0, \\ \partial_t \tau(\mathbf{s}, t) + \mathbf{v}(\mathbf{s}, t) \cdot \nabla \tau(\mathbf{s}, t) - \Delta \tau(\mathbf{s}, t) - \partial_{s_1} \Delta^{-1} b(\mathbf{s}, t) = 0, \end{cases} \tag{5.1}$$

where $\tau(\mathbf{s}, t) \in \mathbb{R}$ and $\mathbf{v}(\mathbf{s}, t) \in \mathbb{R}^2$ are the temperature and the velocity and where the buoyancy $b(\mathbf{s}, t) \in \mathbb{R}$ satisfies

$$\mathbf{v}(\mathbf{s}, t) = \nabla^\perp \Delta^{-1} b(\mathbf{s}, t).$$

The parameters ρ and ν appearing in (5.1) have the following physical meaning. The Rayleigh number $\nu \in \mathbb{R}_+$ controls the balance between thermal diffusion and the tendency for a packet of fluid to rise due to the buoyancy force. The Prandtl number $\rho \in \mathbb{R}_+$ measures the relative importance of viscosity compared to thermal diffusion. These parameters also controlsthe coupling of the buoyancy evolution with the thermal diffusion process. In particular for $\nu = 0$ or $\rho = 0$, the system is decoupled in the sense that the evolution of buoyancy is independent of the temperature.

At initial time $t = 1$, the fluid in the cell is still and subject to a difference of temperature between the bottom and the top. We set the initial condition to:

$$\begin{aligned} b(\mathbf{s}, 1) &= \pi_b \sin(as_1) \sin(\pi s_2) + \epsilon_b, \\ \tau(\mathbf{s}, 1) &= \pi_\tau \cos(as_1) \sin(\pi s_2) - \pi_{\tau'} \sin(2\pi s_2) + \epsilon_\tau, \end{aligned} \tag{5.2}$$

12

where π_b, π_τ, $\pi_{\tau'}$ and $a \in \mathbb{R}$ are parameters. This initial condition is equal to the solution of the Lorenz attractor [23] up to the additive terms $\epsilon = (\epsilon_b, \epsilon_\tau)^\mathsf{T}$.

We apply a a finite difference scheme on (5.1) to obtain a discrete system of the form of (1.1) with $x_t = \begin{pmatrix} b_t \\ \tau_t \end{pmatrix} \in \mathbb{R}^n$, and $n = 1024$, where b_t's and τ_t's are spatial discretisations of buoyancy and temperature fields at time t. This discretised system constitutes the target model we want to reduce.

5.2. Benchmark Problems. We consider different variations of the problem of ROM construction for unknown $p_\mathbf{X}$. For given trajectories length T and number of observations D, the benchmark problems are variations of the definition of the probabilistic models $p_\mathbf{X}$, $\tilde{p}_\mathbf{X}$ and $p_{\mathbf{Y}|\mathbf{X}}$.

We begin by specifying $p_\mathbf{X}$. Let $\theta_1 = (a, \pi_b, \pi_\tau, \pi_{\tau'})^\mathsf{T} \in \mathbb{R}^4$ and $\theta_2 = (\epsilon_b, \epsilon_\tau)^\mathsf{T} \in \mathbb{R}^{40}$ parameterise the initial condition x_1 using (5.2). Let $\theta_3 = (\rho, \nu)^\mathsf{T} \in \mathbb{R}^2$ parametrise the dynamics (5.1). We recall that we consider here a discretised version of (5.1) - (5.2) of the form of (1.1), which is parameterised by $\theta = (\theta_1, \theta_2, \theta_3)$. We specify the density $p_\mathbf{X}$ through the definition of a probabilistic model for parameter θ and the use of the model (1.1). Note that $p_\mathbf{X}$ is in this configuration a particularisation of (3.6) where the transition kernel is a Dirac measure. We choose for θ a uniform distribution on $\mathcal{T} = (\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)$. The set $\mathcal{T}_1$ is chosen so that initial conditions live at a distance at most of γ from a 10-dimensional subspace of $\mathbb{R}^{1024}$. The set $\mathcal{T}_2$ is a centred ball $\mathcal{B}(0, \gamma)$ of $\mathbb{R}^{40}$ of radius γ. We choose the set of parameters $\mathcal{T}_3$ ruling the dynamics to induce either a dynamical decoupling or a coupling of the buoyancy and the temperature.

The surrogate density $\tilde{p}_\mathbf{X}$ is defined in an analogous manner to $p_\mathbf{X}$. The only difference with the definition of density $p_\mathbf{X}$ is that parameter θ is drawn according to a surrogate uniform distribution on $\tilde{\mathcal{T}} = (\tilde{\mathcal{T}}_1, \mathcal{T}_2, \mathcal{T}_3)$ with $\tilde{\mathcal{T}}_1 \supset \mathcal{T}_1$. More precisely, we fix $\tilde{\mathcal{T}}_1$ so that the $\tilde{x}_1$'s live at a distance at most of γ from a 20-dimensional subspace of $\mathbb{R}^{1024}$.

Let us finally specify the conditional density $p_{\mathbf{Y}|\mathbf{X}}$. It is chosen to be a Gaussian distribution with uncorrelated components so that it admits the factorisation (3.7) where $Y_t|X_t = x_t$ is normally distributed according to $\mathcal{N}(\mathbf{h}x_t, \zeta^2\mathbf{i}_n)$ with $\mathbf{h} \in \mathbb{R}^{n/2 \times n}$ and $\zeta \in \mathbb{R}_+$. Matrix $\mathbf{h}$ is chosen to be a discrete approximation of the convolution by a sinus cardinal kernel so that it represents an ideal low-pass filter degrading the resolution by a factor 2.

Using this configuration, we are able to generate D *i.i.d.* realisations $\{\mathbf{x}^{(i)}, i = 1, \cdots, D\}$ of $p_\mathbf{X}$ by uniformly sampling the set $(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)$ and using model (1.1). Sampling the density $p(y_t|x_t^{(i)})$ then yields to the set of observations $\{\mathbf{y}^{(i)}, i = 1, \cdots, D\}$.

We are now ready to present the benchmark problems. We want to judge of the influence of the following parameters: the trajectories length T, the number of observations D, the noise variance ζ^2, the structure of the distribution on the initial condition (a subspace, *i.e.*, $\mathcal{T}_2 = \{0\}$, or a high-dimensional slice, *i.e.*, $\mathcal{T}_2 = \mathcal{B}(0, \gamma)$ with $\gamma > 0$) and the range of $\mathcal{T}_3$, *i.e.*, of the Prandtl number ρ and the Reynolds number ν, which control the system diffusivity and coupling. We thus consider five different ROM construction problems according to the following setups:

$i)$ $D = 30$, $T = 2$, $\zeta = 0$, $\gamma = 0$, $\rho = 0$ and $\nu = 30$,

13

ii) identical to setup (*i*) but with the noise variance ζ^2 set to induce a peak-to-signal-noise-ratio around 26,

iii) identical to setup (*ii*) but with a uniform distribution on a high-dimensional slice for the initial distribution, *i.e.*, $\gamma = 10^{-3}$,

iv) identical to setup (*iii*) but with longer trajectories ($T = 5$) and fewer observations ($D = 10$),

v) identical to setup (*iv*) but with a Prandtl number of $\rho = 0.03$ and a Reynolds number in the interval $\nu \in [30, 300]$.

The choice of $(T, D) = (2, 30)$ and $T = (5, 10)$ will be justified in the Section 5.4.

5.3. ROMs and Sampling Algorithms. We consider the two examples of reduced models exposed previously and their idealised version, namely:

- *ROM-1*, a low-rank linear approximation, presented in Section 4.2,
- *ROM-2*, a POD-Galerkin approximation, presented in Section 4.1,
- *ideal ROM-1*, approximation $\tilde{x}_t = \mathbf{u}x_{t-1}$ for $t = 2, ..., T$ and $\tilde{x}_1 = x_1$ where $\mathbf{u}$ is the parameter of *ROM-1*, *i.e.*, the orthogonal projection of $x_t's$ on *ROM-1* approximation subspace[4]

$$\mathcal{S}_1(\mathbf{u}) = \mathbb{R}^n \times \underbrace{\operatorname{im}(\mathbf{u}\mathbf{w}_\mathbf{u}) \times \cdots \times \operatorname{im}(\mathbf{u}\mathbf{w}_\mathbf{u})}_{T-1 \text{ times}}. \tag{5.3}$$

- *ideal ROM-2*, approximation $\tilde{x}_t = \mathbf{u}\mathbf{u}^\mathsf{T}x_t$ for $t = 1, ..., T$ where $\mathbf{u}$ is the parameter of *ROM-2*, *i.e.*, the orthogonal projection on the approximation subspace of *ROM-2*

$$\mathcal{S}_2(\mathbf{u}) = \underbrace{\operatorname{im}(\mathbf{u}) \times \cdots \times \operatorname{im}(\mathbf{u})}_{T \text{ times}}. \tag{5.4}$$

According to lower bounds in (4.4) and (4.9), *ROM-1* and *ROM-2* will necessarily be less or as accurate as *ideal ROM-1* and *ideal ROM-2*. The loss in accuracy between the ROMs and their idealised versions corresponds to the error committed by the ROMs outside their approximation subspaces.

We evaluate different sampling algorithms for building a ROM in our uncertain context. The parameter $\mathbf{u}$ of the ROMs are obtained using Proposition 4.1 (resp. Proposition 4.2) for *ROM-2* and *ideal ROM-2* (resp. for *ROM-1* and *ideal ROM-1*) with a definition of matrices $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ specific to the sampling algorithm. In the context of our SMC simulations, we observe that a number of particle $N = 40$ is reasonable. Indeed, increasing this number does not impact significantly the value of the inferred ROM parameter. The sampling strategies are as follows.

---

[4]Note that the distance of $\mathbf{x} \in \mathbb{R}^{n \times T}$ to the subspace $\mathcal{S}_1$ is

$$\inf_{(x_1', ..., x_T') \in \mathcal{S}_1(\mathbf{u})} \sum_{t=1}^{T} \|x_t - x_t'\|_2 = \inf_{(z_1, ..., z_{T-1}) \in \mathbb{R}^{k \times (T-1)}} \sum_{t=2}^{T} \|x_t - \mathbf{u}\mathbf{w}_\mathbf{u}z_{t-1}\|_2,$$

$$= \sum_{t=2}^{T} \|x_t - \mathbf{u}\mathbf{w}_\mathbf{u}\mathbf{w}_\mathbf{u}^\mathsf{T}\mathbf{u}^\dagger x_t\|_2,$$

and that this distance vanishes if $x_t = \mathbf{u}\mathbf{w}_\mathbf{u}\mathbf{w}_\mathbf{u}^\mathsf{T}\mathbf{u}^\dagger x_t = \mathbf{u}x_{t-1}$ for $t = 2, ..., T$.

- *Sampling the target density $p_{\mathbf{X}}$.* ROMs are built relying on samples drawn according to the density $p_{\mathbf{X}}$ (supposed unknown). Matrices $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n \times (T-1)D}$ and $\mathbf{c} \in \mathbb{R}^{n \times TD}$ are set in this case to

$$\mathbf{a} = (\mathbf{x}_{1:T-1}^{(1)} \cdots \mathbf{x}_{1:T-1}^{(D)}),$$
$$\mathbf{b} = (\mathbf{x}_{2:T}^{(1)} \cdots \mathbf{x}_{2:T}^{(D)}),$$
$$\mathbf{c} = (\mathbf{x}_{1:T}^{(1)} \cdots \mathbf{x}_{1:T}^{(D)}).$$

  with $\mathbf{x}_{t_1:t_2}^{(i)} = (x_{t_1}^{(i)} \cdots x_{t_2}^{(i)})$ where $x_t^{(i)}$ denotes the state generating observation $y_t^{(i)}$.

- *Sampling the proposed data-enhanced surrogate density $\hat{p}_{\mathbf{X}}$.* ROMs are built relying on a refined version of the surrogate $\tilde{p}_{\mathbf{X}}$ defined in (2.5). This density is approximated using MC and SMC techniques, as presented in Sections 3.1 and 3.2. More precisely, SMC samples are obtained by sequential importance sampling [12] with η_1 as the proposal distribution[5]. Matrices $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are defined in Section 4.1 and Section 4.2.

- *Sampling the initial surrogate density $\tilde{p}_{\mathbf{X}}$.* ROMs are built relying on samples drawn according to $\tilde{p}_{\mathbf{X}}$, *i.e.*, ignoring observations. Matrices $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n \times (T-1)DN}$ and $\mathbf{c} \in \mathbb{R}^{n \times TDN}$ are set in this case to

$$\mathbf{a} = (\boldsymbol{\xi}_{1:T-1}^{(1,1)} \cdots \boldsymbol{\xi}_{1:T-1}^{(D,N)}),$$
$$\mathbf{b} = (\boldsymbol{\xi}_{2:T}^{(1,1)} \cdots \boldsymbol{\xi}_{2:T}^{(D,N)}),$$
$$\mathbf{c} = (\boldsymbol{\xi}_{1:T}^{(1,1)} \cdots \boldsymbol{\xi}_{1:T}^{(D,N)}).$$

- *Point estimates.* ROMs are built relying on MMSE point estimates. Matrices $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{n \times (T-1)D}$ and $\mathbf{c} \in \mathbb{R}^{n \times TD}$ are set in this case to

$$\mathbf{a} = (\hat{\boldsymbol{\xi}}_{1:T-1}^{(1)} \cdots \hat{\boldsymbol{\xi}}_{1:T-1}^{(D)}),$$
$$\mathbf{b} = (\hat{\boldsymbol{\xi}}_{2:T}^{(1)} \cdots \hat{\boldsymbol{\xi}}_{2:T}^{(D)}),$$
$$\mathbf{c} = (\hat{\boldsymbol{\xi}}_{1:T}^{(1)} \cdots \hat{\boldsymbol{\xi}}_{1:T}^{(D)}).$$

  with point estimates $\hat{\boldsymbol{\xi}}_{t_1:t_2}^{(i)} = (\hat{\xi}_{t_1}^{(i)} \cdots \hat{\xi}_{t_2}^{(i)})$ given for $i = 1, ..., D$ by

$$\hat{\xi}_t^{(i)} = \sum_{j=1}^{N} w^{(i,j)} \xi_t^{(i,j)},$$

  for $t = 1, ..., T$.

---

[5]In the case of a small noise variance ζ^2, we slightly modify the density η_1 in (3.6) to avoid spreading samples at initial time too far from observations. We force the initial condition samples to concentrate around the hyperplane $\mathbf{h}^{\dagger} y_1$ by substituting sample $\tilde{x}_1$ by the orthogonal projection of

$$(\mathbf{i} - \mathbf{h}^{\dagger}\mathbf{h})\tilde{x}_1 + \mathbf{h}^{\dagger}(y_1 + \mathrm{w}),$$

on the subspace embedding $\tilde{\mathcal{T}}_1$. We recall that $\mathrm{w} \in \mathbb{R}^n$ is a realisation of a zero-mean and uncorrelated Gaussian random variable of variance ζ^2.

5.4. Results and Discussion. Figure 5.1 and Figure 5.2 present the performances of the different sampling algorithms for building *ROM-1*, *ideal ROM-1*, *ROM-2* and *ideal ROM-2*. Figure 5.1 and Figure 5.2 treat respectively problem setups i) to iii) and setups iv) to v). The plots display the evolution of the average of the error norm $\| \mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u}) \|_F$ over the D trajectories which have generated the observations, with respect to the ROM dimension k.

We have set $(T, D) = (2, 30)$ in Figure 5.1 and $(T, D) = (5, 10)$ in Figure 5.2 to make the error norm comparable in the two figures for a dimension $k < 50$. We mention that setting $T = 2$ (results displayed in Figure 5.1), we obtain that $\|x_2 - \mathbf{u}x_1\|_2 = \|\mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u})\|_F$ for *ROM-1*, implying that the distance to subspace $\mathcal{S}_1$ defined in (5.3) will necessarily be equal to the ROM error norm. *ROM-1* will thus be in this case equivalent to the *ideal ROM-1* for any sampling algorithm. This setting simplifies the understanding and comparison of the different algorithms, as detailed below. Besides, for $T > 2$ (Figure 5.2) we can expect a difference of performance between *ROM-1* and *ideal ROM-1* , and in particular, the more non-linear the f_t's in the dynamical model (1.1), the more the difference of performances. Nevertheless, in our experiments, we will observe that this difference remains reasonable, showing that inequalities (4.4) and (4.9) are almost equalities ($c \simeq 1$), and in consequence this will provide an experimental justification of the strategy of bounding (rather than evaluating precisely) the objective function. For legibility purposes, we will display the performances of *ideal ROM-1* and *ideal ROM-2* only for the algorithm sampling the target density.

We observe in our experiments that, for any problem setup and any ROM, sampling the *initial surrogate* $\tilde{p}_{\mathbf{X}}$ leads to the poorest peformance. The ROMs built from *point estimates* yields to a slight enhancement. Moreover, except in the case of setup v) and its strong non-linearities, the *proposed data-enhanced surrogate* leads to the best approximation accuracy. In what follows, we discuss in details this general analysis.

Let us begin by some comments on the behaviour of the algorithm sampling the *target density* $p_{\mathbf{X}}$ for construction of *ROM-1*, *ROM-2* and their idealised versions. As expected for setups i) and ii) where $T = 2$, the error of *ROM-1* obtained vanishes[6] for a subspace of dimensionality above the initial condition dimension, *i.e.*, for $k \geq 10$. *ROM-2* is less accurate because the error vanishes in the best case for $k \geq 20$ in the case of a projection of the trajectory on subspace $\mathcal{S}_2$ defined in (5.4) (while it can vanish for $k \geq 10$ for a projection on $\mathcal{S}_1$). For setup iii), the initial conditions are not embedded anymore in a 10-dimensional subspace, but in a slice of thickness γ and of dimensionality $10 + 40 = 50$ around this subspace. As expected, we verify that the error is equal to zero for $k \geq \min(50, D) = 30$ (resp. $k \geq \min(100, 2D) = 60$) in the case of *ROM-1* (resp. *ROM-2*). In practice, we note that the error vanishes even around $k \geq 25$ (resp. $k \geq 50$). The last two settings, *i.e.*, setups iv) and v), imply longer sequences $T = 5$. In this situation, the trajectories could be approximated in a worst-case scenario (scenario of linear independence of states at different times of the trajectory) using a subspace of dimension $(T - 1)D = 40$ for *ROM-1* (resp. of dimension $TD = 50$ for *ROM-2*). We observe however that this pessimistic scenario does not occur in practice, especially if non-linearities are moderated (setup iv)). Indeed,

---

[6]The computation of an eigendecomposition or the use of SVD induces a machine precision around $1e - 5$ for trajectories computed with *ROM-2* or *ROM-1*.

between 30 and 40 components (resp. between 35 and 40 components) are sufficient to obtain a zero approximation error for *ROM-1* (resp. *ROM-2*).

We now compare the performances of algorithms which do not rely on the knowledge of $p_{\mathbf{X}}$, namely the algorithms sampling the *initial surrogate*, the *proposed data-enhanced surrogate* and the algorithm based on *point estimates*. First, for setups *i*) and *ii*), since the *initial surrogate* and the *proposed data-enhanced surrogate* are defined for an initial condition of dimensionality twice bigger, we verify that *ROM-1* (resp. *ROM-2*) built sampling the *initial surrogate* or the *proposed data-enhanced surrogate* cancel the error for $k \geq 20$ (resp. $k \geq 40$). We can observe that a ROM built with *point estimates* is slightly less accurate independently of the dimension k, and in particular for $k \geq 20$ by a factor 5. This error saturation effect is the consequence of the non-observation of certain components of the trajectories due to the non-trivial kernel of operator $\mathbf{h}$. In other words, this non-reducible error is due to the fact the method ignores that point estimates have an infinite variance along directions of the subspace defined by this kernel.

Second, we observe a moderate loss of accuracy for setup *ii*), which attests that the data-based algorithms seem to be robust to moderate observation noise.

Third, the target distribution defined in setup *iii*) (taking the form of a high-dimensional slice) turns out to slightly increase the error above $k \geq 20$ by a factor 5 (resp. 2.5) for sampling of the *initial surrogate* or of the *proposed data-enhanced surrogate* (resp. for *point estimates*). This result can be interpreted as the fact that the algorithms are robust to the reduction of trajectories which do not necessarily belong to a subspace, but which are moderately distant from it.

Fourth, we observe in setup *iv*), that sampling the *proposed data-enhanced surrogate* or using *point estimates* induces only a slight deterioration of performances when compared to the algorithm sampling the *target density*. On the contrary, while being reasonable for *ROM-2*, sampling the *initial surrogate* yields dramatic results for *ROM-1*. This effect can be easily understood: in the case of a low-rank linear approximation, increasing the dimension k is not sufficient to obtain a gain in performance; in particular, over-estimating the eigenvalues of matrix $\mathbf{u}^{\star}$ induces by construction of *ROM-1* an exponential increase of the approximation error.

Finally, for setup *v*), this unstable behaviour affects *ROM-1* for any of the algorithms. On the contrary, *ROM-2*, *i.e.*, POD-Galerkin approximation, seems to be nearly insensitive to the presence of strong non-linearities in (1.1).

6. Conclusions. We have proposed a general framework for the construction of ROMs when the set of trajectories of a dynamical system is imperfectly known. This work assumes that we have the following two sources of information at our disposal: 1) samples of an initial surrogate density characterising the set of trajectories of the system of interest; 2) a set of incomplete observations on the target trajectories obeying a known conditional model.

The ROM construction consists in the minimisation of the expectation of the norm of the error between the true and reduced trajectories. The expectation relies on a data-enhanced surrogate density obtained in a Bayesian setting combining the initial surrogate density and the conditional observation model. We show, that under mild conditions and in term of the Kullback-Leibler distance, the proposed data-enhanced surrogate is a better approximation of the target density than the initial surrogate.

We stress the need of approximations to efficiently solve this problem and propose in this context tractable solvers. In particular, we use MC and SMC techniques to
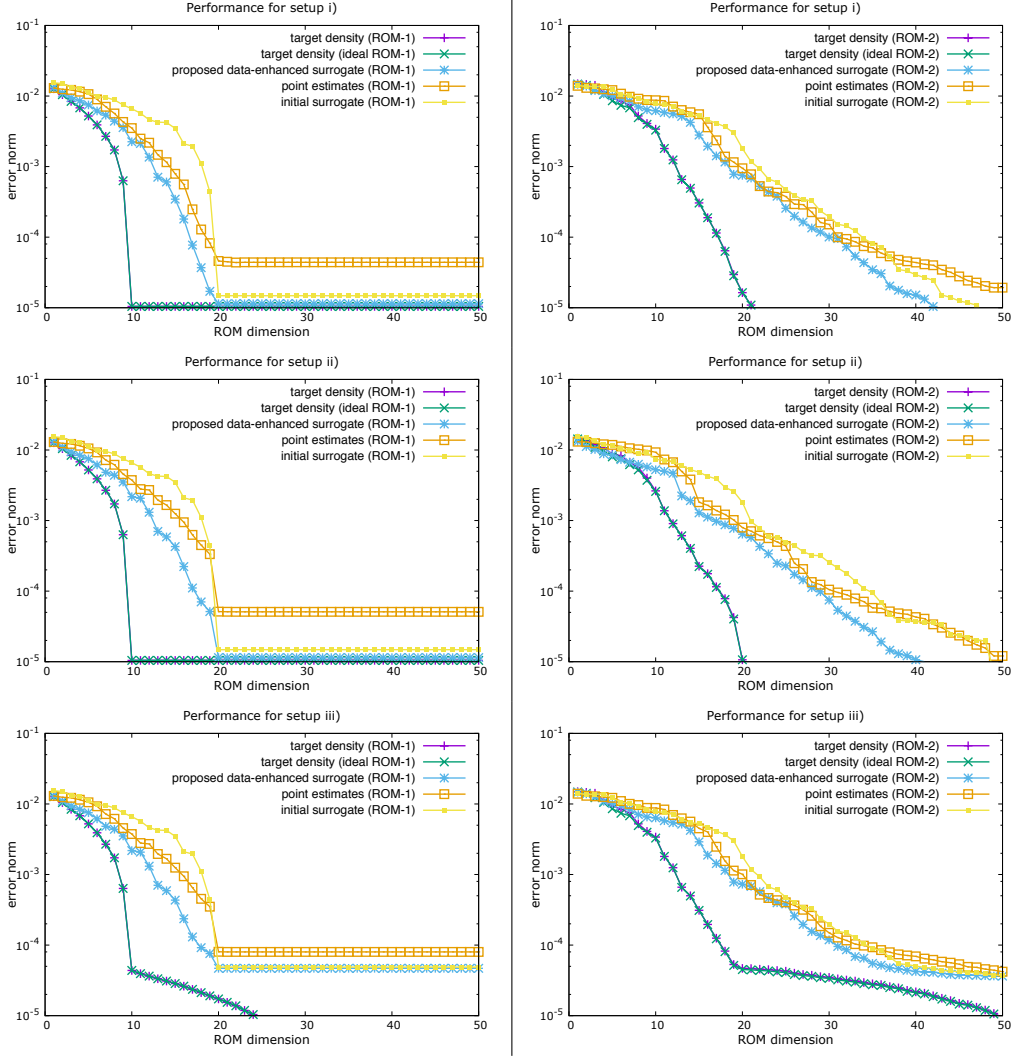
Fig. 5.1. *Algorithms performances for construction of ROM-1 and ideal ROM-1 (on the left) and ROM-2 and ideal ROM-2 (on the right) for setup i) (above), ii) (middle) and iii) (below). See details in Section 5.2 and 5.3*

characterise our data-enhanced surrogate and propose implementations based on the minimisation of a bound on the objective function. We illustrate how the proposed methodology particularises to two different families of ROMs. We show that these particularisations can be seen as generalisations, in a probabilistic framework, of well-known ROM techniques, namely POD and low-rank DMD. We also show that they can be seen as generalisations of standard ROM constructions based on point estimates.

A numerical evaluation, led in the context of the reduction of a geophysical model, reveals that the proposed methodology enhances state of the art.
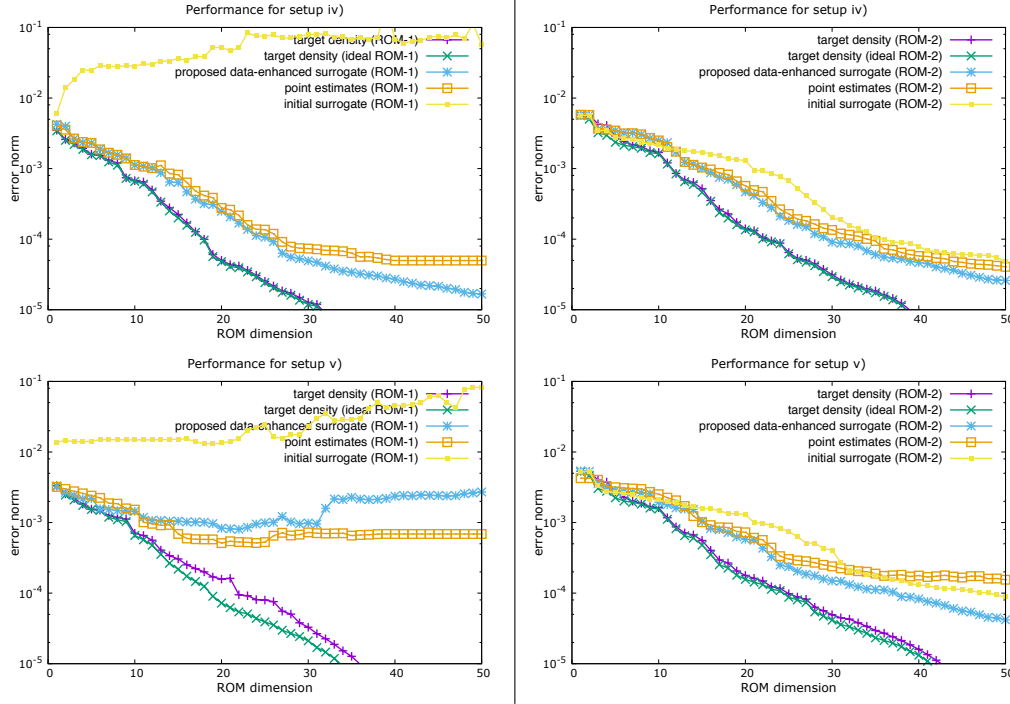
FIG. 5.2. *Algorithms performances for construction of ROM-1 and ideal ROM-1 (on the left) and ROM-2 and ideal ROM-2 (on the right) for setup iv) (above) and v) (below). See details in Section 5.2 and 5.3*

Appendix A. Error Bounds for Low-Rank Linear Approximation. We hereafter show that the error norm for a low-rank linear approximation of (1.1) can be bounded as presented in Section 4.2.

To obtain the lower bound in (4.9), we notice that according to (5.3), we have $\tilde{x}_t \in \mathrm{im}(\mathbf{u}\mathbf{w_u})$ for $t \geq 2$, so that each term contributing to the error norm $\| \mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u}) \|_F^2 = \sum_{t=2}^{T} \|x_t - \tilde{x}_t(\mathbf{u})\|_2^2$ can be decomposed into two orthogonal components

$$x_t - \tilde{x}_t(\mathbf{u}) = \underbrace{\mathbf{u}\mathbf{w_u}\mathbf{w_u^{\mathsf{T}}}\mathbf{u}^{\dagger}x_t - \tilde{x}_t(\mathbf{u})}_{\in\,\mathrm{im}(\mathbf{u}\mathbf{w_u})} + \underbrace{x_t - \mathbf{u}\mathbf{w_u}\mathbf{w_u^{\mathsf{T}}}\mathbf{u}^{\dagger}x_t}_{\in\,\mathrm{im}(\mathbf{u}\mathbf{w_u})^{\perp}}.$$

This implies that

$$\sum_{t=2}^{T} \|x_t - \mathbf{u}\mathbf{w_u}\mathbf{w_u^{\mathsf{T}}}\mathbf{u}^{\dagger}x_t\|_2^2 = \sum_{t=2}^{T} \|x_t - \mathbf{u}x_{t-1}\|_2^2 \leq \| \mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u}) \|_F^2 .$$

The following result shows that the upper bound in (4.9) can be obtained under mild assumptions.

LEMMA A.1. *Assume there exists λ such that $\|\mathbf{u}\|_{2,2} < \lambda$. Then*

$$\| \mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u}) \|_F^2 \leq c \sum_{t=2}^{T} \|x_t - \mathbf{u}x_{t-1}\|_2^2,$$

with $c = \max_{t \in \{2,...,T\}}\{\sum_{k=t}^{T}(1 + 4(k-1))\lambda^{2(k-t)}\}$.

Proof. Since $\tilde{x}_t = \mathbf{u}\tilde{x}_{t-1}$, we have

$$\| \mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u}) \|_F^2 = \sum_{t=2}^{T} \|x_t - \mathbf{u}^{t-1}x_1\|_2^2.$$

Using the triangular inequality and the definition of the induced ℓ_2-norm, each term contribution in this sum can be bounded as follows

$$\|x_t - \mathbf{u}^{t-1}x_1\|_2 \leq \sum_{\ell=2}^{t} \|\mathbf{u}^{\ell-1}x_{t-\ell+1} - \mathbf{u}^{\ell-2}x_{t-\ell+2}\|_2,$$

$$\leq \sum_{\ell=2}^{t} \|\mathbf{u}^{\ell-2}\|_{2,2}\|\mathbf{u}x_{t-\ell+1} - x_{t-\ell+2}\|_2,$$

$$= \sum_{k=2}^{t} \|\mathbf{u}^{t-k}\|_{2,2}\|\mathbf{u}x_{k-1} - x_k\|_2.$$

Therefore, expanding the square of this sum, we obtain

$$\|x_t - \mathbf{u}^{t-1}x_1\|_2^2 \leq \sum_{k=2}^{t} \|\mathbf{u}^{t-k}\|_{2,2}^2\|\mathbf{u}x_{k-1} - x_k\|_2^2$$

$$+ 2 \sum_{i,j=2|i\neq j}^{t} \|\mathbf{u}^{t-i}\|_{2,2}\|\mathbf{u}x_{i-1} - x_i\|_2\|\mathbf{u}^{t-j}\|_{2,2}\|\mathbf{u}x_{j-1} - x_j\|_2,$$

$$\leq \sum_{k=2}^{t} \|\mathbf{u}^{t-k}\|_{2,2}^2\|\mathbf{u}x_{k-1} - x_k\|_2^2$$

$$+ 2 \sum_{i,j=2|i\neq j}^{t} \max\{\|\mathbf{u}^{t-i}\|_{2,2}^2\|\mathbf{u}x_{i-1} - x_i\|_2^2, \|\mathbf{u}^{t-j}\|_{2,2}^2\|\mathbf{u}x_{j-1} - x_j\|_2^2\},$$

$$\leq \sum_{k=2}^{t} \|\mathbf{u}^{t-k}\|_{2,2}^2\|\mathbf{u}x_{k-1} - x_k\|_2^2$$

$$+ 2 \sum_{i,j=2|i\neq j}^{t} \|\mathbf{u}^{t-i}\|_{2,2}^2\|\mathbf{u}x_{i-1} - x_i\|_2^2 + \|\mathbf{u}^{t-j}\|_{2,2}^2\|\mathbf{u}x_{j-1} - x_j\|_2^2,$$

$$\leq \sum_{k=2}^{t} (1 + 4(t-1))\|\mathbf{u}^{t-k}\|_{2,2}^2\|\mathbf{u}x_{k-1} - x_k\|_2^2.$$

In consequence, we conclude remarking that

$$\| \mathbf{x} - \tilde{\mathbf{x}}(\mathbf{u}) \|_F^2 \leq \sum_{t=2}^{T}\sum_{k=2}^{t} (1 + 4(t-1))\|\mathbf{u}^{t-k}\|_{2,2}^2\|\mathbf{u}x_{k-1} - x_k\|_2^2,$$

$$= \sum_{t=2}^{T}\sum_{k=t}^{T} (1 + 4(k-1))\|\mathbf{u}^{k-t}\|_{2,2}^2\|\mathbf{u}x_{t-1} - x_t\|_2^2,$$

$$\leq c \| \mathbf{x}_{2:T} - \mathbf{u}\mathbf{x}_{1:T-1} \|_F^2,$$

20

with $c = \max_{t \in \{2,...,T\}} \{\sum_{k=t}^{T}(1+4(k-1))\|\mathbf{u}^{k-t}\|_{2,2}^2\}$ and where the equality has been obtained by inverting the two sums. $\square$

REFERENCES

[1] Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D., Stuart, A.M.: Importance Sampling: Computational Complexity and Intrinsic Dimension. ArXiv e-prints (2015)

[2] Antoulas, A.C.: An overview of approximation methods for large-scale dynamical systems. Annual Reviews in Control **29**(2), 181–190 (2005)

[3] Barrault, M., Maday, Y., Nguyen, N.C., Patera, A.T.: An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. Comptes Rendus Mathematique **339**(9), 667 – 672 (2004)

[4] Chandrasekhar, S.: Hydrodynamic and hydromagnetic stability. Courier Corporation (2013)

[5] Chaturantabut, S., Sorensen, D.C.: Nonlinear model reduction via discrete empirical interpolation. SIAM Journal on Scientific Computing **32**(5), 2737–2764 (2010)

[6] Chen, K.K., Tu, J.H., Rowley, C.W.: Variants of dynamic mode decomposition: boundary condition, koopman, and fourier analyses. Journal of nonlinear science **22**(6), 887–915 (2012)

[7] Ciarlet, P.: The Finite Element Method for Elliptic Problems. Society for Industrial and Applied Mathematics (2002)

[8] Cohen, A., Devore, R.: Approximation of high-dimensional parametric PDEs. ArXiv e-prints (2015)

[9] Crisan, D., Doucet, A.: A survey of convergence results on particle filtering methods for practitioners. IEEE Transactions on signal processing **50**(3), 736–746 (2002)

[10] Cui, T., Martin, J., Marzouk, Y.M., Solonen, A., Spantini, A.: Likelihood-informed dimension reduction for nonlinear inverse problems. Inverse Problems **30**(11), 114015 (2014)

[11] Cui, T., Marzouk, Y.M., Willcox, K.E.: Data-driven model reduction for the Bayesian solution of inverse problems. International Journal for Numerical Methods in Engineering **102**, 966–990 (2015)

[12] Doucet, A., Godsill, S., Andrieu, C.: On sequential monte carlo sampling methods for bayesian filtering. Statistics and computing **10**(3), 197–208 (2000)

[13] Everson, R., Sirovich, L.: Karhunen-loève procedure for gappy data. J. Opt. Soc. Am. A **12**(8), 1657–1664 (1995)

[14] Fink, J.P., Rheinboldt, W.C.: On the error behavior of the reduced basis technique for nonlinear finite element approximations. ZAMM - Journal of Applied Mathematics and Mechanics **63**(1), 21–28 (1983)

[15] Gunes, H., Sirisup, S., Karniadakis, G.E.: Gappy data: To krig or not to krig? Journal of Computational Physics **212**(1), 358–382 (2006)

[16] Hasselmann, K.: Pips and pops: The reduction of complex dynamical systems using principal interaction and oscillation patterns. Journal of Geophysical Research: Atmospheres **93**(D9), 11,015–11,021 (1988)

[17] Héas, P., Herzet, C.: Low-rank Approximation and Dynamic Mode Decomposition. ArXiv e-prints (2016)

[18] Herzet, C., Drémeau, A., Héas, P.: Model Reduction from Partial Observations. ArXiv e-prints (2016)

[19] Holmes, P., Lumley, J.L., Berkooz, G.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge University Press (1996). Cambridge Books Online

[20] Jolliffe, I.: Principal Component Analysis. Springer Series in Statistics. Springer (2002)

[21] Jovanovic, M., Schmid, P., Nichols, J.: Low-rank and sparse dynamic mode decomposition. Center for Turbulence Research Annual Research Briefs pp. 139–152 (2012)

[22] Kwasniok, F.: The reduction of complex dynamical systems using principal interaction patterns. Phys. D **92**(1-2), 28–60 (1996)

[23] Lorenz, E.N.: Deterministic Nonperiodic Flow. Journal of Atmospheric Sciences **20**, 130–148 (1963)

[24] Maday, Y., Patera, A.T., Penn, J.D., Yano, M.: A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. International Journal for Numerical Methods in Engineering **102**(5), 933–965 (2015)

[25] Peherstorfer, B., Willcox, K.: Data-driven operator inference for nonintrusive projection-based model reduction. Computer Methods in Applied Mechanics and Engineering **306**, 196–215 (2016)

[26] Quarteroni, A., Rozza, G., Manzoni, A.: Certified reduced basis approximation for

parametrized partial differential equations and applications. Journal of Mathematics in Industry **1**(1), 1–49 (2011)

[27] Raviart, P., Thomas, J.: Introduction à l'analyse numérique des équations aux dérivées partielles. Collection Mathématiques appliquées pour la maîtrise. Masson (1983)

[28] Sirovich, L.: Turbulence and the dynamics of coherent structures. Quarterly of Applied Mathematics **45**, 561–571 (1987)

[29] Spantini, A., Solonen, A., Cui, T., Martin, J., Tenorio, L., Marzouk, Y.: Optimal low-rank approximations of Bayesian linear inverse problems. ArXiv e-prints (2014)