Stochastic Data-Driven Variational Multiscale Reduced Order Models

Fei Lu[1], Changhong Mou[2], Honghu Liu[3], and Traian Iliescu[4]

[1]Department of Mathematics, Johns Hopkins University, Baltimore, MD 21218, USA (feilu@math.jhu.edu)
[2]Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53705, USA (cmou3@wisc.edu)
[3]Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA (hhliu@vt.edu)
[4]Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA (iliescu@vt.edu)

September 6, 2022

Abstract

Trajectory-wise data-driven reduced order models (ROMs) tend to be sensitive to training data, and thus lack robustness. We propose to construct a robust stochastic ROM closure (S-ROM) from data consisting of multiple trajectories from random initial conditions. The S-ROM is a low-dimensional time series model for the coefficients of the dominating proper orthogonal decomposition (POD) modes inferred from data. Thus, it achieves reduction both space and time, leading to simulations orders of magnitude faster than the full order model. We show that both the estimated POD modes and parameters in the S-ROM converge when the number of trajectories increases. Thus, the S-ROM is robust when the training data size increases. We demonstrate the S-ROM on a 1D Burgers equation with a viscosity $\nu = 0.002$ and with random initial conditions. The numerical results verify the convergence. Furthermore, the S-ROM makes accurate trajectory-wise predictions from new initial conditions and with a prediction time far beyond the training range, and it quantifies the spread of uncertainties due to the unresolved scales.

Contents

1 Introduction

Data-driven methods, combined with structures in physical systems, provide effective tools for the construction of closure reduced order models (ROMs) [43, 46, 47, 49] that bring down computational cost by orders of magnitude from the full order models (FOM). The data are a single or multiple solutions of a FOM

$$\dot{\boldsymbol{u}} = \boldsymbol{f}(\boldsymbol{u}), \quad \Rightarrow \quad \{\boldsymbol{u}(t)^{(m)}, t \in [0, T]\}_{m=1}^{M}, \tag{1.1}$$

where $\boldsymbol{u}$ can be either a high-dimensional state vector for a system of stochastic/ordinary differential equations or a solution to a partial differential equation. The closure ROMs are low-dimensional models that are constructed from data. They approximate the FOM solution $\boldsymbol{u}$ by $\boldsymbol{u}(t, x) \approx \sum_{i=1}^{r} a_i(t)\boldsymbol{\varphi}_i(x)$, where both the basis functions $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_r\}$ and the ROM closure for the coefficient vector $\boldsymbol{a} = (a_1, \ldots, a_r)$:

$$\dot{\boldsymbol{a}} = \boldsymbol{F}(\boldsymbol{a}) + \text{Closure}(\boldsymbol{a}), \tag{1.2}$$

are estimated from data. Here the term $\boldsymbol{F}$ often comes from a truncated representation of $\boldsymbol{f}$ in the FOM, and the term Closure($\boldsymbol{a}$) accounts for the effects of the unresolved modes $\{\boldsymbol{\varphi}_{r+1}, \ldots, \boldsymbol{\varphi}_n, \ldots\}$.

The closure problem is ubiquitous in dimension-reduced modeling of high- or infinite-dimensional nonlinear dynamics. The task is to account for the unresolved scales crucial for the dynamics, such as the sub-grid scales in the classical numerical discretization of turbulent flows (e.g., finite element or finite volume methods). There are hundreds of closure models in classical CFD, e.g., large eddy simulation (LES) [2–4, 20]. These LES closure models are built around the physical insight stemming from Kolmogorov's statistical theory of turbulence. Unfortunately, this physical insight is generally posed in the Fourier series setting, thus not directly applicable to ROMs whose basis functions are estimated from data. Therefore, these ROM closure models are deprived of many tools that represent the core of most LES closure models.

The combination of physical insights and data provides a promising direction for ROM closure modeling. It is poised to extend the reach of physical insights, reduce the size of training data, and produce robust predictive closure models. This is an extremely active research area and many important research have demonstrated the power of this combination (see e.g., [1, 5, 6, 11, 24, 26, 33, 48, 52, 54, 57–59] for a biased sample). In particular, deterministic ROM closure models have been constructed with proper orthogonal decomposition (POD) basis functions, Galerkin projections and least squares closures (see e.g., [5, 49]). The resulting ROM closures accurately fit the data, but they tend to have limited generalizability beyond the training data set and do not represent the uncertainty due to the unresolved scales. Stochastic ROM closure models have been constructed from multi-trajectory data via statistical inference [16, 40, 41, 43]. While they quantify the spread of uncertainty, they are only for pre-selected large scale variables.

We propose to construct data-driven stochastic ROM (S-ROM) closure for deterministic systems from multi-trajectory data with random initial conditions, so that the S-ROM can make accurate predictions beyond training datasets and quantify the spread of uncertainty. The construction consists of two steps. First, we identify the dominating modes (basis functions) from the data by POD. Then, we infer a stochastic closure model for the evolution of the coefficients of the dominating

modes. The S-ROM approximates the discrete-time flow map of these modes, and its parameter is estimated by maximizing the likelihood of the data. It is a natural probabilistic extension of the variational approach from deterministic ROMs [49, 50] to stochastic models, and an extension of the stochastic closure models in [16, 41, 43] from pre-selected basis to data-deduced basis.

This study focuses on systems with quadratic nonlinearity and considers S-ROMs depending linearly on the parameters. As a result, we can efficiently construct the S-ROM by least squares (with regularization when the normal matrix is ill-conditioned). Furthermore, we prove that the estimators of the POD basis and the parameters converge at the rate $M^{-1/2}$, where M denotes the number of independent trajectories.

We demonstrate this new approach on a 1D Burgers equation with random initial conditions. Numerical results verify the convergence of the estimators of the POD basis and the parameters of the S-ROM. In contrast, the single-trajectory estimator varies largely between trajectories. Furthermore, the S-ROM makes accurate predictions from new random initial conditions for a longer time than the training interval, and it quantifies the spread of the uncertainty. In addition, the S-ROM is adaptive to time-step size, and it can tolerate a relatively large time-step size because it accounts for the discretization error by utilizing the information from data.

This study provides a parametric inference examination of the success of the combination of physical insights and data. As a data-driven approach, this study makes a new contribution to the recent advances on utilizing statistical learning and machine learning tools to construct ROM closure models, ranging from Mori-Zwanzig formalism [14,17,55,60], non-Markovian modeling [16,36,37,40], filtering and model error quantification [23,28,46], and conditional Gaussian models [7,9], to machine learning methods [8, 21, 29, 38, 45, 51, 53], to name just a few. Furthermore, it can be viewed as an approximation of the flow operator of the full order model, in the context of operator inference by machine learning methods [35].

Our main contributions are:

1. We examine a parametric inference approach, which combines physical insights with data, to construct stochastic ROM closures from multi-trajectory data. The parameters are efficiently estimated by regression with regularization. We prove and numerically verify the convergence of the POD modes and the parameter estimators as the number of data trajectories increases.

2. We show that the training trajectories can be short trajectories for autonomous systems, and the constructed S-ROM can make predictions far beyond the training time interval. Furthermore, the S-ROM can represent the spread of the uncertainty due to unresolved scales, making it suitable for data assimilation tasks.

3. We demonstrate that the S-ROM is adaptive to both resolution and time step size. Thus, it leads to reduction in both space and time, achieving efficient simulations that can be orders of magnitudes faster than the FOM.

The rest of the paper is organized as follows: we introduce the method for the construction of the S-ROM in Section 2. Section 3 investigates S-ROM in the context of a Burgers equation, systematically examining the convergence of estimators and the performance of the S-ROM in trajectory-wise prediction and ensemble prediction. Conclusions and final remarks are presented in Section 4.

2   Stochastic DD-VMS-ROM

In this section, we introduce the method for the construction of a stochastic data-driven variational multiscale reduced order model (DD-VMS-ROM). It leads to a robust stochastic ROM (S-ROM)

that quantifies the uncertainty from the unresolved scales and the randomness from the initial condition. We construct the S-ROM from multiple-trajectory data in two stages: (1) construct the basis functions (orthogonal modes) from the data by proper orthogonal decomposition (POD); (2) infer the parameters in the S-ROM, which is a time series model describing the dynamics of the coefficients of the basis functions, by maximizing the likelihood of the multi-trajectory data. Thus, the S-ROM builds on data-deduced basis functions, a parametric structure derived from the full model, and a parametric inference. It is a natural probabilistic extension of the variational approach from deterministic ROMs [49, 50] to stochastic models, and an extension of the stochastic closure models in [16, 41, 43] from pre-selected spectral basis to data-deduced basis.

2.1 Variational multiscale methods and flow map approximation

Variational multiscale methods. The variational multiscale (VMS) methods [30–32] increase the accuracy of classical Galerkin approximation by introducing hierarchical spaces and coupling terms to model the effects of the unresolved scales. To illustrate the VMS methodology, suppose that the FOM in (1.1) is a nonlinear PDE and consider its variational (weak) form

$$\langle \dot{\boldsymbol{u}}, \boldsymbol{v} \rangle = \langle \boldsymbol{f}(\boldsymbol{u}), \mathbf{v} \rangle, \quad \boldsymbol{v} \in \boldsymbol{X}, \tag{2.1}$$

where $\boldsymbol{X}$ is an appropriate Hilbert space and $\langle \cdot, \cdot \rangle$ denotes its inner product. The VMS projects the full model onto subspaces $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \ldots$, which build a sequence of hierarchical spaces of increasing resolutions $\boldsymbol{X}_1, \boldsymbol{X}_1 \oplus \boldsymbol{X}_2, \boldsymbol{X}_1 \oplus \boldsymbol{X}_2 \oplus \boldsymbol{X}_3, \ldots$. The projection leads to separate equations on each space, and the goal is to solve for the components of $\boldsymbol{u}$ that live in the space $\boldsymbol{X}_1$ (which represents the resolved scales), denoted by $\boldsymbol{u}_1$, whose equation is a low-dimensional system

$$\langle \dot{\boldsymbol{u}}_1, \boldsymbol{v}_1 \rangle = \langle \dot{\boldsymbol{u}}, \boldsymbol{v}_1 \rangle = \langle \boldsymbol{f}(\boldsymbol{u}), \boldsymbol{v}_1 \rangle = \langle \boldsymbol{f}(\boldsymbol{u}_1), \boldsymbol{v}_1 \rangle + \langle \boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{f}(\boldsymbol{u}_1), \boldsymbol{v}_1 \rangle, \quad \forall \boldsymbol{v}_1 \in \boldsymbol{X}_1. \tag{2.2}$$

The above system is not closed since, due to the nonlinearity of $\boldsymbol{f}(\boldsymbol{u})$, the term $\langle \boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{f}(\boldsymbol{u}_1), \boldsymbol{v}_1 \rangle$ involves components that do not live in $\boldsymbol{X}_1$. Thus, a VMS closure model of $\boldsymbol{u}_1$ aims to approximate $\langle \boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{f}(\boldsymbol{u}_1), \boldsymbol{v}_1 \rangle$ in terms of $\boldsymbol{u}_1$.

To construct the VMS closure, we identify a closed dynamical system to describe the dynamics of the coefficients of $\boldsymbol{u}_1$ in an orthonormal basis of $\boldsymbol{X}_1$. More precisely, suppose that $\boldsymbol{X}$ has a complete orthonormal basis $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_r, \ldots\}$ and $\boldsymbol{X}_1 = \text{span}\{\boldsymbol{\varphi}_i\}_{i=1}^r$, and write the solution as $\boldsymbol{u}(t, x) = \sum_{i=1}^\infty a_i(t) \boldsymbol{\varphi}_i(x)$ and $\boldsymbol{u}_1(t, x) = \sum_{i=1}^r a_i(t) \boldsymbol{\varphi}_i(x)$. Denote $\boldsymbol{a} = (a_1, a_2, \ldots, a_r)$ and $\boldsymbol{b} = (a_{r+1}, \ldots, a_{r+n}, \ldots)$, where $\boldsymbol{a}_i = \langle \dot{\boldsymbol{u}}, \boldsymbol{\varphi}_i \rangle$ for $i \geqslant 1$. Then, Eq. (2.2) is equivalent to the differential system

$$\dot{\boldsymbol{a}} = \boldsymbol{F}(\boldsymbol{a}) + \boldsymbol{G}(\boldsymbol{a}, \boldsymbol{b}), \tag{2.3}$$

which is a differential system on $\mathbb{R}^r$ with

$$\boldsymbol{F}(\boldsymbol{a})_i = \langle \boldsymbol{f}(\boldsymbol{u}_1), \boldsymbol{\varphi}_i \rangle, \qquad \boldsymbol{G}(\boldsymbol{a}, \boldsymbol{b})_i = \langle \boldsymbol{f}(\boldsymbol{u}), \boldsymbol{\varphi}_i \rangle - \langle \boldsymbol{f}(\boldsymbol{u}_1), \boldsymbol{\varphi}_i \rangle, \qquad i = 1, \ldots, r. \tag{2.4}$$

Thus, the task of a closure problem is to approximate the term $\boldsymbol{G}(\boldsymbol{a}, \boldsymbol{b})$ by either a function of $\boldsymbol{a}$ or a functional of the path of $\boldsymbol{a}$. Similarly, a hierarchical multiscale closure model describes the dynamics of the coefficients of the $\boldsymbol{u}$ in hierarchical spaces.

There are two main challenges: (1) the hierarchical spaces, i.e., their basis functions, can be difficult to construct in numerical methods with non-orthogonal basis functions (e.g., finite element or finite volume methods); and (2) numerical closure models, which account for the effects of the unresolved scales (i.e., the term $\boldsymbol{G}(\boldsymbol{a}, \boldsymbol{b})$), are difficult to construct from the first principles.

4

These challenges promote recent developments that combine data and physical insights in the construction of ROM closure [13, 15, 23, 36, 46, 61]. In particular, the data-driven variational multiscale reduced order model (DD-VMS-ROM) in [49, 59] constructs the basis functions from data by POD and identifies the closure method through regression, leading to an accurate deterministic model for each trajectory. Meanwhile, stochastic closure models for the coefficients of prescribed spectral-Galerkin basis have been constructed in [40, 41, 43] from data consisting of multiple trajectories by statistical inference methods, leading to robust predictive models. In the following, we briefly review these two approaches, highlighting their connections that will foster a combination of their strengths to construct the stochastic DD-VMS-ROM in the next section.

DD-VMS-ROM. The DD-VMS-ROM method introduced in [49, 59] constructs a ROM closure that fits data consisting of a single-trajectory: it first constructs basis functions from data by POD, then it estimates the parameters in the closure model. For example, when $f(u)$ in (2.1) consists of only linear and quadratic terms as typically encountered in fluid flow problems, the two-scale 2S-DD-VMS-ROM [59] is

$$\dot{a} = (A + \widetilde{A})a + a^\top (B + \widetilde{B})a, \tag{2.5}$$

where A and $\widetilde{A}$ are $r \times r$ matrices, while B and $\widetilde{B}$ are $r \times r \times r$ tensors. The term $Aa + a^\top Ba$ corresponds to $F(a)$ in (2.3), which represents the Galerkin projection of $f(u)$ onto the subspace X_1. The additional term $\widetilde{A}a + a^\top \widetilde{B}a$ aims to approximate the model error $G(a, b)$ in (2.3) by quadratic polynomials of the resolved scales, and the entries in $\widetilde{A}$ and $\widetilde{B}$ are estimated from data, consisting of time snapshots in a single solution. The above two-scale model has also been extended to three-scale 3S-DD-VMS-ROM [49] to include finer interactions in hierarchical spaces.

Inference-based closures. The stochastic model reduction method [40, 41, 43] constructs ROM closures by inferring a time series model that approximates the *discrete-time flow map* of the resolved scales. The discrete-time flow map of a is obtained by integrating (2.3) on an time interval $[t_{n-1}, t_n]$,

$$a(t_n) = a(t_{n-1}) + \int_{t_{n-1}}^{t_n} [F(a(s)) + G(a(s), b(s))] \, ds$$
$$= \mathcal{F}(a(t_{n-1}), b(t_{n-1})) \approx F_\theta(a(t_{n-p:n-1}), \xi_{n-p:n-1}), \tag{2.6}$$

where the flow map $\mathcal{F}(a(t_{n-1}), b(t_{n-1}))$ depends on the unresolved scale $b(t_{n-1})$. It is approximated by a functional F_θ depending on the past of a, denoted by a vector with time lags $a(t_{n-p:n-1}) = (a(t_{n-p}), \ldots, a(t_{n-1}))$, since $b(t_{n-1})$ depends on the history of a. Here F_θ is a parametric function with parameter θ estimated from data, and ξ_n is assumed to be sequence of independent identically distributed (IID) Gaussian noise to quantify the uncertainty. In particular, utilizing the quadratic terms, a parametric form of F_θ is shown to be effective for a stochastic Burgers equation [43]:

$$F_\theta(a(t_{n-p:n-1}), \xi_{n-p:n-1}) \approx c_1 a(t_{n-1}) + c_R R^\delta(a) + c^w G(a) + \xi_{n-1}, \tag{2.7}$$

where the parameters (c_1, c_R, c^w) are diagonal matrices, R^δ comes from the RK4 (the fourth-order Runge-Kutta) integrator of the r-mode truncated system, and $G(a)$ denotes a vector with entries

$$G(a)_k = \sum_{l \in \mathcal{A}_{k,r}} \widetilde{a}_l^{n-1} \widetilde{a}_{k-l}^{n-j} \ \text{ with } \widetilde{a}_k^{n-j} = \begin{cases} a_k^{n-j}, & 1 \leqslant k \leqslant r; \\ \sum_{|l| \leqslant r, |k-l| \leqslant r} a_{k-l}^{n-j} a_l^{n-j}, & r < k \leqslant 2r, \end{cases} \tag{2.8}$$

5

where $\mathcal{A}_{k,r} = \{l : |k - l| \leqslant r, r < |l| \leqslant 2r \text{ or } |l| \leqslant r, r < |k - l| \leqslant 2r\}$ and $\boldsymbol{a}_l^m = \boldsymbol{a}_l(t_m)$ for each l, m. The time series model, in the form of a nonlinear autoregression moving average model (NARMA), is inferred from multiple-trajectory data. It takes into account both the model error and the numerical discretization error. Thus, it can reproduce the statistics and the dynamics of the resolved scales. In particular, it can be used for ensemble prediction with uncertainty quantified [9, 43].

Note that when $\boldsymbol{f}(\boldsymbol{u})$ consists of quadratic nonlinearities, the structure of the DD-VMS-ROM in [49] is similar to the NARMA model in [43]. The DD-VMS-ROM allows for fine interactions between the resolved scales through the $\widetilde{B}$ matrix. In NARMA closure, these interactions are represented in a bundle $G(\boldsymbol{a})$ of polynomials of degree up to four, and this bundle has fewer parameters, allowing for time lags to represent the memory effects while avoiding overfitting.

2.2 Stochastic DD-VMS-ROM by inference

Combining the strengths of both the DD-VMS-ROM and the inference-based approaches, we introduce a method to construct robust data-driven stochastic ROM (S-ROM) by inference.

The premise is that the FOM (1.1) is subject to random initial conditions, which are sampled from a given probability measure μ:

$$\dot{\boldsymbol{u}} = \boldsymbol{f}(\boldsymbol{u}), \quad \boldsymbol{u}(0) = \boldsymbol{u}_0 \sim \mu. \tag{2.9}$$

While the method presented below is applicable to general nonlinear systems, to fix ideas, we assume that the nonlinear term $\boldsymbol{f}(\boldsymbol{u})$ is quadratic. For a given reduced dimension r, the S-ROM consists of a quadratic drift part and an additive noise term. Besides the Galerkin projection of $\boldsymbol{f}(\boldsymbol{u})$, the drift part involves additional linear and bilinear terms whose coefficients are optimized based on multiple-trajectory FOM data.

More precisely, the proposed S-ROM fits the FOM data to the time series model

$$\boldsymbol{a}(t_{l+1}) - \boldsymbol{a}(t_l) \approx \left[(A + \widetilde{A})\boldsymbol{a} + \boldsymbol{a}^\top (B + \widetilde{B})\boldsymbol{a} \right] \delta + \sqrt{\delta} \Sigma \boldsymbol{\xi}_l, \quad l = 1, \ldots, n_l, \tag{2.10}$$

where $\boldsymbol{a}$ is $\mathbb{R}^r$-valued for a chosen reduced dimension r, the noise term $\{\boldsymbol{\xi}_l\}$ is an IID sequence of Gaussian random variables with distribution $N(0, Id)$, and $\{t_l = l\delta\}_{l=1}^{n_l}$ are the time instances with a time-step δ which can be relatively large (see Section 3.5 for an exploration). The parameters $\widetilde{A}, \widetilde{B}$ and Σ are estimated by maximizing the likelihood of data consisting of multiple trajectories. We assume Σ to be diagonal for computational efficiency. This model approximates the discrete-time flow map of the process $\boldsymbol{a}$ in (2.6). It is similar to an Euler-Maruyama discretization of the stochastic differential equation

$$\dot{\boldsymbol{a}} = (A + \widetilde{A})\boldsymbol{a} + \boldsymbol{a}^\top (B + \widetilde{B})\boldsymbol{a} + \Sigma \dot{\mathbf{W}}_t, \tag{2.11}$$

where $\dot{\mathbf{W}}_t$ is white noise. This equation is a stochastic version of the 2S-DD-VMS-ROM in (2.5).

The above S-ROM differs from the 2S-DD-VMS-ROM in three aspects: (i) its parameters $\widetilde{A}, \widetilde{B}$ fit multiple-trajectories instead of a single trajectory; (ii) it has an additional noise term to account for the uncertainty in the residual; (iii) it is a discrete-time model instead of a differential system and its parameters are time-step adaptive. As a result, the stochastic ROM can make predictions for new initial conditions sampled from the distribution and it is suitable for ensemble predictions with uncertainty quantification. In other words, it generates a new stochastic process approximating the original stochastic process of the FOM in distribution. Also, it accounts for the discretization error and can tolerate a larger time-step size.

The construction of the stochastic ROM consists of three steps: data generation, extraction of basis functions and inference of the parameters from data. First, we generate multiple-trajectory data by the FOM, which resolves the system with high-resolution in space and time, with initial conditions sampled from μ. Second, we construct basis functions from data, leading to a data-adaptive selection of hierarchical spaces. Third, we infer the parameters in the time series model (2.10) from multiple trajectory data of $\boldsymbol{a}$.

Data generation. We generate data using the FOM with a time-step Δt. The data consists of many trajectories with the initial conditions sampled from the initial distribution:

$$\text{Data of multiple trajectories:} \quad \{Y^{(m)} = \boldsymbol{u}(t^F_{1:N_t})^{(m)} \in \mathbb{R}^{N_x \times N_t}\}^M_{m=1}, \tag{2.12}$$

where $t^F_{1:N_t} = (t^F_1, \ldots, t^F_{N_t})$, with $t^F_l = l\Delta t$ being the time instances. Hereafter, $\boldsymbol{u}(t^F_l) \in \mathbb{R}^{N_x}$ denotes the FOM solution with N_x space grid points at time t^F_l. These fine time instances will be downsampled to infer the S-ROM (2.10) when a time-step δ, larger than Δt, is used to construct the S-ROM.

Extraction of POD basis and solution coefficients. We construct the basis functions from data via POD and extract the coefficients $\boldsymbol{a}$ by projecting the FOM solutions onto the POD basis.

1. (Extract basis functions by ensemble-POD): We get basis functions by computing the eigenvalues and eigenvectors of the square matrix $\overline{K}_M$:

$$\overline{K}_M = \frac{1}{M} \sum_{m=1}^M K^{(m)}, \quad \text{with } K^{(m)} = \frac{1}{N_t} Y^{(m)}(Y^{(m)})^\top \in \mathbb{R}^{N_x \times N_x}. \tag{2.13}$$

Here $K^{(m)}$ can be computed for each trajectory in parallel. We sort the eigenvalues in descending order, denote them by λ_i and denote their eigenvectors by $\varphi_i \in \mathbb{R}^{N_x \times 1}$ for $i = 1, \ldots, N_x$.

2. (Extract trajectory data of the low modes). Projecting the solution $\boldsymbol{u}$ to the POD basis:

$$\boldsymbol{a}_i^{(m)} = \varphi_i^\top Y^{(m)} \in \mathbb{R}^{1 \times N_t}, \quad , i = 1, \ldots, r;\ m = 1, \ldots, M,$$

we obtain the coefficients $\{\boldsymbol{a}_i(t^F_l) = (\boldsymbol{u}(\cdot, t^F_l), \varphi_i(\cdot), i = 1, \ldots, r\}^{N_t}_{l=0}$, which is a multivariate time series encoding the dynamics.

The above construction of the POD basis is almost the same as the classical approach of principle component analysis, which identifies the basis by the eigen–decomposition of the covariance matrix. The minor difference here is the time averaging step that discards the temporal correlations of the process Y, so that we can focus on the dominating modes of the dynamics. It is similar to the ensemble POD in [25, 26], which uses multiple trajectories from systems with different parameters, in the sense that the initial condition can be viewed as a parameter of the system. The eigen–decomposition of $\overline{K}_M$ can be done directly (e.g., by Cholesky factorization) when its dimension is not too large, and more advanced methods (e.g., the nested-POD [34]) are available for higher dimensions.

Inference of the parameters. We estimate the parameters by maximizing the likelihood of data for the model (2.10). We start from downsampling the data trajectories in time to fit the time-step of the S-ROM. That is, each projected FOM trajectory $\boldsymbol{a}(t_{1:N_t}^F)$ with $t_j^F = j\Delta t$ is downsampled to $\boldsymbol{a}(t_{1:n_t})$ with $t_l = l\delta$. In other words, if $\delta = \mathrm{Gap}\Delta t$, we make observations every Gap steps from the fine data. Denote the downsampled data by

$$\textbf{Data } \{\boldsymbol{a}^{(m)}(t_{1:n_t})\}_{m=1}^M, \ t_l = l\delta = l\mathrm{Gap}\Delta t.$$

Since the S-ROM depends on the parameters linearly, the parameters $(\widetilde{A}, \widetilde{B})$ are estimated by regression, which is equivalent to maximizing the likelihood of the data trajectories:

$$(\widehat{A}, \widehat{B}) = \underset{(\widetilde{A}, \widetilde{B})}{\arg\min} \, l(\widetilde{A}, \widetilde{B}) := \frac{1}{Mn_t} \sum_{m=1}^M \sum_{l=1}^{n_t} \|F^{(m)}(t_l) - \widetilde{A}\boldsymbol{a}^{(m)}(t_l) - (\boldsymbol{a}^{(m)}(t_l))^\top \widetilde{B}\boldsymbol{a}^{(m)}(t_l)\|^2, \quad (2.14)$$

where, with the notation $\Delta\boldsymbol{a}(t_l) = \boldsymbol{a}(t_{l+1}) - \boldsymbol{a}(t_l)$, the variable F is defined by

$$F^{(m)}(t_l) = \frac{1}{\delta}\Delta\boldsymbol{a}^{(m)}(t_l) - (A\boldsymbol{a}^{(m)} + \boldsymbol{a}^\top B\boldsymbol{a}^{(m)})(t_l) \in \mathbb{R}^{r\times 1}. \quad (2.15)$$

The diagonal Σ is estimated by the residual of the regression.

Note that $\widetilde{B} \in \mathbb{R}^{r\times r\times r}$ is a 3D array such that for each k, the matrix $(\widetilde{B}(i, i', k))_{1\leqslant i, i'\leqslant r}$ is symmetric because it is the coefficient matrix of the terms $\boldsymbol{a}_i\boldsymbol{a}_{i'}$. Thus, including $\widetilde{A} \in \mathbb{R}^{r\times r}$, we have $n_r = r + (r+1)r/2$ parameters to be estimated for each of the r modes. To write in the form of least squares, we denote the estimator by $\boldsymbol{c} = (\widehat{A}(i, k)_{1\leqslant i, k\leqslant r}, (\widehat{B}(i, i', k))_{1\leqslant i'\leqslant i\leqslant r, 1\leqslant k\leqslant r}) \in \mathbb{R}^{n_r\times r}$, and denote the normal matrix and vector of regression by

$$\boldsymbol{A}^{(m)}(j, j') = \frac{1}{n_t} \sum_{l=1}^{n_t} \psi_j(\boldsymbol{a}^{(m)}(t_l))\psi_{j'}(\boldsymbol{a}^{(m)}(t_l)), \ 1 \leqslant j, j' \leqslant n_r,$$
$$\boldsymbol{b}^{(m)}(j) = \frac{1}{n_t\delta} \sum_{l=1}^{n_t} \psi_j(\boldsymbol{a}^{(m)}(t_l))F^{(m)}(t_l) \in \mathbb{R}^r, \ 1 \leqslant j \leqslant n_r, \quad (2.16)$$

where $(\psi_j(\boldsymbol{a}(t_l)), 1 \leqslant j \leqslant n_r)$ is defined by

$$\psi_j(\boldsymbol{a}(t_l)) = \begin{cases} \boldsymbol{a}_j(t_l), & 1 \leqslant j \leqslant r; \\ \boldsymbol{a}_i(t_l)\boldsymbol{a}_{i'}(t_l), & j = i \times r + r' \text{ with } 1 \leqslant i' \leqslant i \leqslant r. \end{cases}$$

Then, the estimator is solved by

$$\boldsymbol{c}_M = \boldsymbol{A}_M^{-1}\boldsymbol{b}_M, \quad \text{with } \boldsymbol{A}_M = \frac{1}{M}\sum_{m=1}^M \boldsymbol{A}^{(m)}, \quad \boldsymbol{b}_M = \frac{1}{M}\sum_{m=1}^M \boldsymbol{b}^{(m)}, \quad (2.17)$$

where $\boldsymbol{A}_M^{-1}$ denote the Moore–Penrose pseudo-inverse when the normal matrix is singular. In practice, when $\boldsymbol{A}_M$ is ill-conditioned or singular, a regularization term often helps to lead to a robust estimator (see Section A for more details). The residual provides us the estimator for Σ:

$$\Sigma_M = \mathrm{Diag}(\sigma(1), \ldots, \sigma(r)), \quad \text{with } \sigma(k)^2 = \|\boldsymbol{A}_M\boldsymbol{c}_M(\cdot, k) - \boldsymbol{b}_M(\cdot, k)\|_{\mathbb{R}^{n_r}}^2. \quad (2.18)$$

The resulted discrete-time S-ROM is designed to account for both the model error and discretization error. It has a consistent estimator (i.e., the estimator converges as the number of trajectories increases (see Theorem 2.2 in the next section). Furthermore, it can tolerate a relatively large time step size [16, 39, 43], and we will demonstrate it in Section 3.5 for a viscous Burgers equation.

Computational complexity. At the learning stage, the major computation cost occurs when reading the FOM data twice: one for the construction of the POD basis $\{\varphi_k\}_{k=1}^r$ and the other for the computation of the coefficients $\{\boldsymbol{a}(t_{1:n_t}))^{(m)}\}_{m=1}^M$. The rest of the computation uses only these coefficients, whose dimension is significantly lower than the dimension of the FOM data. We implement all the trajectory-wise computation in parallel, from the evaluation of $K^{(m)}$ in (2.13) to the estimator in (2.17). In particular, during the computation of the estimator, we first compute $\boldsymbol{A}^{(m)}$ and $\mathbf{b}^{(m)}$ in (2.16) for each trajectory in parallel, then we assemble them as in (2.17).

2.3 Convergence of the estimators

We show that the POD modes (the eigenvectors) and the parameter estimator converge as the number of trajectories increases. Such a convergence follows from the Central Limit Theorem, because the trajectories are independent and identically distributed with the randomness comes from the initial conditions.

The following assumption requires the solution field to be uniformly bounded almost surely, and it holds true for the weak solution of Equation (2.1) with a wide range of nonlinearities and initial distributions, including the Burgers equation with random initial conditions to be studied in the next section.

Assumption A: Assume that the FOM solution $\boldsymbol{u}$ of (2.9), defined for $t \in [0, T]$ and $x \in D \subset \mathbb{R}^d$ with D being a bounded domain, satisfies $\|\boldsymbol{u}\|_\infty = \sup_{t \in [0,T], x \in D} |u(t, x)| < \infty$ for almost all initial conditions sampled from the initial distribution μ.

The next theorem shows that the POD eigenvectors estimated in the previous section converge as the number of trajectories increases.

Theorem 2.1 (Convergence of POD eigenvectors) *The POD eigenvectors and eigenvalues converge as $M \to \infty$ under* Assumption A. *More precisely,*

- *Each eigenvalue of the averaged covariance matrix $\overline{K}_M$ in (2.13), denoted by λ_k^M, converges to the corresponding eigenvalue λ_k of $\mathbb{E}[\overline{K}_M] = \lim_{M \to \infty} \overline{K}_M$, and $M^{-1/2}(\lambda_k^M - \lambda_k)$ is asymptotically normal.*

- *Each eigenvector of the eigenvalues with multiplicity one, denoted by φ_k^M, converges in $L^2(D)$ almost surely to the corresponding eigenvector φ_k of $\mathbb{E}[\overline{K}_M]$, and $M^{-1/2}(\varphi_k^M - \varphi_k)$ is asymptotically normal.*

Proof. This is a classical result in principle component analysis, see e.g., [19, Proposition 8-10]. ∎

Remark 2.1 *When the data is continuous (i.e., not on the spatial grid points), the principal component analysis shows that similar convergence holds true for the eigenvalues and the eigenfunctions of the operators [19]. That is, let $K_M(x, y) = \frac{1}{M} \sum_{m=1}^M \frac{1}{T} \int_0^T [u(t,x)u(t,y)]dt$ and define the operator $L_{K_M} : L^2(D) \to L^2(D)$ by $L_{K_M}\varphi(x) = \int_D K_M(x,y)\varphi(y)dy$. Then, the eigenvalues and eigenfunctions of L_{K_M} converge as M increases.*

The next theorem shows that the estimator of $(\widetilde{A}, \widetilde{B})$ is asymptotically normal as M increases.

Theorem 2.2 (Convergence of parameter estimator) *Suppose that Assumption A holds true. Suppose that the smallest eigenvalue of $\boldsymbol{A}_\infty = \lim_{M \to \infty} \boldsymbol{A}_M = \mathbb{E}[\boldsymbol{A}_M]$, the expectation of the random matrix $\boldsymbol{A}^{(m)}$ in (2.16), is positive. Then, the estimator $\boldsymbol{c}_M$ in (2.17) converges to $\boldsymbol{c}_* = \boldsymbol{A}_\infty^{-1}\mathbf{b}_\infty$, where $\mathbf{b}_\infty = \mathbb{E}[\mathbf{b}_M]$, almost surely and $M^{-1/2}(\boldsymbol{c}_M - \boldsymbol{c}_*)$ is asymptotically normal.*

9

Proof. Note that by the strong Law of Large Numbers, $\boldsymbol{A}_M \to \boldsymbol{A}_\infty$ and $\mathbf{b}_M \to \mathbf{b}_\infty$ almost surely as $M \to \infty$. Thus, $\boldsymbol{A}_M^{-1} \to \boldsymbol{A}_\infty^{-1}$ almost surely (using the fact that $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$). Then, $\boldsymbol{c}_M = \boldsymbol{A}_M^{-1}\mathbf{b}_M \to \boldsymbol{c}_* = \boldsymbol{A}_\infty^{-1}\mathbf{b}_\infty$ almost surely, i.e. the estimator is consistent. Meanwhile, note that $\sqrt{M}(\mathbf{b}_M - \mathbf{b}_\infty)$ is asymptotically normal by the Central Limit Theorem. Together with the almost sure convergence of $\boldsymbol{A}_M^{-1}$, we obtain the asymptotic normality of $M^{-1/2}(\boldsymbol{c}_M - \boldsymbol{c}_*)$. ∎

We conclude this section by noting that the above convergence is in the number of independent trajectories, and it does not take into account of the length of the trajectories. While the length of the trajectories plays a limited role in the convergence in M, long trajectories help to identify the dominating modes for the longer-term dynamics. Also, the parameter estimator and the POD basis depend on the length of the data (in addition to the initial distribution).

3   Numerical results for a viscous Burgers equation

In this section, we perform a numerical investigation of the S-ROM for the one-dimensional viscous Burgers equation. In Section 3.1, we present the mathematical and computational setups. Sections 3.2-3.3 examine the convergence of the POD basis and the parameters learnt from data. Section 3.4 investigates the performance of the S-ROM in making predictions and quantifying the uncertainties. Finally, taking advantage of the S-ROM's features of being efficient and adaptive to spatial resolution and time-step, we explore optimal space-time reduction in Section 3.5.

3.1   The one-dimensional viscous Burgers equation and the numerical setup

As an illustration of the stochastic DD-VMS-ROM framework presented in Section 2, we consider the viscous Burgers equation posed on $(0, 1)$ and supplemented with homogeneous Dirichlet boundary conditions and a random initial condition:

$$
\begin{aligned}
u_t &= \nu u_{xx} - u u_x, \quad 0 < x < 1,\ t > 0,\\
u(0, t) &= u(1, t) = 0, \quad t \geqslant 0,\\
u(\cdot, 0) &= u_0(\cdot, \omega) \sim \mu.
\end{aligned}
\tag{3.1}
$$

Here $u_0(\cdot, \omega) \sim \mu$ means that the initial condition is sampled from the measure μ on $L^2(0, 1)$. The randomness from initial conditions is important for data-driven modeling of the dynamics because it generates data trajectories that can sufficiently explore the dynamics of system, and it has been utilized in [10, 12, 56].

In the numerical experiments, we set the viscosity constant to be $\nu = 2 \times 10^{-3}$, and consider smooth random initial conditions in the form

$$
u_0(x, \omega) = \sum_{k=1}^{K} \frac{w_k(\omega)}{k} \sin(\pi k x),
\tag{3.2}
$$

where we set $K = 50$ to allow for fast oscillations in the sampled initial data profiles, and each w_k is a random number sampled from the normal distribution with mean $\mu_u = 0.5$ and standard deviation $\sigma_u = 0.2$. A few typical such random initial conditions are shown in the left panel of Figure 1.

For each random initial condition, the initial boundary value problem (3.1) is solved with a finite element method, in which the spatial domain is discretized using piecewise linear finite elements with a uniform spatial mesh size $h = 2^{-8}$, and the temporal discretization is performed by the implicit Euler method with a step size $\Delta t = 5 \times 10^{-3}$. The FOM solution is computed over the time window $[0, 2]$, leading to $N_t = 401$ snapshots for each random initial condition. The number

10

of spatial grid points is $N_x = 257$ for the chosen spatial resolution h. Thus, each FOM trajectory is stored in an $N_x \times N_t$ matrix, denoted by $Y^{(m)}$ for the m-th trajectory according to (2.12). A typical FOM solution is shown in the right panel of Figure 1.
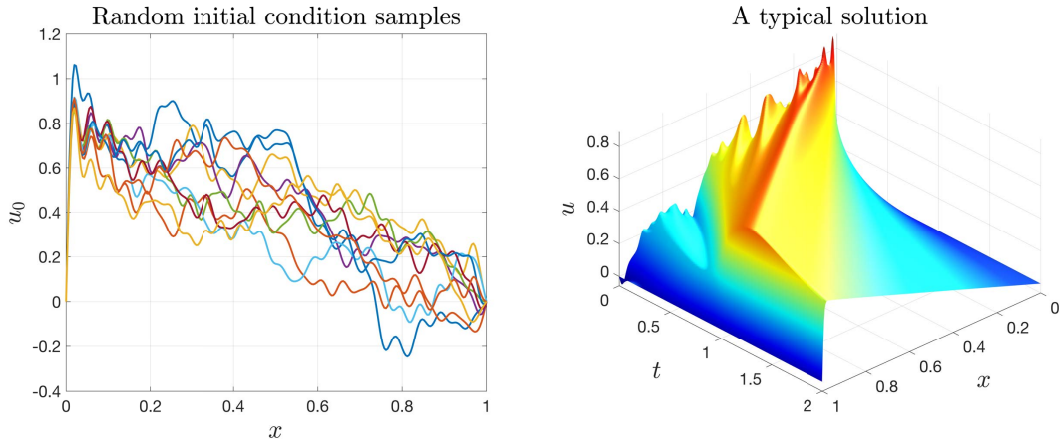

Figure 1: **Left panel**: 10 random initial conditions sampled according to (3.2). **Right panel**: A solution of (3.1) over the time window $[0, 2]$ with a random initial condition sampled from (3.2).

We examine the construction of S-ROM of the form (2.11) for the problem (3.1) in two groups of assessments. First, we investigate the convergence of the basis functions and the parameter estimator as the number of data trajectories increases. Second, we examine the performance of the S-ROM in two scenarios: trajectory-wise prediction and ensemble prediction. The trajectory-wise prediction aims to predict the deterministic dynamics, and the ensemble prediction aims to quantify the spread of the uncertainty from the unresolved modes. We will report the root mean square errors (RMSEs) of these predictions. In each simulation, we solve the estimated discrete-time S-ROM in (2.10) exactly as it is: we set $\Sigma = 0$ when making trajectory-wise prediction, and we use the estimated Σ with randomly sampled $\{\boldsymbol{\xi}_l\}$ when making ensemble prediction.

3.2 POD Basis from multiple trajectories

We first check how the estimated POD modes and their corresponding eigenvalues stabilize as the number of training trajectories, M, increases. It turns out that the dominant POD modes estimated with a relatively small number of trajectories (e.g. $M = 20$) already capture qualitatively the shape of those obtained with significantly more trajectories. To facilitate a quantitative assessment, we denote the POD modes learnt from the first M trajectories by $\{\varphi_j^M\}$, and we take those estimated with $\overline{M} = 1000$ as the reference. The numerics reveal that the first 10 POD modes, $\varphi_1^{\overline{M}}, \ldots, \varphi_{10}^{\overline{M}}$, already capture above 99.9% of the averaged kinetic energy in each of the 1000 training trajectories. We thus focus on these first 10 modes.

Panels (A)–(C) of Figure 2 show the estimated modes φ_1, φ_5, and φ_{10} as M increases. As can be observed, the estimated modes with $M = 400$ are already almost indistinguishable with their respective reference ($\overline{M} = 1000$). The eigenvalues also stabilize quickly as shown in Panel (D) of Figure 2. We also computed the L^2-error for each POD mode, $\|\varphi_j^M - \varphi_j^{\overline{M}}\|_{L^2(0,1)}$, as the number of training trajectories M increases towards $\overline{M}$. The L^2-error for each mode follows a decreasing trend as M increases with although local fluctuations; also, for each fixed M, the lower indexed modes have smaller errors. Panel (E) of Figure 2 show these L^2-errors for the modes φ_1, φ_5, and φ_{10}.
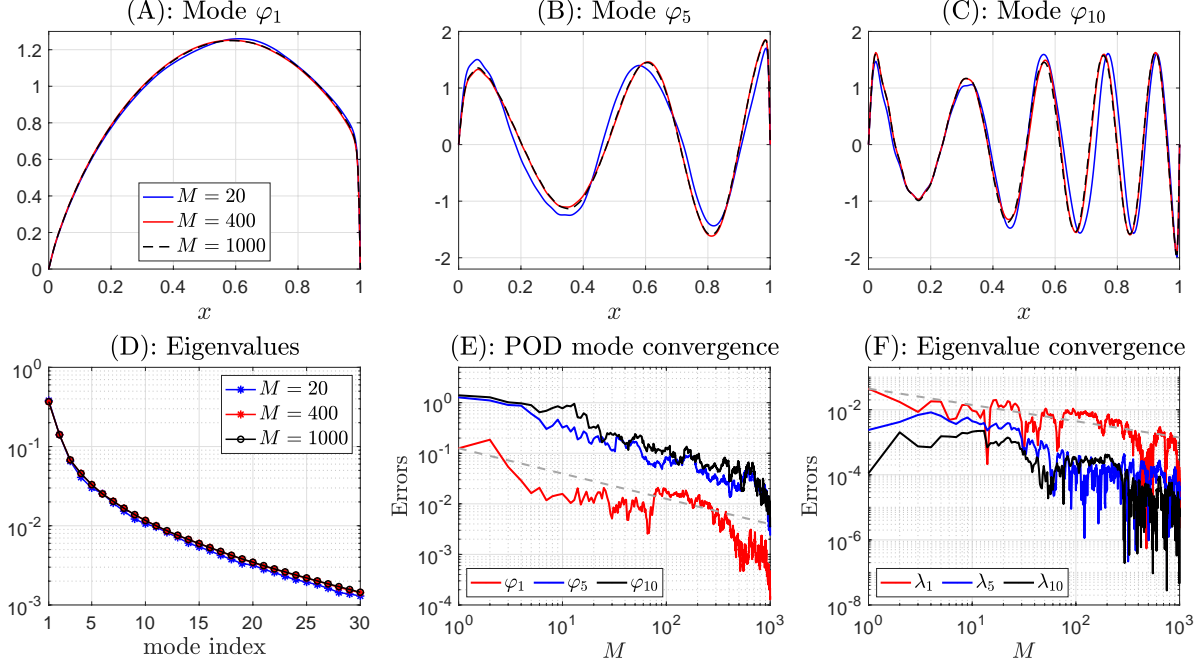
11

Figure 2: **Panels (A-C)**: POD modes φ_1, φ_5 and φ_{10}, computed from the FOM solution datasets with $M = 20$, 400, and 1000 trajectories, respectively. **Panels (D)**: The first 30 POD eigenvalues. **Panel (E)**: L^2-error for the j-th POD mode, $\|\varphi_j^M - \varphi_j^{\overline{M}}\|_{L^2(0,1)}$, as the number of training trajectories M increases towards $\overline{M} = 1000$, shown for $j = 1, 5$, and 10. **Panel (F)**: Absolute error for the j-th POD eigenvalue, $|\lambda_j^M - \lambda_j^{\overline{M}}|$, as the number of training trajectories M increases towards $\overline{M} = 1000$, shown for $j = 1, 5$, and 10. The computed POD modes and their corresponding eigenvalues all show a clear trend of convergence as the number of training trajectories M increases, at a rate close to the theoretical rate $M^{-1/2}$ in Theorem 2.1, which is indicated by the dashed gray lines with slope -0.5. See Section 3.2 for details.

A noteworthy feature of these L^2-errors is a power law decay with exponent -0.5 as marked by the dashed gray line in Panel (E) of Figure 2. As shown in Theorem 2.1, such a power law decay with exponent -0.5 is just a manifestation of the Central Limit Theorem since the trajectories are independent and identically distributed with randomness from the initial condition. Also, in agreement with Theorem 2.1, this power law decay is also visible in the absolute error for the POD eigenvalues, $|\lambda_j^M - \lambda_j^{\overline{M}}|$, as shown in Panel (F) of Figure 2, with an even more negative scaling exponent for some of the higher indexed modes.

3.3 Convergence of parameter estimator

We show next that the parameter estimator converge as the number of trajectory increases. Meanwhile, we show that the trajectory-wise estimator can vary largely between trajectories. Thus, it is important to estimate the parameters using multiple trajectories when constructing ROM for predictions from different initial conditions, particularly for random or stochastic systems.

Figure 3 shows the estimators of the S-ROM with $r = 10$ POD modes (those estimated from 1000 trajectories in the previous section) and with a time gap Gap = 5. Panel (A) shows that the trajectory-wise estimators of $\widetilde{A}(1, 1)$ and $\widetilde{B}(1, 1, 1)$ vary largely between trajectories, and Panel
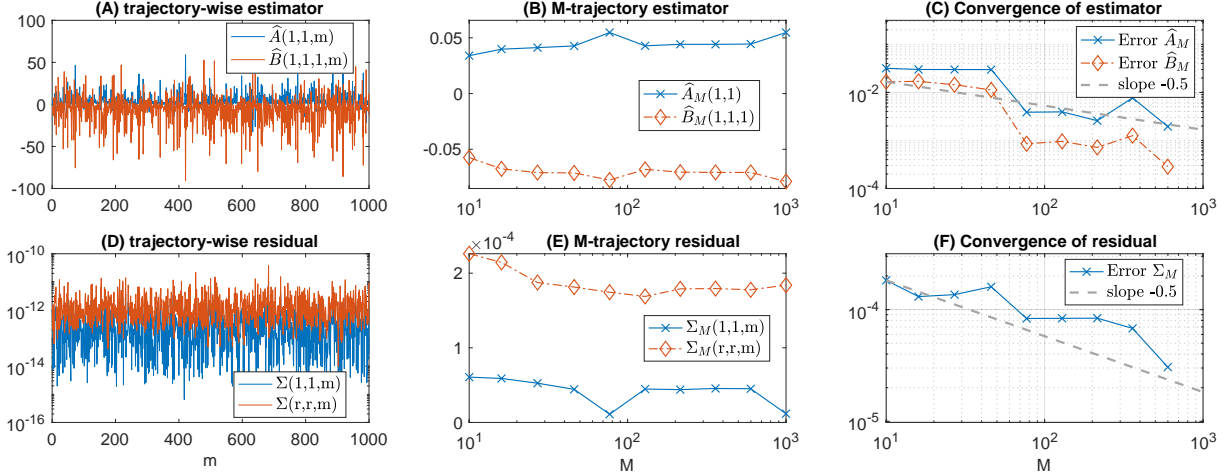
Figure 3: Convergence of parameter estimator. **Panel (A)** shows that the single-trajectory estimators vary largely between trajectories, and **Panel (D)** shows that the corresponding residuals are extremely small. Together, they indicate an overfitting by the single-trajectory ROM. On the contrary, **Panel (B) and (E)** show that the M-trajectory estimators stabilize quickly as M increases. **Panel (C) and (F)** show the errors of the estimators in comparison to the reference estimator with $\overline{M} = 1000$. The errors decay at a rate close to the theoretical rate $M^{-0.5}$ in Theorem 2.2. See Section 3.3 for details.

(D) shows that the residuals of the trajectory-wise estimators are extremely small, indicating an overfitting. Together they show that while a trajectory-wise ROM can fit each specific trajectory data well with a negligible residual, it is sensitive to the data and is not generalizable. On the other hand, Panel (B) shows that the M-trajectory estimators of $\widetilde{A}(1,1)$ and $\widetilde{B}(1,1,1)$ stabilizes very quickly: as can be observed, the estimators with $M = 10$ are already close to those with $M = 1000$ trajectories. Similar convergence trend is observed for the corresponding residuals in Panel (E). We remark that such a quick stabilization is partially due to that the POD modes are estimated from 1000 trajectories for all the estimators; if the POD modes are estimated for each sample size, which will oscillate as we have seen in Figure 2, we expect a sightly slower stabilization.

Figure 3 (C) and (F) further show that the estimators converge at a rate close to $M^{-0.5}$, in agreement with Theorem 2.2. Here we compute the errors of the estimators in comparison to the reference estimators using $\overline{M} = 1000$ trajectories. That is, for each of $M \in \{\lfloor 10^{1+j\alpha} \rfloor, j = 0, \ldots, 9\}$ with $\alpha = 2/9$, the errors of $\widehat{A}_M$, $\widehat{B}_M$ and Σ_M are computed by the Frobenius norms defined by

$$\|\widehat{A}_M - \widehat{A}_{\overline{M}}\|^2 = \frac{1}{r^2} \sum_{1 \leq i,k \leq r} |\widehat{A}_M(i,k) - \widehat{A}_{\overline{M}}(i,k)|^2,$$

$$\|\widehat{B}_M - \widehat{B}_{\overline{M}}\|^2 = \frac{2}{r^2(r+1)} \sum_{1 \leq i \leq i' \leq r, 1 \leq k \leq r} |\widehat{B}_M(i,i',k) - \widehat{B}_{\overline{M}}(i,i',k)|^2,$$

$$\|\Sigma_M - \Sigma_{\overline{M}}\|^2 = \frac{1}{r} \sum_{1 \leq k \leq r} |\Sigma_M(k,k) - \Sigma_{\overline{M}}(k,k)|^2.$$

The parameters of S-ROM estimated from $\overline{M}$ trajectories are presented in Appendix B.

13

3.4 S-ROM performance

We examine the performance of the S-ROM in two scenarios: deterministic single trajectory prediction and stochastic ensemble prediction. In the trajectory-wise prediction, the S-ROM makes a single trajectory prediction for each given initial condition by setting the stochastic force to be zero (i.e., $\Sigma = 0$). We compare a typical solution field of S-ROM with those of the r-mode projection of the FOM solution and G-ROM (the Galerkin ROM with r-modes, i.e., $\dot{\boldsymbol{a}} = A\boldsymbol{a} + \boldsymbol{a}^\top B\boldsymbol{a}$). We also report the statistics of the root mean square errors (RMSE) of the S-ROM to the r-mode projection of the FOM in multiple predictions. In the stochastic ensemble prediction, we turn on the stochastic force in the S-ROM and generate an ensemble of trajectories for each initial condition. The ensemble represents the spread of the uncertainty from the unresolved scales (i.e., the conditional distribution of the process), which is important for data assimilation [9, 42].

The numerical settings are as follows. The S-ROM and G-ROM have $r = 10$ POD modes (since it captures 99.9% of the averaged kinetic energy of almost each trajectory, see Section 3.2) and a time step 0.025 (i.e., with time gap = 5). The S-ROM uses the parameters estimated from Section 3.3. The solutions are on the time interval $[0, 4]$, twice the length of the training time interval. Their initial condition is the r-mode projection of the FOM's initial condition.
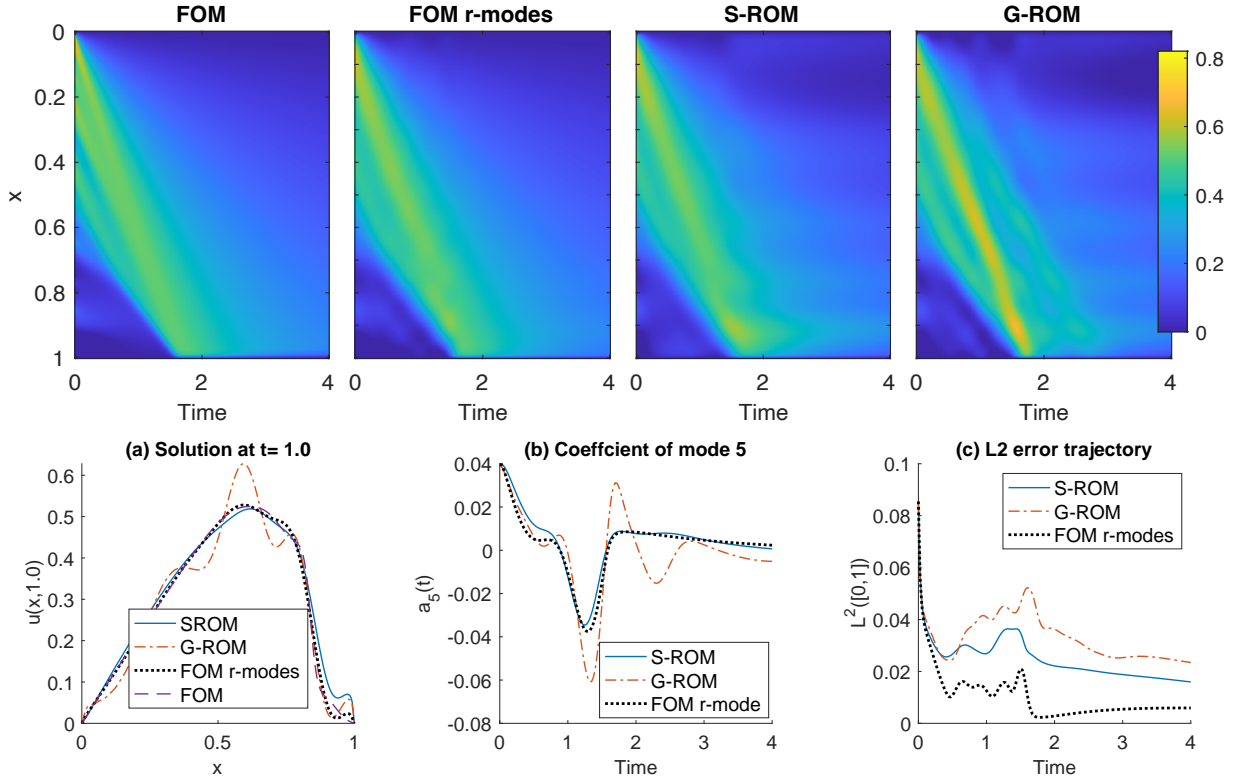

Figure 4: A typical solution trajectory of the S-ROM, in comparison with the FOM, the r-mode projection of the FOM, the G-ROM solutions. The **top row** compares the four solution fields. The **bottom row** compares the solutions at time $t = 1.0$ in **(a)**, the trajectory of $\boldsymbol{a}_5(t)$ in **(b)** and the trajectory of $L^2(0, 1)$ errors, in comparison to the FOM solution in **(c)**. As can be seen, the S-ROM solution is more accurate than the G-ROM in approximating the FOM.

Deterministic single trajectory prediction Figure 4 compares a typical solution field of S-ROM with those of the FOM, the r-mode projection of FOM (denoted by "FOM r-modes"), and the G-ROM. It shows that the S-ROM solution is more accurate than the G-ROM in approximating the r-mode projection of the FOM. More specifically, the top row shows that the S-ROM has a solution field significantly closer to the r-mode projection of the FOM than the G-ROM's. The bottom row provides detailed profiles of the solution field: the spatial profile at time $t = 1.0$ in **(a)** and the trajectory of $\boldsymbol{a}_5(t)$ in **(b)**. The spatial curves of $u(x, 1.0)$ in **(a)** shows that the S-ROM is significantly closer to the FOM's r-mode projection than the G-ROM. Similar superior performance is observed in **(b)**, which presents the trajectory of $\boldsymbol{a}_5(t)$. Furthermore, **(c)** shows that S-ROM has $L^2(0,1)$ errors smaller than those of the G-ROM. Here the $L^2(0,1)$ error is computed as $\|\widehat{u}_r(\cdot, t) - u(\cdot, t)\|^2_{L^2(0,1)}$, with $\widehat{u}_r$ coming from the S-ROM, the G-ROM and the r-mode projection of the FOM. We note that the r-mode projection of the FOM has non-negligible errors because of the missing higher modes.
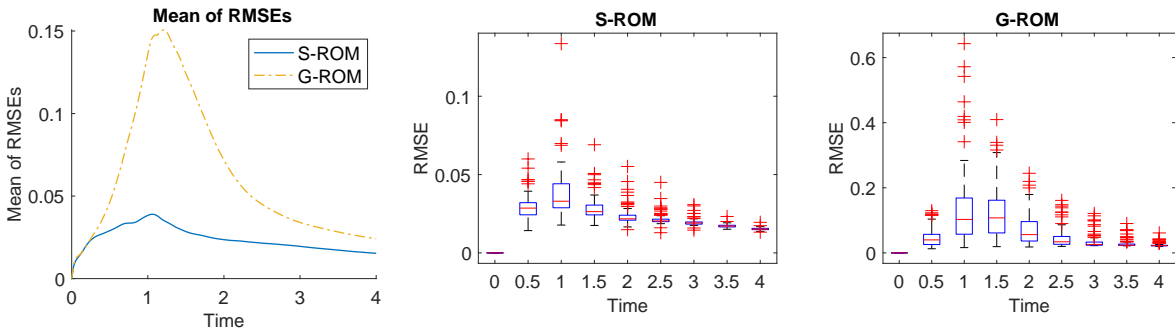

Figure 5: RMSEs of 100 prediction trajectories by the S-ROM and G-ROM. **Left:** the mean of the RMSEs; **Middle and Right:** the box-plots of the RMSEs in the time interval $[0, 4]$. For each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25-th and 75-th percentiles, respectively. The whiskers extend to the most extreme data points without outliers, and the outliers are plotted individually using the "+" marker symbol. The S-ROM has RMSEs significantly smaller than those of the G-ROM.

We further examine the superior performance of S-ROM over G-ROM in the prediction with 100 new initial conditions, and report the statistics of the RMSEs of the trajectories in Figure 5. The RMSEs are computed as

$$\text{RMSE(t)} = \|\widehat{\boldsymbol{a}}(t) - \boldsymbol{a}(t)\| = \Big(\sum_{j=1}^{r} |\widehat{\boldsymbol{a}}_j(t) - \boldsymbol{a}_j(t)|^2\Big)^{1/2}, \tag{3.3}$$

for $t \in [0, 4]$ for each trajectory, where $\widehat{\boldsymbol{a}}(t)$ comes from the S-ROM or G-ROM, and $\boldsymbol{a}(t)$ is the r-mode projection of the FOM. These plots clearly show the improvement brought by the S-ROM: (1) the S-ROM's median RMSEs are less than 0.04, while those of G-ROM are about 0.15, about three times larger. Similar improvements are observed for the 25-th and the 75-th percentiles. (2) the outliers of S-ROM are less than 0.15, while those of the G-ROM exceeds 0.6. Thus, the S-ROM is significantly more accurate than the G-ROM in approximating the r-mode projection of the FOM.

Stochastic ensemble prediction. Next, we investigate the stochastic ensemble prediction by the S-ROM from an initial condition when the stochastic force is turned on. This is the setting
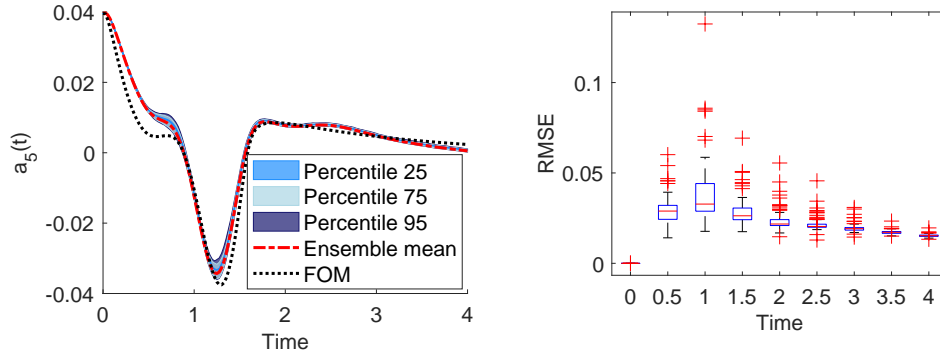
Figure 6: Stochastic ensemble prediction. **Left:** An ensemble of 100 random trajectories of the S-ROM from an initial condition, as well as its ensemble percentiles (25,75 and 95) and mean (in red color), in comparison with the FOM's trajectory. **Right:** the box-plot of the RMSEs of the ensemble mean (as shown in left) in 100 independent simulations.

when we have only the partial observation of the r-model of the FOM, and we would like to predict the future dynamics without resolving the FOM. In this setting, the S-ROM ensemble is the output corresponding to the realizations of the stochastic force. The ensemble represents the spread of the uncertainty from the unresolved scales, which is important for data assimilation [9, 42]. The left panel in Figure 6 shows the ensemble with 100 trajectories of $\boldsymbol{a}_t(t)$ (in cyan color) from a fixed initial condition, along with the ensemble mean and the trajectory of the FOM. It can be seen that the ensemble spreads out in the time interval $[0, 1]$ and the spread is the widest when the trajectory changes convexity, which is due to nonlinear interaction between modes; and the spread becomes narrow after $t = 2$ when the shock is almost formulated (see Figure 4 top row and Figure 1 right panel). The right panel of Figure 6 further shows the statistics of the RMSEs of the ensemble mean in 100 independent such simulations (that is, in each simulation, we generate an ensemble of 100 prediction trajectories, obtain the ensemble mean, and then compute the RMSE of the ensemble mean as in (3.3)). The statistics of the RMSEs are similar to those of deterministic prediction Figure 5 (middle), reflecting the fact that the noise strength parameter Σ is very small (see Figure 3) compared with other parts of the vector field in the ROM.

3.5 Discussion on space-time reduction

The most attractive feature of the S-ROM is that it is adaptive to the resolution r and time step size δ. In the previous section, we focused on demonstrating the performance of S-ROM with $r = 10$ and $\delta = 0.025$ (i.e., time gap = 5). A natural question is that when (r, δ) changes, how will the S-ROM adapt. In particular, for a given resolution r, what is the maximal time step size that the S-ROM can remain stable and what is the optimal time step size such that the S-ROM makes the most accurate predictions.

To answer these questions, we test 90 S-ROMs with $r \in \{6, 8, 10, 12, 14, 16\}$ and with $\delta =$ Gap $\times 0.005$ with Gap $\in \{1, \ldots, 15\}$. All the S-ROMs are trained from the dataset consisting of 1000 data trajectories, downsampled with time gap from the FOM solution with time step size $\Delta t = 0.005$. For each (r, δ), we estimate an S-ROM, compute the mean RMSEs of 200 deterministic predictions by the estimated S-ROM, and report the time-average of the mean RMSE trajectory in Figure 7. Here the darkest blue color with the "NAN" marker indicates that the S-ROM has
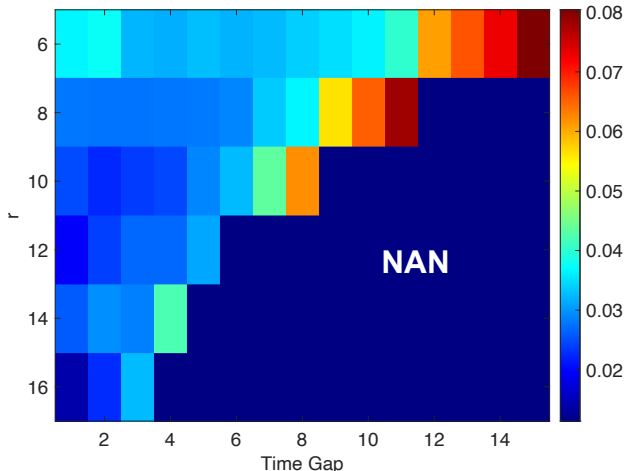
Figure 7: Average RMSEs of trajectory prediction from 90 S-ROMs with resolutions $r \in \{6, 8, \ldots, 16\}$ and time step sizes $\delta = \text{Gap} \times 0.005$ with $\text{Gap} \in \{1, \ldots, 15\}$. For each pair (r, Gap), an S-ROM is estimated from data consisting of 1000 trajectories on time interval $[0, 2]$, and it is used to make predictions on the time interval $[0, 4]$ from 200 random initial conditions; the average RMSE is the time-average of the RMSEs of these 200 predicted trajectories. The darkest blue color with the "NAN" marker indicates that the S-ROM has at least one blowup solution among the 200 predictions. As r increases, the maximal time step size of stability decreases, from $\text{Gap} = 11$ for $r = 8$ to $\text{Gap} = 3$ for $r = 16$, due to the increasing stiffness. Note that for $r \in \{6, 8, 10\}$, the average RMSE first decreases then increases as the time gap increases, thus, suggesting that the S-ROM makes the most accurate prediction when the time step size is medium large.

at least one blowup solution among 200 prediction trajectories. As r increases, the stiffness of the S-ROM increases, and its maximal time step size of stability decreases, from $\text{Gap} = 11$ for $r = 8$ to $\text{Gap} = 3$ for $r = 16$. When the resolution is $r = 6$, the S-ROMs with all the time gaps are stable, but the average RMSE first decreases then increases as the time gap increases. Such a pattern is also observed for the S-ROMs with $r = 8$ and $r = 10$. This pattern suggests that the S-ROM makes the most accurate prediction when the time step size is medium large.

In summary, the S-ROMs are adaptive to resolution and time-step size. Its performance is best when there is a balanced space-time reduction. It remains open to understand the mechanism of such a balanced space-time reduction, and its connection with numerical error and statistical learning.

4 Conclusions

We have proposed and investigated a parametric inference approach to construct data-driven stochastic reduced order models (S-ROM) that combines physical insights with data. The S-ROM is trained from multiple-trajectory data with random initial conditions and it is robust to make accurate predictions from new initial conditions. The framework consists of two stages. First, it constructs basis function from data by POD. Then, it infers a closure model for the coefficients of the first POD modes from the projected data trajectories. The current study focuses on systems with quadratic nonlinearities and constructs closure terms with linear and quadratic terms. We prove that the estimated POD basis and the parameters in the S-ROM converge at a rate $M^{-1/2}$, with M being

17

the number of data trajectories.

We demonstrate the performance of the S-ROM on a viscous Burgers equation with random initial conditions. Numerical results verify the convergence of the POD basis and the parameters. They also confirm the superior performance of the S-ROM in making accurate predictions and representing the spread of the uncertainty due to the unresolved scales. Furthermore, the S-ROM is adaptive to the spatial resolution and time-step size. For a given spatial resolution, we demonstrate that the S-ROM makes the most accurate prediction when the time step size is medium large.

We conclude by reiterating some key features of the novel S-ROM framework and the issues to be further investigated in future work. The framework of constructing S-ROM via discrete-time flow map approximation has four major advantages:

1. It is applicable to general data-driven models for high-dimensional dynamical systems, either random or stochastic. In particular, when there are physical insights to provide parameter forms, the ROM can be efficiently trained by regression with proper regularization to achieve performance guarantee.
2. The reduction is in both space and time. Thus, the S-ROM achieves efficient simulation that can be orders of magnitudes faster than the FOM.
3. The training can be done using only short trajectories for autonomous systems, and the constructed S-ROM can make predictions far beyond the training time interval, because the S-ROM learns a flow map that does not vary in time. Furthermore, the S-ROM can make prediction for new initial conditions sampled from the initial distribution.
4. With a small noise term, the S-ROM captures the short-term dynamics of the resolved modes by its deterministic parts, and can represent the spread of the uncertainty due to unresolved scales. Thus, the S-ROM is well-suited for predictive modeling that requires uncertainty quantification, e.g., data assimilation.

There are also a few aspects to be further investigated:

1. The constructed S-ROM depends on the initial distribution through the training data. The reason is that, although the flow map of the FOM (the solution operator) is independent of the initial condition, the flow map of the ROM variables depends on the initial condition of the unresolved variables. Thus, when the initial distribution changes, the S-ROM must be trained again. It is of interest to understand how the S-ROM parameters depend on the initial distribution.
2. It remains open to understand the mechanism of the balanced space-time reduction that achieves the most accurate prediction. This question is in the same spirit as the bias-variance tradeoff in statistical learning theory [18].

A   Regularization

Regularization plays an important role in the estimation of the parameter in S-ROM, because the normal matrix in the regression is often ill-conditioned or even singular [33]. In our numerical tests, the normal matrix $\boldsymbol{A}_M$ in (2.17) is nonsingular but mostly ill-conditioned, with conditional numbers mostly in the range from 10^6 to 10^{12} when $M > 10$. Thus, the parameter estimation by solving the linear equation $\boldsymbol{A}_M \boldsymbol{c} = \mathbf{b}_M$ in (2.17) is an ill-posed inverse problem. Then, regularization is necessary to avoid the numerical and model errors in $\mathbf{b}_M$ being overly amplified. We regularize the problem using the Euclidean norm. That is, we minimize the regularized loss function

$$\mathcal{E}_\lambda(\boldsymbol{c}) = \mathcal{E}(\boldsymbol{c}) + \lambda \|\boldsymbol{c}\|^2, \quad \text{with } \mathcal{E}(\boldsymbol{c}) = \boldsymbol{c}^\top \boldsymbol{A}_M \boldsymbol{c} - 2\mathbf{b}_M^\top \boldsymbol{c}$$

where $\|\cdot\|$ is the Euclidean norm, and the $\mathcal{E}(\boldsymbol{c})$ comes from the loss function in (2.14). We select the optimal regularization parameter λ by the L-curve method (see [27], and we refer to [22, 44] for recent developments) as follows. First, we solve $c_\lambda = (\boldsymbol{A}_M + \lambda I)^{-1}\mathbf{b}_M$ by the minimum norm least square solution. Then, we find the regularization parameter that maximizes the curvature of the curve:

$$l(\lambda) = (x(\lambda), y(\lambda)) := (\log(\mathcal{E}(\boldsymbol{c}_\lambda)), \log(\|\boldsymbol{c}_\lambda\|)).$$

Recall that the curvature of l is $\kappa(\lambda) = \frac{x'y'' - x'y''}{(x'^2 + y'^2)^{3/2}}$. Thus, we compute the curvature by central difference approximation of these derivatives with a mesh for λ between the minimal and maximal eigenvalues of $\boldsymbol{A}_M$, and we select the λ with the maximal curvature.

B   Parameters in the S-ROM model

Table 1 and Figure 8 show the estimated $\widetilde{A}$ and $\widetilde{B}$ of S-ROM estimated from $\overline{M} = 1000$ trajectories. This S-ROM has $r = 10$ POD modes, and its time step is $\delta = 0.025$ (i.e., its time gap is Gap = 5). We present the parameter $\widetilde{A} \in \mathbb{R}^{r \times r}$ in Table 1 and present $\widetilde{B} \in \mathbb{R}^{r \times r \times r}$ in scaled images in Figure 8.

Table 1: Parameter $\widetilde{A}$ in S-ROM with $r = 10$ and $\delta = 0.025$, estimated from $\overline{M} = 1000$ trajectories.

| j	1	2	3	4	5	6	7	8	9	10
$\widetilde{A}(1,j)$	0.05	-0.03	-0.05	-0.11	0.14	-0.13	0.15	-0.16	0.17	-0.18
$\widetilde{A}(2,j)$	-0.08	0.01	0.08	0.11	-0.21	0.17	-0.25	0.25	-0.30	0.25
$\widetilde{A}(3,j)$	-0.10	0.00	-0.02	0.05	-0.18	0.22	-0.28	0.31	-0.40	0.34
$\widetilde{A}(4,j)$	-0.06	0.00	0.02	-0.15	-0.05	0.19	-0.09	0.26	-0.30	0.36
$\widetilde{A}(5,j)$	0.07	-0.00	-0.01	-0.06	-0.30	0.03	0.29	-0.14	0.36	-0.10
$\widetilde{A}(6,j)$	-0.04	0.00	0.00	0.08	0.01	-0.59	0.19	0.31	-0.02	0.36
$\widetilde{A}(7,j)$	0.07	-0.00	-0.00	-0.04	0.22	0.02	-0.98	0.32	0.45	-0.03
$\widetilde{A}(8,j)$	-0.06	0.00	0.00	0.07	-0.07	0.30	0.18	-1.57	0.65	0.80
$\widetilde{A}(9,j)$	0.07	0.00	-0.00	-0.01	0.15	0.03	0.45	0.47	-2.50	1.75
$\widetilde{A}(10,j)$	-0.06	0.00	0.00	0.04	-0.03	0.30	-0.03	0.80	1.24	-3.39
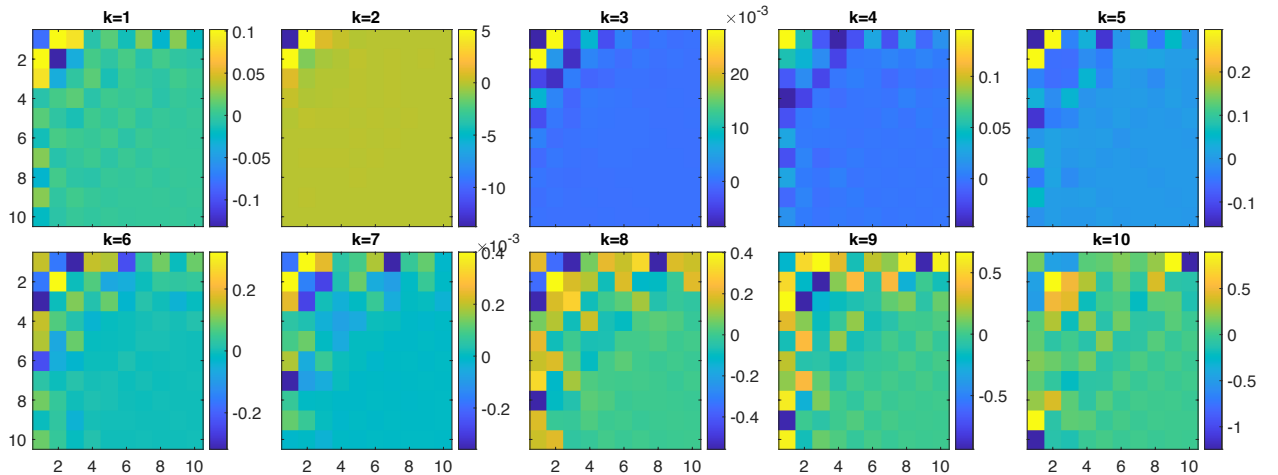
Figure 8: Parameter $\widetilde{B}$ in S-ROM with $r = 10$ and $\delta = 0.025$, estimated from $\overline{M} = 1000$ trajectories.

19

References

[1] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *Journal of Fluid Mechanics*, 192(-1):115, 1988. Eng4.

[2] L. C. Berselli, P. F. Fischer, T. Iliescu, and T. M. Özgökmen. Horizontal large eddy simulation of stratified mixing in a lock-exchange system. *J. Sci. Comput.*, 228(2):516–538, 2011.

[3] L. C. Berselli, T. Iliescu, and W. J. Layton. *Mathematics of Large Eddy Simulation of Turbulent Flows*. Scientific Computation. Springer-Verlag, Berlin, 2006.

[4] L. Berselli and C. Grisanti. On the consistency of the Rational large eddy simulation model. *Comput. Vis. Sci.*, 6:75–82, 2004.

[5] K. Carlberg, C. Bou-Mosleh, and C. Farhat. Efficient non-linear model reduction via a least-squares Petrov-Galerkin projection and compressive tensor approximations. *International Journal for Numerical Methods in Engineering*, 86(2):155–181, 2011. Eng4.

[6] M. D. Chekroun, H. Liu, and S. Wang. *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II*. Springer Briefs in Mathematics, Springer, New York, 2015.

[7] N. Chen and A. J. Majda. Conditional Gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification. *Entropy*, 20(7):509, 2018.

[8] N. Chen and D. Qi. A physics-informed data-driven algorithm for ensemble forecast of complex turbulent systems. *arXiv preprint arXiv:2204.08547*, 2022.

[9] N. Chen, Y. Li, and H. Liu. Conditional gaussian nonlinear system: A fast preconditioner and a cheap surrogate model for complex nonlinear systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(5):053122, 2022.

[10] H. Cho, D. Venturi, and G. E. Karniadakis. Statistical analysis and simulation of random shocks in stochastic Burgers equation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 470(2171):20140080–20140080, 2014.

[11] Y. Choi and K. Carlberg. Space–time least-squares Petrov-Galerkin projection for nonlinear model reduction. *SIAM Journal on Scientific Computing*, 41(1):A26–A58, 2019.

[12] A. J. Chorin. Averaging and renormalization for the Korteveg-deVries-Burgers equation. *Proc. Natl. Acad. Sci. USA*, 100(17):9674–9679, 2003.

[13] A. J. Chorin and O. H. Hald. *Stochastic Tools in Mathematics and Science*. Springer, New York, NY, 3rd edition, 2013.

[14] A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction with memory. *Physica D*, 166(3):239–257, 2002.

[15] A. J. Chorin, A. P. Kast, and R. Kupferman. Optimal prediction of underresolved dynamics. *Proc. Natl. Acad. Sci. USA*, 95:4094–4098, 1998.

[16] A. J. Chorin and F. Lu. Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics. *Proc. Natl. Acad. Sci. USA*, 112(32):9804–9809, 2015.

[17] A. J. Chorin. Conditional expectations and renormalization. *Multiscale Modeling & Simulation*, 1(1):105–118, 2003.

[18] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin of the American mathematical society*, 39(1):1–49, 2002.

[19] J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of multivariate analysis*, 12(1):136–154, 1982.

[20] J. Duan and B. T. Nadiga. Stochastic parameterization for large eddy simulation of geophysical flows. *Proc. Amer. Math. Soc.*, 135(4):1187–1196 (electronic), 2007.

[21] L. Feng, T. Gao, M. Dai, and J. Duan. Auto-sde: Learning effective reduced dynamics from data-driven stochastic dynamical systems. *arXiv preprint arXiv:2205.04151*, 2022.

[22] S. Gazzola, P. C. Hansen, and J. G. Nagy. Ir tools: a matlab package of iterative regularization methods and large-scale test problems. *Numerical Algorithms*, 81(3):773–811, 2019.

[23] B. Gershgorin, J. Harlim, and A. J. Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comput. Phys.*, 229(1):1–31, 2010.

[24] S. Giere, T. Iliescu, V. John, and D. Wells. SUPG reduced order models for convection-dominated convection-diffusion-reaction equations. *Comput. Methods Appl. Mech. Engrg.*, 289:454–474, 2015.

[25] M. Gunzburger, T. Iliescu, and M. Schneier. A Leray regularized ensemble-proper orthogonal decomposition method for parameterized convection-dominated flows. *IMA J. Numer. Anal.*, 40(2):886–913, 2020.

[26] M. Gunzburger, N. Jiang, and M. Schneier. An ensemble-proper orthogonal decomposition method for the nonstationary Navier-Stokes equations. *SIAM J. Numer. Anal.*, 55(1):286–304, 2017.

[27] P. C. Hansen. The L-curve and its use in the numerical treatment of inverse problems. In *in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering*, pages 119–142. WIT Press, 2000.

[28] J. Harlim. Model error in data assimilation. In C. Franzke and T. O'Kane, editors, *Nonlinear and Stochastic Climate Dynamics*, page in press. Cambridge University Press, Oxford, 2016.

[29] J. Harlim, S. W. Jiang, S. Liang, and H. Yang. Machine learning for prediction with missing dynamics. *Journal of Computational Physics*, page 109922, 2020.

[30] T. J. R. Hughes, G. R. Feijóo, L. Mazzei, and J.-B. Quincy. The variational multiscale method – a paradigm for computational mechanics. *Comput. Methods Appl. Mech. Engrg.*, 166(1):3–24, 1998.

[31] T. J. R. Hughes, L. Mazzei, and K. E. Jansen. Large eddy simulation and the variational multiscale method. *Comput. Vis. Sci.*, 3:47–59, 2000.

[32] T. J. R. Hughes, A. Oberai, and L. Mazzei. Large eddy simulation of turbulent channel flows by the variational multiscale method. *Phys. Fluids*, 13(6):1784–1799, 2001.

[33] T. Iliescu, H. Liu, and X. Xie. Regularized reduced order models for a stochastic Burgers equation. *Int. J. Numer. Anal. Mod.*, 15(4–5):594–607, 2018.

[34] T. Kadeethum, F. Ballarin, and N. Bouklas. Non-intrusive reduced order modeling of poroelasticity of heterogeneous media based on a discontinuous galerkin approximation. *arXiv preprint arXiv:2101.11810*, 2021.

[35] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé. Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, 28(3):985–1010, 2018.

[36] D. Kondrashov, M. D. Chekroun, and M. Ghil. Data-driven non-Markovian closure models. *Physica D*, 297:33–55, 2015.

[37] H. Lei, N. A. Baker, and X. Li. Data-driven parameterization of the generalized Langevin equation. *Proc. Natl. Acad. Sci. USA*, 113(50):14183–14188, 2016.

[38] M. E. Levine and A. M. Stuart. A framework for machine learning of model error in dynamical systems. *arXiv preprint arXiv:2107.06658*, 2021.

[39] X. H. Li, F. Lu, and F. X.-F. Ye. ISALT: Inference-based schemes adaptive to large time-stepping for locally Lipschitz ergodic systems. *Discrete & Continuous Dynamical Systems - S*, 15(4):747–771, 2022.

[40] K. K. Lin and F. Lu. Data-driven model reduction, wiener projections, and the koopman-mori-zwanzig formalism. *Journal of Computational Physics*, 424:109864, 2021.

[41] F. Lu, K. K. Lin, and A. J. Chorin. Data-based stochastic model reduction for the Kuramoto–Sivashinsky equation. *Physica D*, 340:46–57, 2017.

[42] F. Lu, X. Tu, and A. J. Chorin. Accounting for model error from unresolved scales in ensemble kalman filters by stochastic parameterization. *Mon. Wea. Rev.*, 145(9):3709–3723, 2017.

[43] F. Lu. Data-driven model reduction for stochastic Burgers equations. *Entropy*, 22(12):1360, Nov 2020.

[44] F. Lu, Q. Lang, and Q. An. Data adaptive RKHS Tikhonov regularization for learning kernels in operators. *arXiv preprint arXiv:2203.03791*, 2022.

[45] C. Ma, J. Wang, and W. E. Model reduction with memory and the machine learning of dynamical systems. *Commun. Comput. Phys.*, 25(4):947–962, 2018.

[46] A. J. Majda and J. Harlim. Physics constrained nonlinear regression models for time series. *Nonlinearity*, 26(1):201–217, 2013.

[47] A. J. Majda and B. Khouider. Stochastic and mesoscopic models for tropical convection. *Proc. Natl. Acad. Sci. USA*, 99:1123–1128, 2002.

[48] M. Mohebujjaman, L. G. Rebholz, and T. Iliescu. Physically-constrained data-driven correction for reduced order modeling of fluid flows. *Int. J. Num. Meth. Fluids*, 89(3):103–122, 2019.

[49] C. Mou, B. Koc, O. San, and T. Iliescu. Data-driven variational multiscale reduced order models. *Comput. Methods Appl. Mech. Engrg.*, 373:113470, 2020.

[50] C. Mou, H. Liu, D. R. Wells, and T. Iliescu. Data-driven correction reduced order models for the quasi-geostrophic equations: A numerical investigation. *Int. J. Comput. Fluid Dyn.*, pages 1–13, 2020.

[51] E. J. Parish and K. Duraisamy. A paradigm for data-driven predictive modeling using field inversion and machine learning. *J. Comput. Phys.*, 305:758–774, 2016. Eng0.

[52] O. San and T. Iliescu. Proper orthogonal decomposition closure models for fluid flows: Burgers equation. *Int. J. Numer. Anal. Mod., Series B*, 5(3):285–305, 2014.

[53] O. San, R. Maulik, and M. Ahmed. An artificial neural network framework for reduced order modeling of transient flows. *Communications in Nonlinear Science and Numerical Simulation*, 77:271–287, 2019.

[54] S. Sirisup and G. Karniadakis. A spectral viscosity method for correcting the long-term behavior of POD models. *J. Comput. Phys.*, 194(1):92–116, 2004.

[55] P. Stinis. Renormalized Mori-Zwanzig-reduced models for systems without scale separation. *Proc. Royal Soc. A*, 471(2176):20140446–20140446, 2015.

[56] P. Stinis. Mori-Zwanzig reduced models for uncertainty quantification II: Initial condition uncertainty. *arXiv:1212.6360 [math]*, 2012.

[57] Z. Wang, I. Akhtar, J. Borggaard, and T. Iliescu. Proper orthogonal decomposition closure models for turbulent flows: A numerical comparison. *Computer Methods in Applied Mechanics and Engineering*, 237-240:10–26, 2012. Eng3.

[58] K. Willcox. Unsteady flow sensing and estimation via the gappy proper orthogonal decomposition. *Computers & Fluids*, 35(2):208–226, 2006. Eng3.

[59] X. Xie, M. Mohebujjaman, L. G. Rebholz, and T. Iliescu. Data-driven filtered reduced order modeling of fluid flows. *SIAM J. Sci. Comput.*, 40(3):B834–B857, 2018.

[60] Y. Zhu, J. M. Dominy, and D. Venturi. On the estimation of the mori-zwanzig memory integral. *Journal of Mathematical Physics*, 59(10):103501, 2018.

[61] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, USA, 2001.