

Research



Cite this article: Babae H. 2019 An observation-driven time-dependent basis for a reduced description of transient stochastic systems. *Proc. R. Soc. A* **475**: 20190506. <http://dx.doi.org/10.1098/rspa.2019.0506>

Received: 6 August 2019

Accepted: 18 October 2019

Subject Areas:

computational physics, computational mathematics, computer modelling and simulation

Keywords:

observation driven, matrix free, transient dynamics, finite-time instabilities, time-dependent modes, dimension reduction

Author for correspondence:

H. Babae

e-mail: h.babae@pitt.edu

An observation-driven time-dependent basis for a reduced description of transient stochastic systems

H. Babae

Department of Mechanical Engineering and Materials Science, University of Pittsburgh, 3700 O'Hara Street, Pittsburgh, PA 15261, USA

 HB, 0000-0002-6318-2265

We present a variational principle for the extraction of a time-dependent orthonormal basis from random realizations of transient systems. The optimality condition of the variational principle leads to a closed-form evolution equation for the orthonormal basis and its coefficients. The extracted modes are associated with the most transient subspace of the system, and they provide a reduced description of the transient dynamics that may be used for reduced-order modelling, filtering and prediction. The presented method is matrix free; it relies only on the observables of the system and ignores any information about the underlying system. In that sense, the presented reduction is purely observation driven and may be applied to systems whose models are not known. The presented method has linear computational complexity and memory storage requirement with respect to the number of observables and the number of random realizations. Therefore, it may be used for a large number of observations and samples. The effectiveness of the proposed method is tested on four examples: (i) stochastic advection equation, (ii) stochastic Burgers equation, (iii) a reduced description of transient instability of Kuramoto–Sivashinsky, and (iv) a transient vertical jet governed by the incompressible Navier–Stokes equation. In these examples, we contrast the performance of the time-dependent basis versus static basis such as proper orthogonal decomposition, dynamic mode decomposition and polynomial chaos expansion.

1. Introduction

The ability to discover intrinsic structure from data has been called the *fourth paradigm of scientific discovery* [1]. Dimension reduction enables such discoveries and clearly belongs to the fourth paradigm. The goal of dimension reduction is to map high-dimensional data into a meaningful representation of reduced dimensionality. As a result, dimension reduction has been widely used in classification, visualization and compression of high-dimensional data [2]. In the scientific and engineering applications dimension reduction has been used to build reduced-order models to propagate uncertainty, control and optimization [3,4].

This paper focuses on dimension reduction of vector-valued observables $\mathbf{x}(t; \omega) \in \mathbb{R}^n$ generated by a time-dependent stochastic system where ω denotes a random observation. A linear reduction model seeks to approximate the target \mathbf{x} with

$$\mathbf{x} = \mathbf{U}\mathbf{z} + \epsilon, \quad (1.1)$$

where $\mathbf{U} \in \mathbb{R}^{n \times r}$ is a deterministic matrix whose columns can be regarded as *modes* and $\mathbf{z} \in \mathbb{R}^r$ are the reduced stochastic variable, where r is the reduction size and ϵ is the reduction error. Both \mathbf{U} and \mathbf{z} are in general time dependent. In view of equation (1.1), polynomial chaos expansion (PCE) can be seen as a special case where \mathbf{z} is represented by PCE—a *static* (i.e. time independent) basis—while $\mathbf{U} = \mathbf{U}(t)$ is time dependent. PCE has reached widespread popularity especially for steady-state and parabolic stochastic partial differential equations (SPDEs) [5]. However, for time-dependent dynamical systems with transient instabilities the number of PCE terms required for a given level of accuracy must rapidly increase [6]. Similarly, it was shown in [7] that the reduction size of PCE must increase in time to maintain a given accuracy. On the other hand, data-driven methods such as proper orthogonal decomposition (POD) and dynamic mode decomposition (DMD) belong to a class of methods where the modes \mathbf{U} are static and they are computed from numerical/experimental observations of the dynamical system and $\mathbf{z}(t; \omega)$ is often obtained by projection of the data to these modes [8]. For highly transient systems, DMD and POD require a large number of modes to capture transient time behaviour [9,10]. This is partly because both DMD and POD decompositions seek to obtain an optimal reduction of the data in a *time-average* sense. In the case of DMD a best-fit linear dynamical system matrix \mathbf{A} is sought so that $\|\mathbf{x}^{k+1} - \mathbf{A}\mathbf{x}^k\|_2$ is minimized across all snapshots, where k shows the time snapshots. Similarly, the POD is an optimal basis for all snapshots. However, transient instabilities, for example, by definition have a short lifetime. The effect of these transient instabilities may be lost in a reduction based on time average.

More recently dynamically orthogonal (DO) [11] and bi-orthogonal (BO) [12] expansions have proven promising in tackling time-dependent SPDEs. In these reduction techniques both $\mathbf{U}(t)$ and $\mathbf{z}(t; \omega)$ are *dynamic* and they evolve with time. In a DO/BO framework, exact evolution equations for the evolution of $\mathbf{U}(t)$ and $\mathbf{z}(t; \omega)$ are derived. In the case of SPDEs, the evolution equation for $\mathbf{U}(t)$ is a system of deterministic PDEs that are solved fully coupled with a system of stochastic ordinary differential equations (SODEs) and an evolution equation for the mean equation. Although DO and BO algorithms lead to different evolution equations for the stochastic coefficient and the spatial basis, it is shown in [13,14] that both of these methods are equivalent, i.e. the BO and DO reductions span the same subspaces. In deterministic applications, time-dependent bases have been used for dimension reduction of deterministic Schrödinger equations by the multi-configuration time-dependent Hartree (MCTDH) method [15,16], and later by Koch & Lubich [17]. Recently, optimally time-dependent (OTD) reduction was introduced [18], in which an evolution equation for a set of orthonormal modes is found that captures the most instantaneously transient subspace of a dynamical system [19].

The motivation for this paper is to obtain a reduced description of transient dynamics that relies only on random observations—i.e. an equation-free reduction. This is in contrast to model-dependent reduction techniques, for example DO and BO methodologies, which require high-dimensional ODEs/PDEs to be solved to determine the time-dependent modes. When stochastic governing equations are complex, the derivation of evolution equations for the DO/BO

coefficients can become very difficult, if not impossible. The reliance on the model in these techniques impedes their application in complex multi-physics/multi-component systems. For example, it is still unknown how to determine the boundary conditions for DO/BO modes in a case of a single-physics/single-domain PDE subject to random boundary conditions [20]. Moreover, the evolution equation for the DO/BO modes is tied to the equation that governs the dynamical system at hand, which would require a case-by-case implementation. On the other hand, in the era of extreme-scale computing, concurrent high-fidelity numerical simulations are expected to be routinely possible [21]. These simulations generate an unprecedented amount of streaming data, whose low-rank structure can be exploited using a data-driven approach. The low-rank structure can then be used for diagnostic purposes or to build reduced-order models.

The objective of this paper is to find a reduced description of the transient dynamics that is observation driven and discards any information about the underlying model. To this end, we present a variational principle for the determination of a set of time-dependent orthonormal modes that capture the low-dimensional structure from observations of the stochastic system. We note that the presented method is used to perform dimension reduction and not reduced-order modelling. The structure of the paper is as follows. In §2, we present the variational principle for the extraction of the time-dependent modes and their coefficients. In §3, we present four demonstration cases: (i) the stochastic advection equation, (ii) the stochastic Burgers equation, (iii) the Kuramoto–Sivashinsky equation, and (iv) a transient vertical jet in quiescent background flow governed by incompressible Navier–Stokes equations. In §4, we present the conclusions of our work.

2. Methodology

(a) Set-up

We consider a stochastic dynamical system in the form of

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t; \boldsymbol{\xi}), \quad (2.1)$$

where \mathbf{x} is the state space and the above dynamical system is finite dimensional, i.e. $\mathbf{x} \in \mathbb{R}^m$, or infinite dimensional, expressed, for example, by SPDEs, and $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_d) \in \mathbb{R}^d$ is a d -dimensional random space with the joint probability given by $\rho(\boldsymbol{\xi})$. The dynamical system may not be known and the proposed reduction relies only on vector-valued observations of the above system. We base our formulation on continuous-time observations, represented by matrix $\mathbf{T}(t) \in \mathbb{R}^{n \times s}$, whose columns contain vector-valued observations of the stochastic system at time t ,

$$\mathbf{T}(t) = \left[\mathbf{x}_1(t) \mid \mathbf{x}_2(t) \mid \dots \mid \mathbf{x}_s(t) \right]_{n \times s}.$$

Each column of this matrix is obtained by first drawing a random sample of $\boldsymbol{\xi}$ with joint probability distribution of $\rho(\boldsymbol{\xi})$ and solving the forward model given by equation (2.1). Before we proceed we introduce the sample mean operator,

$$\mathbb{E}[\mathbf{T}(t)] = \mathbf{T}(t)\mathbf{w}_{\boldsymbol{\xi}},$$

where $\mathbf{w}_{\boldsymbol{\xi}} = [\mathbf{w}_{\boldsymbol{\xi}_1}, \mathbf{w}_{\boldsymbol{\xi}_2}, \dots, \mathbf{w}_{\boldsymbol{\xi}_s}]^T \in \mathbb{R}^{s \times 1}$ represents the sample weights. For example, for standard Monte Carlo samples, $\mathbf{w}_{\boldsymbol{\xi}} = \mathbf{1}/s$, where $\mathbf{1} \in \mathbb{R}^{s \times 1}$ is a column vector with all elements equal to 1. For cases where controlled observations of system (2.1) is possible, the probabilistic collocation strategy may be used for sampling. In that case $\mathbf{w}_{\boldsymbol{\xi}}$ represents the associated probabilistic collocation weights. Similarly, $\mathbf{w}_{\mathbf{x}} = [\mathbf{w}_{\mathbf{x}_1}, \mathbf{w}_{\mathbf{x}_2}, \dots, \mathbf{w}_{\mathbf{x}_n}] \in \mathbb{R}^{n \times 1}$ represents the associated weights in the state space. For applications governed by SPDEs, $\mathbf{w}_{\mathbf{x}}$ may be chosen as space integration weights (e.g. quadrature weights), such that $\mathbf{w}_{\mathbf{x}}^T \mathbf{u}$ approximates $\int_{\Omega} u d\Omega$, where Ω is the spatial domain and \mathbf{u} contains vector-valued observations of spatial function u . We assume that $\mathbf{T}(t)$ has

zero mean, i.e. $\mathbb{E}[\mathbf{T}(t)] = 0$. This is accomplished by subtracting the ensemble mean from all the realizations. Let us define the inner product in both state and random spaces,

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbf{u}_i^T \mathbf{W}_x \mathbf{u}_j$$

and

$$\mathbb{E}[\mathbf{y}_i^T \mathbf{y}_j] = \mathbf{y}_i^T \mathbf{W}_\xi \mathbf{y}_j,$$

where $\mathbf{W}_x = \text{diag}(\mathbf{w}_x)$ and $\mathbf{W}_\xi = \text{diag}(\mathbf{w}_\xi)$ are matrices of size $n \times n$ and $s \times s$, respectively. Therefore, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle$ is an inner product of \mathbf{u}_i and \mathbf{u}_j in the state space, and $\mathbb{E}[\mathbf{y}_i^T \mathbf{y}_j]$ is an inner product of \mathbf{y}_i and \mathbf{y}_j in the random space. Our goal is to build a low-rank approximation that can capture transient behaviour in the stochastic data $\mathbf{T}(t)$. To this end, we seek a reduction in the form of

$$\mathbf{T}(t) = \mathbf{U}(t)\mathbf{Y}(t)^T + \mathbf{E}(t), \quad (2.2)$$

with

$$\mathbf{Y}(t) = \left[\mathbf{y}_1(t) \mid \mathbf{y}_2(t) \mid \cdots \mid \mathbf{y}_r(t) \right]_{s \times r}$$

and

$$\mathbf{U}(t) = \left[\mathbf{u}_1(t) \mid \mathbf{u}_2(t) \mid \cdots \mid \mathbf{u}_r(t) \right]_{n \times r},$$

where $\mathbf{Y}(t) \in \mathbb{R}^{s \times r}$ and $\mathbf{Y}(t) = [\mathbf{y}_1(t) \mid \mathbf{y}_2(t) \mid \dots \mid \mathbf{y}_r(t)]$ and $\mathbf{U}(t) \in \mathbb{R}^{n \times r}$ is a set of orthonormal modes, i.e. $\langle \mathbf{u}_i(t), \mathbf{u}_j(t) \rangle = \delta_{ij}$ and \mathbf{E} is the reduction error. In the proposed form $\mathbf{U}(t)$ represents a time-dependent low-rank subspace, which we will refer to as the *dynamic basis* (DB), and $\mathbf{Y}(t)$ are the stochastic coefficients of the DB; we will refer to $\mathbf{Y}(t)$ as the *stochastic coefficients*. In the following we present a variational principle that results in closed-form evolution equations for the DB $\mathbf{U}(t)$ and the stochastic coefficients $\mathbf{Y}(t)$.

(b) Variational principle

In this section we present a variational principle from which the evolution equations of the low-rank approximation are derived. Our goal is to build a low-rank approximation that can capture transient data. To this end, we first introduce a weighted Frobenius norm,

$$\left\| \mathbb{E}[\mathbf{T}]^2 \right\|_F^{1/2} = \left(\sum_{i=1}^n \sum_{j=1}^s \mathbf{w}_{x_i} \mathbf{w}_{\xi_j} \mathbf{T}_{i,j}^2 \right)^{1/2}, \quad (2.3)$$

where $\mathbf{T}_{i,j}$ is the entry of matrix \mathbf{T} at the i th row and the j th column. Now we consider the variational principle in the form of

$$\mathcal{F}(\dot{\mathbf{U}}(t), \dot{\mathbf{Y}}(t)) = \left\| \mathbb{E} \left[\frac{d(\mathbf{U}(t)\mathbf{Y}(t)^T)}{dt} - \dot{\mathbf{T}}(t) \right] \right\|_F^2, \quad (2.4)$$

subject to the orthonormality condition of the modes, i.e. $\langle \mathbf{u}_i(t), \mathbf{u}_j(t) \rangle = \delta_{ij}$. The variational principle (2.4) seeks to minimize the difference between the rate of change of all observations and the DB reduction, and the control parameters are $\dot{\mathbf{U}}(t)$ and $\dot{\mathbf{Y}}(t)$. To incorporate the orthonormality condition, we observe that

$$\langle \dot{\mathbf{u}}_i(t), \mathbf{u}_j(t) \rangle + \langle \mathbf{u}_i(t), \dot{\mathbf{u}}_j(t) \rangle = 0, \quad (2.5)$$

and we denote $\phi_{ij}(t) := \langle \mathbf{u}_i(t), \dot{\mathbf{u}}_j(t) \rangle$. It is clear from the above equation that $\phi_{ij}(t)$ is a skew-symmetric matrix, i.e. $\phi_{ij}(t) = -\phi_{ji}(t)$. We incorporate the orthonormality constraint into the variational principle via Lagrange multipliers. This follows:

$$\mathcal{G}(\dot{\mathbf{U}}(t), \dot{\mathbf{Y}}(t)) = \left\| \mathbb{E} \left[\sum_{i=1}^r \frac{d(\mathbf{u}_i(t)\mathbf{y}_i(t)^T)}{dt} - \dot{\mathbf{T}}(t) \right] \right\|_F^2 + \sum_{i,j=1}^r \lambda_{ij}(t) (\langle \mathbf{u}_i(t), \dot{\mathbf{u}}_j(t) \rangle - \phi_{ij}(t)), \quad (2.6)$$

where $\lambda_{ij}(t)$, $i, j = 1, \dots, r$ are the Lagrange multipliers. In appendix A, we show that the first-order optimality condition of the above functional leads to the closed-form evolution equation of $\mathbf{U}(t)$

and $\mathbf{Y}(t)$ as given by

$$\dot{\mathbf{U}}(t) = \prod_{\perp \mathbf{U}(t)} \mathbb{E}[\dot{\mathbf{T}}(t)\mathbf{Y}(t)]\mathbf{C}(t)^{-1} + \mathbf{U}(t)\boldsymbol{\phi}(t) \quad (2.7)$$

and

$$\dot{\mathbf{Y}}(t) = \langle \mathbf{T}(t)^T, \mathbf{U}(t) \rangle + \mathbf{Y}(t)\boldsymbol{\phi}(t), \quad (2.8)$$

where $\mathbf{C}(t) = \mathbb{E}[\mathbf{Y}(t)^T\mathbf{Y}(t)]$ is the reduced covariance matrix ($\mathbf{C}(t) \in \mathbb{R}^{r \times r}$) and $\prod_{\perp \mathbf{U}}$ is the orthogonal projection matrix to the space spanned by the complement of $\mathbf{U}(t)$, i.e. $\prod_{\perp \mathbf{U}} \mathbf{V} = \mathbf{V} - \mathbf{U}\mathbf{U}^T\mathbf{W}_x\mathbf{V} = (\mathbf{I} - \mathbf{U}\mathbf{U}^T\mathbf{W}_x)\mathbf{V}$.

Solving equations (2.7) and (2.8) requires determination of $\boldsymbol{\phi}(t)$. We show in appendix B that for any skew-symmetric matrix $\boldsymbol{\phi}(t)$ the above reduction remains equivalent for all times. Equivalent reductions can be defined as follows: the two reduced subspaces $\{\mathbf{Y}(t), \mathbf{U}(t)\}$ and $\{\tilde{\mathbf{Y}}(t), \tilde{\mathbf{U}}(t)\}$, where $\mathbf{U}(t), \tilde{\mathbf{U}}(t) \in \mathbb{R}^{r \times r}$ and $\mathbf{Y}(t), \tilde{\mathbf{Y}}(t) \in \mathbb{R}^{s \times r}$, are equivalent if there exists a transformation matrix $\mathbf{R}(t) \in \mathbb{R}^{r \times r}$ such that $\mathbf{U}(t) = \tilde{\mathbf{U}}(t)\mathbf{R}(t)$ and $\mathbf{Y}(t) = \tilde{\mathbf{Y}}(t)\mathbf{R}(t)$, where $\mathbf{R}(t)$ is an orthogonal rotation matrix, i.e. $\mathbf{R}(t)^T\mathbf{R}(t) = \mathbf{I}$. Therefore, changing $\boldsymbol{\phi}(t)$ does not affect any of the subspaces that $\mathbf{U}(t)$ and $\mathbf{Y}(t)$ span, and it will only lead to an in-subspace rotation. In the rest of this paper, we take the simplest case of $\boldsymbol{\phi}(t) \equiv \mathbf{0}$. In that case the evolution of the DB reduces to

$$\dot{\mathbf{U}}(t) = \prod_{\perp \mathbf{U}(t)} \mathbb{E}[\dot{\mathbf{T}}(t)\mathbf{Y}(t)]\mathbf{C}(t)^{-1} \quad (2.9)$$

and

$$\dot{\mathbf{Y}}(t) = \langle \dot{\mathbf{T}}(t)^T, \mathbf{U}(t) \rangle. \quad (2.10)$$

To compute the right-hand side of equation (2.9), $\mathbf{U}(t)\mathbf{U}(t)^T$ may not be computed or stored explicitly. This can be done by observing that

$$\begin{aligned} \prod_{\perp \mathbf{U}(t)} \mathbb{E}[\dot{\mathbf{T}}(t)\mathbf{Y}(t)]\mathbf{C}(t)^{-1} &= (\mathbf{I} - \mathbf{U}(t)\mathbf{U}(t)^T\mathbf{W}_x) \dot{\mathbf{T}}(t)\mathbf{W}_\xi\mathbf{Y}(t)\mathbf{C}(t)^{-1} \\ &= (\dot{\mathbf{T}}(t)\mathbf{W}_\xi\mathbf{Y}(t) - \mathbf{U}(t)\dot{\mathbf{T}}_r(t))\mathbf{C}(t)^{-1}, \end{aligned}$$

where $\dot{\mathbf{T}}_r = \mathbf{U}^T\mathbf{W}_x\dot{\mathbf{T}}\mathbf{W}_\xi\mathbf{Y}$ and $\dot{\mathbf{T}}_r \in \mathbb{R}^{r \times r}$. Therefore, the evolution equations for the DB and the reduced manifold are given by

$$\dot{\mathbf{U}}(t) = (\dot{\mathbf{T}}(t)\mathbf{W}_\xi\mathbf{Y}(t) - \mathbf{U}(t)\dot{\mathbf{T}}_r(t))\mathbf{C}(t)^{-1} \quad (2.11)$$

and

$$\dot{\mathbf{Y}}(t) = \dot{\mathbf{T}}(t)^T\mathbf{W}_x\mathbf{U}(t). \quad (2.12)$$

Besides the simplicity of the above evolution equations, the choice of $\boldsymbol{\phi}(t) = \mathbf{0}$ also leads to an insightful interpretation of the DB reduction, namely the update of the DB $\dot{\mathbf{U}}(t)$ is always orthogonal to the space spanned by $\mathbf{U}(t)$, i.e. $\dot{\mathbf{U}}(t) \perp \mathbf{U}(t)$, and this update is driven by the projection of $\mathbb{E}[\dot{\mathbf{T}}(t)\mathbf{Y}(t)]$ onto the orthogonal complement of $\mathbf{U}(t)$ (see equation (2.11)). This means that if $\mathbb{E}[\dot{\mathbf{T}}(t)\mathbf{Y}(t)]$ is in the span of $\mathbf{U}(t)$, then

$$\dot{\mathbf{U}}(t) = \prod_{\perp \mathbf{U}(t)} \mathbb{E}[\dot{\mathbf{T}}(t)\mathbf{Y}(t)]\mathbf{C}(t)^{-1} = \mathbf{0}.$$

However, when $\mathbb{E}[\dot{\mathbf{T}}(t)\mathbf{Y}(t)]$ is not in the span of $\mathbf{U}(t)$, it results in $\dot{\mathbf{U}}(t) \neq \mathbf{0}$. This built-in adaptivity of the DB allows $\mathbf{U}(t)$ to 'follow' the data instantaneously. The $\mathbf{Y}(t)$ coefficients are updated by projection of $\dot{\mathbf{T}}(t)$ onto the space spanned by $\mathbf{U}(t)$ (see equation (2.12)). Therefore, the evolution equation of $\mathbf{Y}(t)$ captures the evolution of $\mathbf{T}(t)$ within the subspace $\mathbf{U}(t)$.

The model-driven counterpart of equations (2.9) and (2.10) is the DO decomposition [11]. In the DO decomposition of SPDEs, a system of deterministic PDEs are derived for the evolution of the time-dependent orthonormal modes and a system of SODEs are derived for the evolution of the stochastic coefficients. These two systems coupled with the mean evolution equation must be solved simultaneously. The DO decomposition does not require observations of the system, since in the DO formulation evolution equations for $\mathbf{U}(t)$ and $\mathbf{Y}(t)$ are derived from the model instead

of the observation. DO is intrusive and obtaining the time-dependent basis requires solving coupled ODE/PDE systems, which must be derived and solved for the application at hand. On the other hand, equations (2.9) and (2.10) are not PDEs and they only involve simple matrix vector operations irrespective of the data generator, i.e. the dynamical system. In other words, the DB reduction is agnostic with regard to the underlying model that generates the observation and observations may come from a system whose model is not known. As we demonstrate in §3c, even if the DB observations are chosen to be the solution of the entire computational domain, DO and DB are, in general, not equivalent, i.e. DO and DB bases do not span the same subspace. This is mainly because the DB samples, i.e. the columns of $\mathbf{T}(t)$, are the solution of the state equation and therefore they represent the true solution, while in the DO decomposition the samples are generated in the reduced subspace and therefore they are affected by the reduced-order modelling error. Finally, the DB reduction can be applied to nonlinear observables of the dynamical system or partial observations. See §3d for the demonstration of the latter feature of DB.

(c) Workflow and computational complexity

The DB reduction is governed by time-continuous evolution equations (2.11) and (2.12). For time-discrete data these equations must be discretized in time. Here we summarize all the steps required for solving equations (2.11) and (2.12).

Step 1. Given the time-discrete realizations in the form of $\tilde{\mathbf{T}}_k, k = 0, 1, 2, \dots, K$ we compute zero-mean by subtracting the instantaneous sample mean from the data: $\mathbf{T}_k = \tilde{\mathbf{T}}_k - \mathbb{E}[\tilde{\mathbf{T}}_k] \mathbf{1}_{1 \times s}$, where $\mathbf{1}_{1 \times s}$ is a row vector of size s whose all elements are equal to 1.

Step 2. We compute the initial condition for \mathbf{Y}_0 and \mathbf{U}_0 by performing Karhunen–Loève (KL) decomposition of the \mathbf{T}_0 at $t = 0$. This follows:

$$\mathbf{T}_0 = \sum_{i=1}^r \sqrt{\lambda_i} \mathbf{v}_i \mathbf{z}_i^T + \mathbf{E}_0,$$

where vectors $\mathbf{v}_i \in \mathbb{R}^n$ are an orthonormal basis, $\mathbf{z}_i \in \mathbb{R}^s$ are the stochastic coefficients, λ_i are their corresponding eigenvalues and \mathbf{E}_0 is the reduction error. The initial conditions for $\mathbf{U}(t)$ and the stochastic coefficients $\mathbf{Y}(t)$ are then given by: $\mathbf{u}_{i_0} = \mathbf{v}_i$ and $\mathbf{y}_{i_0} = \sqrt{\lambda_i} \mathbf{z}_i$ for $i = 1, 2, \dots, r$.

Step 3. The time derivative of the observations $\dot{\mathbf{T}}(t)$ can be estimated via finite difference schemes. For example, the fourth-order central finite difference approximation takes the form of

$$\dot{\mathbf{T}}_k = \frac{-\mathbf{T}_{k+2} + 8\mathbf{T}_{k+1} - 8\mathbf{T}_{k-1} + \mathbf{T}_{k-2}}{12\Delta t} + \mathcal{O}(\Delta t^4).$$

For the first two time steps one-sided finite-difference schemes may be used.

Step 4. Evolve the DB and the stochastic coefficients to the next time step k using a time-integration scheme. We use the fourth-order Runge–Kutta scheme. A comparison with a lower-order time integration scheme is given in the last demonstration case.

Step 5. Perform a Gram–Schmidt orthonormalization on the DB $\mathbf{U}(t)$. Although the DB remains orthonormal by construction, the numerical orthonormality of $\mathbf{U}(t)$ may not be preserved owing to the nonlinearity of the evolution equations (2.11) and (2.12). An enforcement of orthonormality of $\mathbf{U}(t)$ resolves this issue. Now steps 3–5 can be repeated in a time-loop until completion of all time steps.

A quick inspection of the evolution equations (2.11) and (2.12) reveals two favourable properties with regard to the computational cost and memory storage requirement of the DB reduction. (i) The computational complexity of solving the evolution equations (2.11) and (2.12) is $\mathcal{O}(rms)$, i.e. linear computational complexity with respect to the reduction size r , the size of the observations in state space n and the sample size s . We note that given that r is often a small number, i.e. $r \ll n$ and $r \ll s$, the cost of inverting the covariance matrix $\mathbf{C}(t)$ is often negligible. (ii) The memory storage requirement is also linear with respect to n, r and s . This is because at each time instant only computation of $\dot{\mathbf{T}}(t)$ is required, which can be computed from a few snapshots of $\mathbf{T}(t)$. This allows for loading matrix $\mathbf{T}(t)$ in small time batches. The linear computational

and memory complexity of the DB reduction may be exploited for applications with very large datasets.

(d) Mode ranking within the subspace

In this section we present a ranking criterion within the dynamic subspace $\mathbf{U}(t)$. This can be accomplished by using the reduced covariance matrix $\mathbf{C}(t)$, which, in general, is a full matrix and this implies that the $\mathbf{Y}(t)$ coefficients are in general correlated. The uncorrelated $\mathbf{Y}(t)$ coefficients may be computed via the eigen-decomposition of the covariance matrix $\mathbf{C}(t)$. To this end, let

$$\mathbf{C}(t)\boldsymbol{\psi}(t) = \boldsymbol{\psi}(t)\mathbf{A}(t) \quad (2.13)$$

be the eigenvalue decomposition of matrix $\mathbf{C}(t)$, where $\boldsymbol{\psi}(t) \in \mathbb{R}^{r \times r}$ is a matrix of eigenvectors of $\mathbf{C}(t)$ and $\mathbf{A}(t) = \text{diag}(\lambda_1(t), \lambda_2(t), \dots, \lambda_r(t))$ is a diagonal matrix of the eigenvalues of matrix $\mathbf{C}(t)$. The uncorrelated $\mathbf{Y}(t)$ coefficients and the DB associated with these uncorrelated coefficients, denoted by $\hat{\mathbf{Y}}(t)$ and $\hat{\mathbf{U}}(t)$, are obtained by an in-subspace rotation,

$$\hat{\mathbf{Y}}(t) = \mathbf{Y}(t)\boldsymbol{\psi}(t), \quad \hat{\mathbf{U}}(t) = \mathbf{U}(t)\boldsymbol{\psi}(t).$$

It is easy to verify that $\hat{\mathbf{Y}}(t)$ coefficients are uncorrelated, i.e. $\mathbb{E}[\hat{\mathbf{Y}}(t)\hat{\mathbf{Y}}(t)^T] = \mathbf{A}(t)$. We adopt a variance-based ranking for $\{\hat{\mathbf{Y}}(t), \hat{\mathbf{U}}(t)\}$. Since matrix $\mathbf{C}(t)$ is positive semidefinite, without loss of generality, let $\lambda_1(t) \geq \lambda_2(t) \geq \dots \geq \lambda_r(t) \geq 0$, where $\lambda_i(t)$ is the variance of $\hat{\mathbf{y}}_i(t)$.

Since $\{\hat{\mathbf{Y}}(t), \hat{\mathbf{U}}(t)\}$ are obtained by an in-subspace rotation of $\{\mathbf{Y}(t), \mathbf{U}(t)\}$, they are equivalent representations of the low-rank dynamics, i.e. $\hat{\mathbf{U}}(t)\hat{\mathbf{Y}}(t)^T = \mathbf{U}(t)\boldsymbol{\psi}(t)\boldsymbol{\psi}(t)^T\mathbf{Y}(t)^T = \mathbf{U}(t)\mathbf{Y}(t)^T$, where we have used $\boldsymbol{\psi}(t)\boldsymbol{\psi}(t)^T = \mathbf{I}$, since eigenvectors of a symmetric matrix ($\mathbf{C}(t)$) are orthonormal. In all the demonstration cases in this paper, we use the ranked uncorrelated representation of the DB for the analysis of the results. For the sake of simplicity in the notation, we use $\{\mathbf{Y}(t), \mathbf{U}(t)\}$ to refer to the ranked uncorrelated representation instead of $\{\hat{\mathbf{Y}}(t), \hat{\mathbf{U}}(t)\}$, noting that both of these representations are equivalent.

3. Demonstration cases

(a) Stochastic advection equation

As the first example, we apply DB reduction to a one-dimensional stochastic advection equation. As we will show in this section, the stochastic advection equation chosen here is exactly two dimensional, when expressed versus the DB. It also has an exact analytical solution for the DB \mathbf{U} and the stochastic coefficients \mathbf{Y} , which enables an appropriate verification of the DB algorithm. The stochastic advection equation is given by

$$\frac{\partial u}{\partial t} + V(\xi)\frac{\partial u}{\partial x} = 0, \quad x \in [-1, 1], \quad t \in [0, T_f], \quad (3.1)$$

subject to periodic boundary conditions at the ends of the spatial domain $x \pm 1$ and the initial condition of $u_0(x; \xi)$. In the above equation T_f is the final time. The solution to the above equation is a stochastic random field in the form of $u(x, t; \xi)$. The stochastic advection speed $V(\xi)$ and the deterministic initial condition are chosen to be

$$V(\xi) = \bar{v} + \sigma\xi, \quad u_0(x; \xi) = \sin \pi x,$$

where ξ is a one-dimensional uniform random variable defined on $[-1, 1]$, σ is a constant and \bar{v} is the mean of the transport velocity. The exact solution of equation (3.1) is

$$u(x, t; \xi) = \sin \pi(x - (\bar{v} + \sigma\xi)t),$$

and its mean and variance are

$$\mathbb{E}[u(x, t; \xi)] = \frac{1}{\sigma\pi t} \sin \pi(x - \bar{v}t) \sin \sigma\pi t$$

and

$$\mathbb{E}[u(x, t; \xi)^2] - \mathbb{E}[u(x, t; \xi)]^2 = \frac{1}{2} - \frac{1}{4\sigma\pi t} \cos 2\pi(x - \bar{v}t) \sin 2\sigma\pi t - \frac{1}{\sigma^2\pi^2 t^2} \sin^2 \pi(x - \bar{v}t) \sin^2 \sigma\pi t.$$

The expectation operator is given by $\mathbb{E}[u(x, t; \xi)] = \int_{-1}^1 u(x, t; \xi) \rho(\xi) d\xi$, where $\rho(\xi) = 1/2$ is the probability density function. The total variance is equal to

$$\Sigma(t) = \int_{-1}^1 (\mathbb{E}[u(x, t; \xi)^2] - \mathbb{E}[u(x, t; \xi)]^2) dx = 1 - \frac{\sin^2 \sigma\pi t}{(\sigma\pi t)^2}.$$

The mean-subtracted stochastic field can be expressed with the following two dimensional expansion:

$$u(x, t; \xi) - \mathbb{E}[u(x, t; \xi)] = u_1 y_1 + u_2 y_2, \quad (3.2)$$

where

$$\begin{aligned} u_1(x, t) &= \sin \pi(x - \bar{v}t), & y_1(\xi, t) &= \cos \pi \sigma \xi t - \frac{1}{\sigma\pi t} \sin \pi \sigma t, \\ u_2(x, t) &= \cos \pi(x - \bar{v}t), & y_2(\xi, t) &= \sin \pi \sigma \xi t. \end{aligned}$$

It is easy to verify that the two-dimensional expansion of the above random field satisfies the DB orthonormality condition in the continuous form, i.e.

$$\int_{-1}^1 u_i(x, t) u_j(x, t) dx = \delta_{ij}, \quad i, j = 1, 2.$$

The expansion of the stochastic field as expressed above in equation (3.2) is in BO form since $\mathbb{E}[y_i y_j] = 0, i \neq j$.

The details of solving the evolution equations of the DB $\mathbf{U}(t)$ and the stochastic coefficients $\mathbf{Y}(t)$ are as follows. The matrix-valued random observations $\mathbf{T}(t)$ are generated by solving the stochastic advection equation (3.1) at $s = 65$ collocation points. These sampling points are chosen to be the roots of the Legendre polynomial of order 65, defined on the interval of $\xi \in [-1, 1]$. The 65 realizations of equation (3.1) are generated using a deterministic solver that uses Fourier modes of order $n = 128$ for space discretization and fourth-order Runge–Kutta with $\Delta t = 0.001$ for time advancement. The final time is set to $T_f = 10$ for all realizations.

We first performed dimension reduction of size $r = 2$. In figure 1a, the total variance $\Sigma(t)$ obtained from the reduction is compared against the exact variance. The total variance of the low-rank approximation is obtained by $\Sigma(t) = \sum_{i=1}^r \lambda_i(t)$, where $\lambda_i(t)$ are the eigenvalues of the covariance matrix \mathbf{C} . We also solved the stochastic advection equation (3.1) using the probabilistic collocation method (PCM), in which the stochastic solution is approximated as

$$u(x, t; \xi) - \mathbb{E}[u(x, t; \xi)] = \sum_{i=1}^r \phi_i(\xi) v_i(x, t) + e(x, t; \xi),$$

where $\phi_i(\xi)$ are the Legendre chaos polynomials of order i , $v_i(x, t)$ are space–time coefficients and $e(x, t; \xi)$ is the expansion error. In figure 1a, the total variance obtained from different reduction sizes r of the PCM are also shown, where it can be seen that as the size of the PCM expansion increases the PCM variance follows the exact solution for a longer time, but eventually the PCM error increases and even a 20-dimensional PCM reduction under-performs the DB reduction with $r = 2$. For the long-term error analysis for this problem using the PCE, the reader may refer to [7]. We note that in the PCM reduction the stochastic basis is fixed *a priori*, i.e. the polynomial chaos $\phi_i(\xi)$, while in the DB reduction the stochastic coefficients $y_i(\xi, t)$ are time dependent and intuitively this enables the stochastic coefficient to adapt instantaneously to the data according to the variational principle (2.6).

In figure 1b the covariance eigenvalues of the DB reduction with $r = 3$ are compared against the exact solution. Since the solution is two dimensional when expressed versus the DB, the third mode must remain inactive. As shown in figure 1b, this is confirmed numerically as the third eigenvalue does not pick up energy, while the first two eigenvalues are in excellent

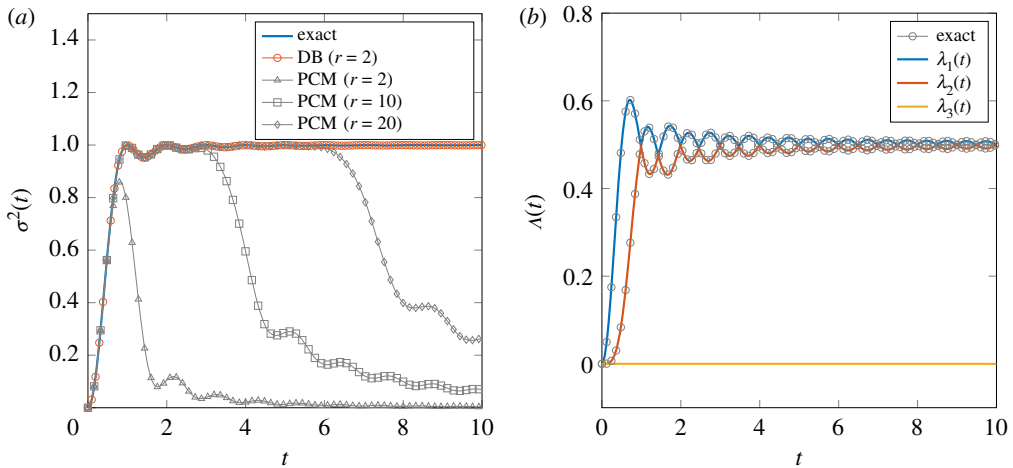


Figure 1. Stochastic advection equation. (a) Comparison of the total variance of the dynamic basis (DB) reduction and polynomial chaos method (PCM) with the exact solution. (b) Comparison of the eigenvalues of the covariance matrix obtained from the DB reduction with those of the exact solution. The exact solution is exactly two dimensional when expressed in terms of the DB, and the variance picked up by the third mode is zero (order of round-off error). (Online version in colour.)

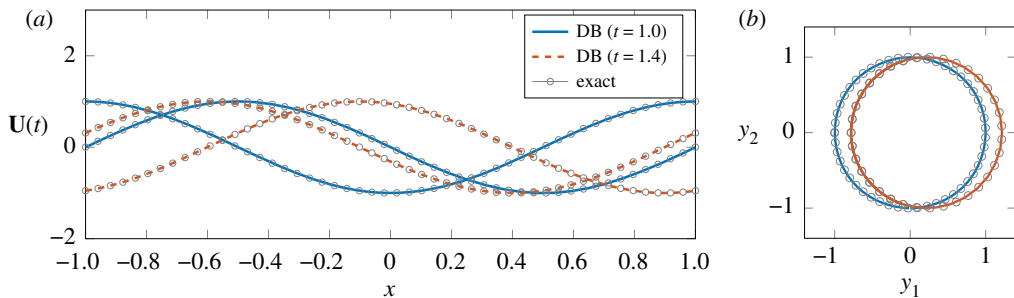


Figure 2. Reduction of the solution of the stochastic advection equation using the dynamic basis (DB): (a) DB and (b) stochastic coefficient (phase portrait). (Online version in colour.)

agreement with the exact solution. In general small eigenvalues of the covariance matrix may cause numerical instability owing to the computation of C^{-1} in equation (2.11). The issue of singularity of matrix C can be resolved by computing the pseudo-inverse of the covariance matrix. See [14] for similar treatment of the singularity of the covariance matrix. In figure 2, the DB and the stochastic coefficients are compared against the analytical values for two different times. In both cases excellent agreement is observed.

(b) Nonlinear dynamics: stochastic Burgers

The purpose of this example is to study the effect of nonlinearity in the context of comparing the DB reduction with the DO decomposition [11]. In both of these reductions an evolution equation for a time-dependent basis $U(t)$ is derived and a low-rank approximation is obtained by projecting the full-dimensional model (for DO) or data (for DB). A comparison with BO is not presented as DO and BO are equivalent [13]. To this end, we consider the Burgers equation:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2}, \quad x \in [0, \pi], \quad t \in [0, 1], \quad (3.3)$$

with $\nu = 0.005$ and subject to periodic boundary conditions and random initial conditions,

$$u(x, 0; \xi) = u_b(x) + u'(x; \xi), \quad (3.4)$$

where

$$u_b(x) = 0.5(\exp(\cos(x)) - 1.5) \sin(x + 2\pi \cdot 0.37)$$

and

$$u'(x; \xi) = \sigma_1 \xi_1 \sin(2\pi x) + \sigma_2 \xi_2 \sin(3\pi x),$$

where ξ_1 and ξ_2 are two independent uniform random variables with unit variance in the interval $[-1, 1]$ and σ_1 and σ_2 are the respective standard deviations in each random dimension. To generate data $\mathbf{T}(t)$ for the DB, equation (3.3) is sampled at 64 collocation points in the $\xi_1 - \xi_2$ space. These points were generated by the tensor product of the roots of the seventh-order Legendre chaos polynomials in each random direction. For the sake of comparison, the same collocation points are used to solve the evolution $d\mathbf{Y}/dt$ in the DO formulation. For the DO formulation of the Burgers equation, see [14]. Third-order Runge–Kutta is used to integrate both DO and DB equations in time with the same $\Delta t = 0.001$ and the Fourier spectral method with $n = 1024$ points is used for the space discretization. The same number of Fourier modes is used to solve the DO equation and to generate samples for $\mathbf{T}(t)$.

We also compared the DB/DO reduction with the KL decomposition, which is the optimal L_2 reduction in the sense of the reconstruction error,

$$\mathcal{E}(t) = \|\mathbb{E}[\mathbf{U}\mathbf{Y}^T - \mathbf{T}]\|_F. \quad (3.5)$$

The KL decomposition requires an eigenvalue problem to be solved for each time instant, which for $s \ll n$ has $\mathcal{O}(n^2)$ and $\mathcal{O}(s^3)$ computational complexity to compute and invert the full covariance matrix $\mathbf{C} \in \mathbb{R}^{s \times s}$, respectively, and similarly for $n \ll s$ has $\mathcal{O}(n^3)$ and $\mathcal{O}(s^2)$ computational complexity. This operation can be cost prohibitive for large datasets, e.g. extreme-scale computing applications [21]. On the other hand, in the DB formulation, the variational principle is formed with respect to the rate of change in the data (see equation (2.4)), whose optimality condition yields computationally scalable evolution equations for $\mathbf{U}(t)$ and $\mathbf{Y}(t)$.

In figure 3 the results for DB, DO and KL reductions are shown. We chose a reduction size of $r = 5$ in all cases. Two cases are considered, as follows. Case I: $\sigma_1 = \sigma_2 = 0.05$ and case II: $\sigma_1 = \sigma_2 = 0.5$. In the first row of figure 3 the reconstruction error $\mathcal{E}(t)$ (figure 3a) and the eigenvalues $\boldsymbol{\Lambda}(t)$ of the covariance matrix $\mathbf{C}(t)$ (figure 3b) are shown. For the KL reduction, r largest eigenvalues of the full covariance matrix are shown. Similarly, the second row shows the results for the reconstruction error and the eigenvalues of the covariance matrix for case II. In figure 3b,d the energy of the mean $\lambda_m(t) = \langle \bar{\mathbf{u}}(t), \bar{\mathbf{u}}(t) \rangle$ is also shown, where $\bar{\mathbf{u}}(t)$ is the sample mean, i.e. $\bar{\mathbf{u}}(t) = \mathbb{E}[\mathbf{T}(t)]$. It is clear that for case I the energy of the mean is roughly two orders of magnitude larger than the variance, i.e. the sum of all eigenvalues. In that sense, the random initial condition amounts to a small perturbation around the mean state and, therefore, the evolution of the random initial condition can be approximated by linearized dynamics around the mean state $\bar{\mathbf{u}}(t)$. It is clear that for this case the DB, DO and KL reductions are nearly identical. However, for case II, the nonlinear effects are much stronger, as evidenced by observing that the energy of the mean is comparable to the eigenvalues of the covariance matrix; see figure 3d. In this case a significant discrepancy between DO and DB reductions is observed, while DB is closer to KL with smaller reconstruction error; see figure 3c.

One source of discrepancy between DO and DB is that, in the DO formulation, the samples are generated *on the fly* by the reduced-order model. Therefore, the samples generated by DO deviate from the truth. This is not, however, the case in DB, where the samples, represented by $\mathbf{T}(t)$, are the truth. On the other hand, in the case of linear dynamics, the discrepancy between the samples generated by DO and the truth approaches zero. For the proof, see appendix A of [14] and theorem 2.4 in [18]. This explains why the DO and DB are nearly identical for linear regimes, i.e. case I, and are different for nonlinear regimes, i.e. case II.

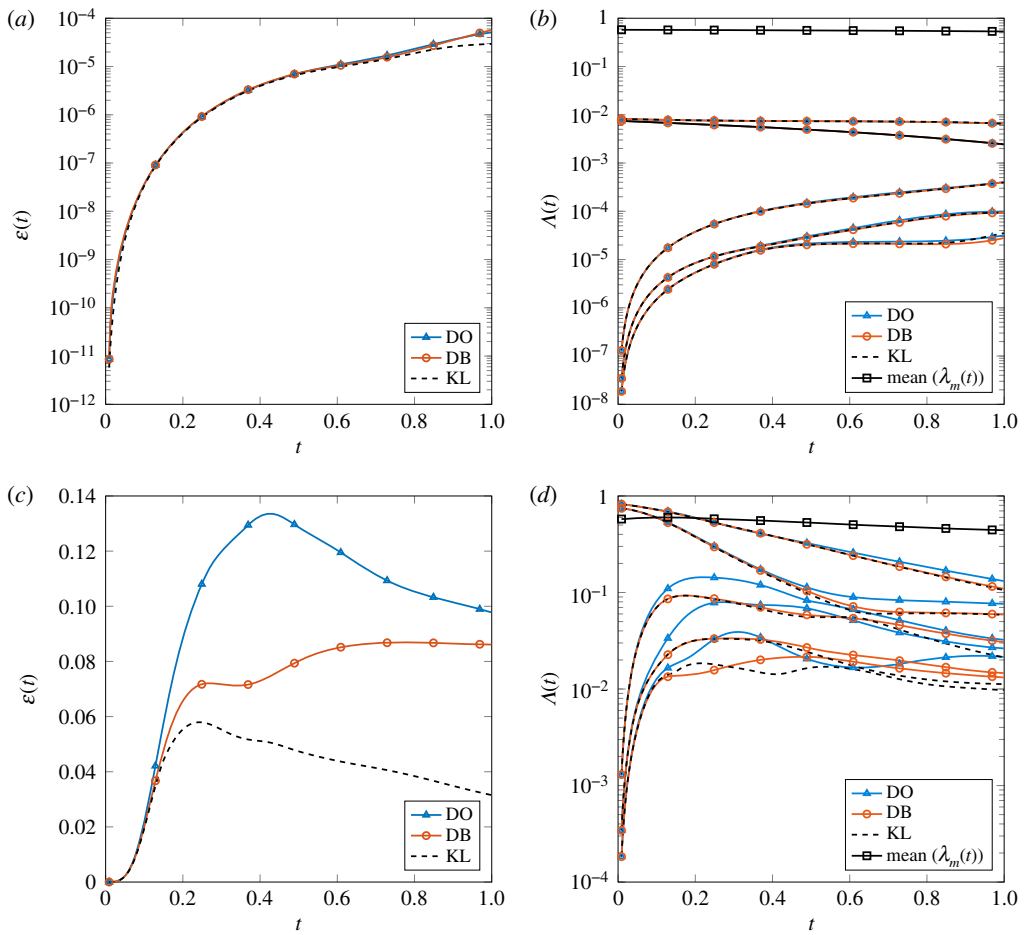


Figure 3. The effect of nonlinearity and comparison of dynamic basis (DB) reduction with dynamically orthogonal (DO) decomposition and Karhunen–Loève (KL) decomposition. (a,c) Reconstruction error and (b,d) eigenvalues of the covariance matrix and the energy of the mean of the samples: $\lambda_m(t) = \langle \bar{\mathbf{u}}(t), \bar{\mathbf{u}}(t) \rangle$, where $\bar{\mathbf{u}}(t) = \mathbb{E}[\mathbf{T}(t)]$. (a,b) $\sigma_1 = \sigma_2 = 0.05$ and (c,d) $\sigma_1 = \sigma_2 = 0.5$. (Online version in colour.)

(c) Transient instability: Kuramoto–Sivashinsky

One of the main utilities of a low-rank approximation of high-dimensional dynamical systems is to provide an interpretable explanation of complex spatio-temporal dynamics. In this section, we demonstrate the application of the DB reduction to the one-dimensional stochastic Kuramoto–Sivashinsky equation. This equation has been widely studied for chaotic spatio-temporal dynamics in wide-ranging applications, including reaction–diffusion systems [22], fluctuations in fluid films on inclines [23] and instabilities in laminar flame fronts [24]. The Kuramoto–Sivashinsky equation exhibits *intrinsic instabilities* and as a result the stochastic perturbations may undergo a significant growth. The spontaneity of the instability makes it a challenging problem for any reduction technique as the reduced subspace during the transition could be very different from that of the equilibrium states. We apply the DB reduction to realizations of the Kuramoto–Sivashinsky system to demonstrate how the $\mathbf{U}(t)$ ‘follows’ the dynamics as the observations undergo a transient instability. This transient instability is visualized in figure 4a, where a sudden transition occurs near $t = 0.75$. The solution shown in figure 4a is one realization of the Kuramoto–Sivashinsky equation with the following set-up. We consider Kuramoto–Sivashinsky

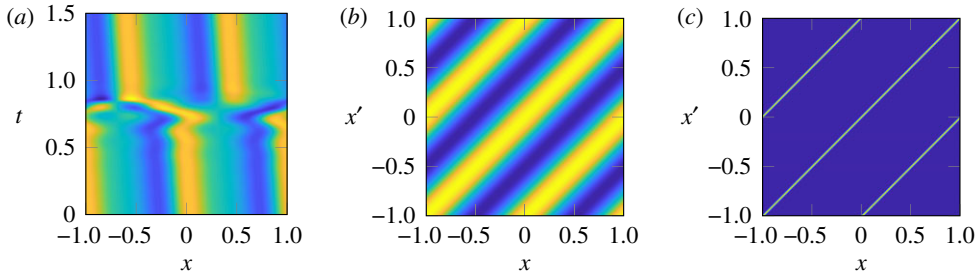


Figure 4. Kuramoto–Sivashinsky: (a) solution $u(x, t)$ shows a sudden transition; covariance kernel of the random perturbation with correlation length (b) $l_c = 2.5$ and (c) $l_c = 0.05$. (Online version in colour.)

equation

$$\frac{\partial u}{\partial t} = u \frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \epsilon \frac{\partial^4 u}{\partial x^4}, \quad x \in [-1, 1], \quad t \in [0, T_f], \quad (3.6)$$

with $\epsilon = 0.01$ and subject to random initial condition

$$u(x, 0; \xi) = u_b(x) + u'(x; \xi), \quad x \in [-1, 1]. \quad (3.7)$$

In the above equation $u_b(x)$ is chosen to be the terminal solution of equation (3.6) after long-time integration. The random perturbation of the initial condition, denoted by $u'(x; \xi)$, is set to have periodic covariance given by

$$K(x, x') = \sigma^2 \exp\left(-\frac{2 \sin^2(\pi(x - x'))}{l_c^2}\right),$$

where σ is the standard deviation and l_c is the correlation length. We perform KL decomposition

$$\int_{-1}^1 K(x, x') \phi_i(x') dx' = \lambda_i \phi_i(x),$$

where λ_i and $\phi_i(x)$ are the i th eigenvalue and eigenfunction of the covariance kernel. The random perturbation is then approximated as

$$u'(x; \xi) = \sum_{i=1}^d \sqrt{\lambda_i} \xi_i \phi_i(x) + \epsilon,$$

where $\xi_i \in [0, 1]$ are independent uniform random numbers with unit variance. The dimension d is chosen such that 99% of the energy is captured by the KL decomposition. Decreasing l_c increases the number of dimensions required to capture the energy threshold. We consider two cases of low-dimensional and high-dimensional random perturbations with $l_c = 2.5$ and $l_c = 0.05$, respectively. The covariance kernels for these two correlation lengths are shown in figure 4.

Three-dimensional random perturbation. We consider a correlation length of $l_c = 2.5$ and $\sigma = 0.1$. The three-dimensional ($d = 3$) expansion of the KL decomposition captures more than 99% of the energy of the covariance kernel. We sample the Kuramoto–Sivashinsky equation at probabilistic $Q = 5$ collocation points in each ξ_i direction; therefore, the total number of samples is $s = 5^3$. Correspondingly, the sample weights \mathbf{w}_ξ are the probabilistic collocation weights. We use *chebfun* [25] as a blackbox solver to solve the Kuramoto–Sivashinsky equation (3.6) for $s = 5^3$ realizations of the random initial condition. The solution is post-processed with fixed Chebyshev polynomial order $n = 256$. Matrix \mathbf{T} is sampled at discrete times with a constant time interval of $dt = 0.001$ from $t = 0$ to $T_f = 1.2$. This time interval encompasses the transition period that occurs, $0.6 < t < 1.0$.

We perform DB reduction with size $r = 4$. In figure 5a, the eigenvalues of the reduced covariance matrix versus time are shown. The eigenvalues of the covariance matrix represent

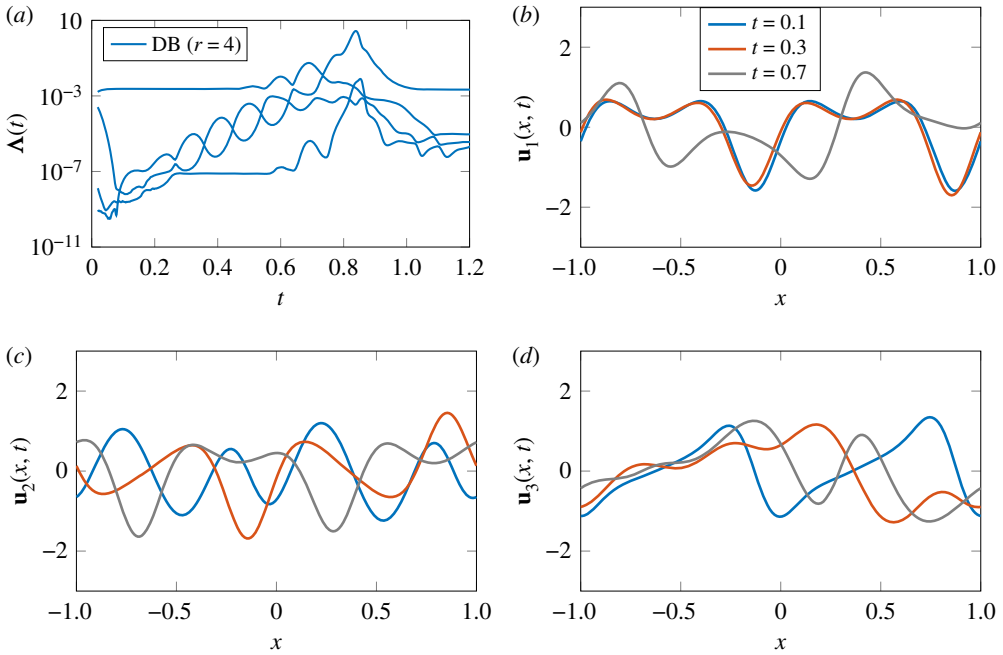


Figure 5. Dynamic basis (DB) reduction for the Kuramoto–Sivashinsky equation with reduction size $r = 4$. (a) Eigenvalues of reduced covariance matrix. The first three modes at different time instants: (b) \mathbf{u}_1 , (c) \mathbf{u}_2 and (d) \mathbf{u}_3 . (Online version in colour.)

variance or the energy associated with the DB. The first three DBs associated with the three largest eigenvalues of the covariance matrix are shown in figure 5*b–d*.

We first observe that the energy of the first mode remains roughly constant for $0 < t < 0.5$. However, the second and third modes undergo four orders of magnitude growth during the same period. Correspondingly, the first DB remains nearly static for $0 < t < 0.5$ (see figure 5*b*). However, \mathbf{u}_2 and \mathbf{u}_3 are evolving during the same time interval (figure 5*c,d*). The second and third modes have near-eigenvalue crossings periodically in this interval, which causes very rapid changes in the modes when their eigenvalues become close. The significant growth of the energy of the second and third modes reveals the transient unstable directions of the dynamical system. However, these two modes are ‘buried’ underneath the most energetic mode, e.g. the first mode. We note that detecting the unstable directions is crucial for control of these systems [26]. Detecting these directions can be used in building precursors of sudden transitions or rare events in transient systems.

We also computed DMD modes to compare against the DB reduction. DMD has been very successful as a diagnostic tool to characterize complex dynamical systems. However, it is difficult for DMD to handle transient phenomena. To this end, we computed the DMD modes for a deterministic realization of the Kuramoto–Sivashinsky equation. The realization is generated by solving the Kuramoto–Sivashinsky equation for the collocation point $\xi_0 = (-0.91, -0.54, 0.91)$. This collocation point is chosen arbitrarily and any other collocation point would result in the same qualitative observations that we make in this section. To compute the DMD modes we first performed singular value decomposition (SVD) of the mean-subtracted snapshots of the Kuramoto–Sivashinsky equation in the span of $t \in [0, T_f]$ for $\xi_0 = (-0.91, -0.54, 0.91)$. The SVD truncation threshold $e = \sum_{i=1}^n \sigma_i / \sum_{i=1}^N \sigma_i$ was set to 99%, where σ_i are the singular values and N is the total number of snapshots. The 99% threshold resulted in $n = 11$ low-rank approximations. We then computed the DMD modes Φ according to the procedure explained in [8]. The DMD-reconstructed data were then computed using: $\mathbf{u}(t) \simeq \sum_{k=1}^n \Phi_k \exp(\omega_k t) b_k$, where b_k is the initial amplitude of each mode, Φ_k ’s are the DMD eigenvectors and ω_k ’s are the eigenvalues associated

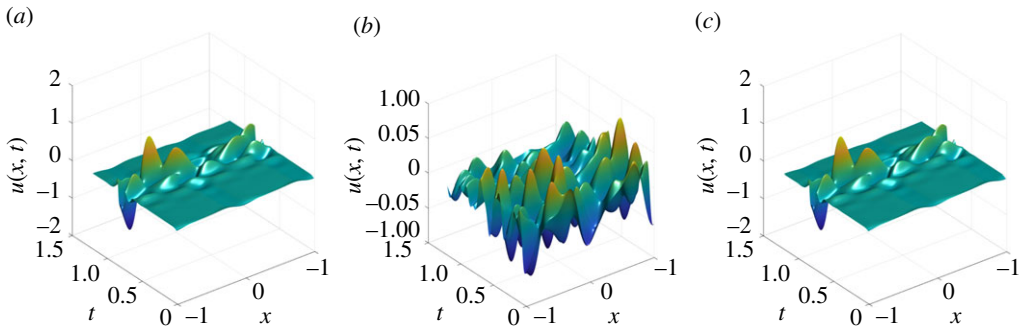


Figure 6. Reconstruction of a Kuramoto–Sivashinsky sample with: (a) true solution; (b) reconstruction with dynamic mode decomposition and (c) reconstruction with dynamic basis. (Online version in colour.)

Table 1. Reconstruction error of the Kuramoto–Sivashinsky equation: comparison between dynamic mode decomposition (DMD) and dynamic basis (DB).

r	2	4	6	8
error (DMD)	9.01×10^{-2}	9.027×10^{-2}	9.18×10^{-2}	9.24×10^{-2}
error (DB)	2.10×10^{-1}	6.92×10^{-3}	1.88×10^{-3}	8.92×10^{-4}

with DMD modes. In figure 6, we compare the reconstruction of the DMD with $n = 11$ and DB with $r = 4$. The DB reconstruction obtained by: $\mathbf{u}(t) \simeq \sum_{i=1}^r \mathbf{Y}_{s_0, i}(t) \mathbf{u}_i(t)$, where $\mathbf{y}_i(t, \xi_0)$ are the stochastic coefficients associated with $\xi_0 = (-0.91, -0.54, 0.91)$, i.e. a fixed row (s) of matrix \mathbf{Y} . It is clear that DMD fails to capture the transient instability, while DB reconstruction looks qualitatively accurate.

We perform a quantitative comparison between DMD and DB. We rank the $n = 11$ DMD modes based on their growth rates from increasing to decreasing, and we perform DMD reconstruction $\mathbf{u}(t) \simeq \sum_{i=1}^r \Phi_k \exp(\omega_k t) b_k$ for $r = 2, 4, 6$ and 8 . We also perform DB for the same reduction sizes. We then compute the reconstruction error: $\epsilon^2 = \sum_{k=1}^K \|\mathbf{u}_{true}(t_k) - \mathbf{u}(t_k)\|^2 / K$, where K is the total number of snapshots and $\mathbf{u}(t_k)$ is the reconstructed solution based on DMD or DB. In table 1, the reconstruction errors ϵ for DB and DMD are shown. DMD is incapable of converging to the true solution as the size of DMD modes increases; however, DB converges to the true solution as the reduction size increases. The limitation of DMD for capturing transient dynamics has been recognized before; see for example [9], chapter 1, section 1.5.2.

One-hundred-dimensional random perturbation. In this section we present the results of DB reduction of a 100-dimensional random initial condition with a correlation length of $l_c = 0.05$ (figure 5c). To capture the small scales of the random initial condition the state space is represented with a Chebyshev polynomial of order $n = 1024$. We draw Monte Carlo samples of the initial condition with sample sizes of $s = 1000$ and $s = 5000$. We note that these sample sizes are quite small given the high dimensionality of the random space. However, as we demonstrate here the DB reduction extracts the low-dimensional structure of the observables despite the severely low number of samples. We choose a reduction size of $r = 6$. In figure 7 the eigenvalues of the covariance matrix for the sample sizes are shown. As is clear, the DB reductions with sample sizes of 1000 and 5000 are very close before and during the transition. This implies that, despite the high dimensionality of the random space, the transient response has a low-dimensional structure—albeit a highly time-dependent one—that is effectively captured with the proposed reduction with severely sparse samples.

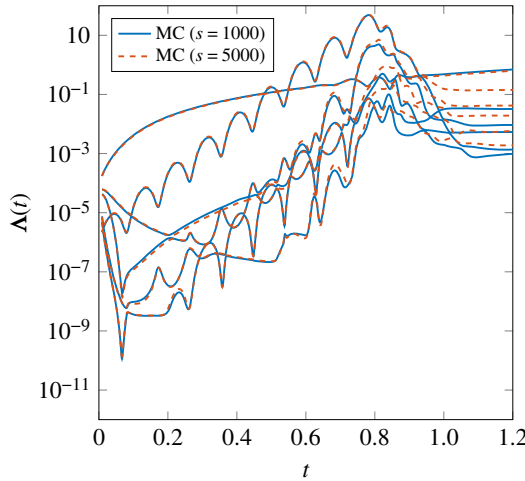


Figure 7. Dynamic basis reduction of the Kuramoto–Sivashinsky equation subject to 100 random initial conditions. Two Monte Carlo (MC) sample sizes of $s = 1000$ and $s = 5000$ are shown. The reduction size is $r = 6$. (Online version in colour.)

(d) Transient vertical jet

In this section we perform dynamic reduction for a two-dimensional transient vertical jet flushed into quiescent background flow. The jet is driven by random boundary conditions as described in this section. We note that, in the DO/BO formulation, it is still not known how to determine the boundary condition for the DO/BO modes in the case of random boundary conditions [20]. However, in the DB formulation the boundary conditions for the DB modes are optimally determined via the variational principle. The schematic of the problem is shown in figure 8*a*. To generate the samples, we performed direct numerical simulation (DNS) of the two-dimensional incompressible Navier–Stokes equations,

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \frac{1}{Re} \nabla^2 \mathbf{u}$$

and

$$\nabla \cdot \mathbf{u} = 0,$$

where $\mathbf{u}(\mathbf{x}, t; \xi)$ is the random velocity field with two components of $\mathbf{u} = (u, v)$ and the spatial coordinates of $\mathbf{x} = (x, y)$, and $p(\mathbf{x}, t; \xi)$ is the random pressure field and $Re = 500$ is the Reynolds number. We consider a stochastic jet velocity given by

$$u(x, t; \xi) = (1 + u'(x, \xi)) \left(1 - (x/r)^2\right) \exp(-0.823(x/r)^4). \quad (3.8)$$

The mean of the prescribed jet boundary condition is similar to that found in previous studies [27,28]. The random component of the boundary condition is represented by a uniform random field with the covariance kernel specified by

$$K(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{l_c^2}\right),$$

where $l_c = 0.5$ and $\sigma = 0.01$. The dimension $d = 2$ is chosen such that 99% of the energy is captured by the KL decomposition. We performed DNS using the spectral/hp element method with $Ne = 1604$ and a spectral polynomial order of $p = 5$. With these settings, the size of the discrete phase space is roughly $n = 115\,000$ when accounting for the two components of the velocity vector field. The time integration was performed using the third-order stiffly stable scheme [29] with time intervals of $\Delta t = 5 \times 10^{-3}$. We sample the DNS at PCM points with $Q = 6$ quadrature points in each random dimension. Therefore, the total number of samples is $s = Q^2 = 36$.

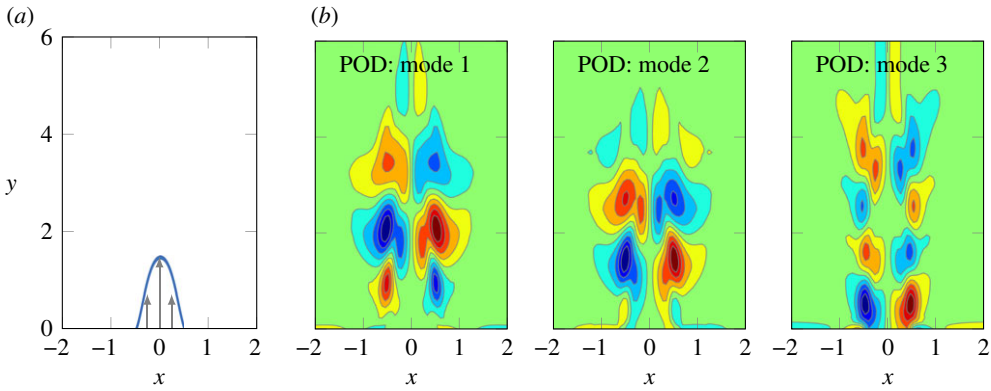


Figure 8. Transient vertical jet in quiescent flow. (a) Schematic of the problem. (b) The three most energetic proper orthogonal decomposition (POD) modes. (Online version in colour.)

As a reference reduction, we first computed the POD modes. Since POD modes are often used in deterministic flows, we explain how these modes are computed in a stochastic setting. To do so, we first formed the matrix \mathbf{T} that contains all mean-subtracted snapshots of all 36 samples, i.e. the entire DNS data. We then performed SVD of this matrix. The POD modes are then computed as the left singular eigenvectors of matrix \mathbf{T} . Therefore, the computed POD modes are an optimal basis for the entire DNS data in the second norm sense. In figure 8b, the first three POD modes associated with the largest singular values are shown.

We solved the evolution equations (2.11) and (2.12) with $r = 3$ with a fourth-order Runge–Kutta integration scheme. The time snapshots of matrix \mathbf{T} were obtained by skipping every three time steps of the DNS, i.e. the time interval of DB evolution is $\Delta t = 2 \times 10^{-2}$.

In figure 9, we show the results of the state of the DB for three time snapshots at $t_1 = 2$, $t_2 = 6$ and $t_3 = 10$ time units, with each row showing the results at a fixed time. For the purpose of visualizing the mean, we solved a passive scalar transport equation along with the Navier–Stokes equation with Schmidt number $Sc = 1$. See [18] for more details of a similar approach for visualization. In the leftmost column, the mean data are visualized by the scalar transport intensity. The mean shows the puff growing in size as the jet entrains the quiescent background flow while advecting upward. In the middle three columns, the first three most dominant dynamic bases are visualized by their vorticity: $\omega_{z_i} = \partial u_i / \partial y - \partial v_i / \partial x$. The evolution of the DB along with the data is clearly seen. The DB reveals the *spatio-temporal coherent structures* in the flow. Similar to the POD modes (figure 8b), the higher modes of the DB show finer spatial structures. However, the POD modes are static and they are optimal in a time-average sense, as opposed to the DB that evolves according to $\dot{\mathbf{T}}$. We will compare the POD reduction error with that of the DB. In the rightmost column, the low-dimensional attractor is visualized by plotting the iso-surface of the joint probability density function $\pi_{y_1 y_2 y_3}$ of the three most dominant stochastic coefficients: y_1 , y_2 and y_3 . Since the DNS are sampled at PCM points, the PCE is used to generate more samples for visualizing $\pi_{y_1 y_2 y_3}$ in the post-processing stage. The attractor clearly shows a low-dimensional manifold as it evolves in time.

We now leverage the data-driven feature of the DB reduction and we apply it to partial observations. To this end, we applied the DB with a reduction size of $r = 3$ to a subset of the observations, as shown by the circle in figure 10. Therefore, the data matrix \mathbf{T} only includes points that are inside the circle in figure 10 and points outside of the circle are not included in the DB calculation. The extracted DBs associated with the three modes are also shown in the three middle columns of figure 10 for $t = 2, 6$ and 10 . In this case the dynamics represent a locally orthonormal basis. At $t = 2$, the vortex has not reached the region of interest (ROI); however, owing to the incompressibility of the flow, the random boundary conditions affect the entire domain and the DB modes capture this effect with more strength in the lower part of the ROI. The attractor at

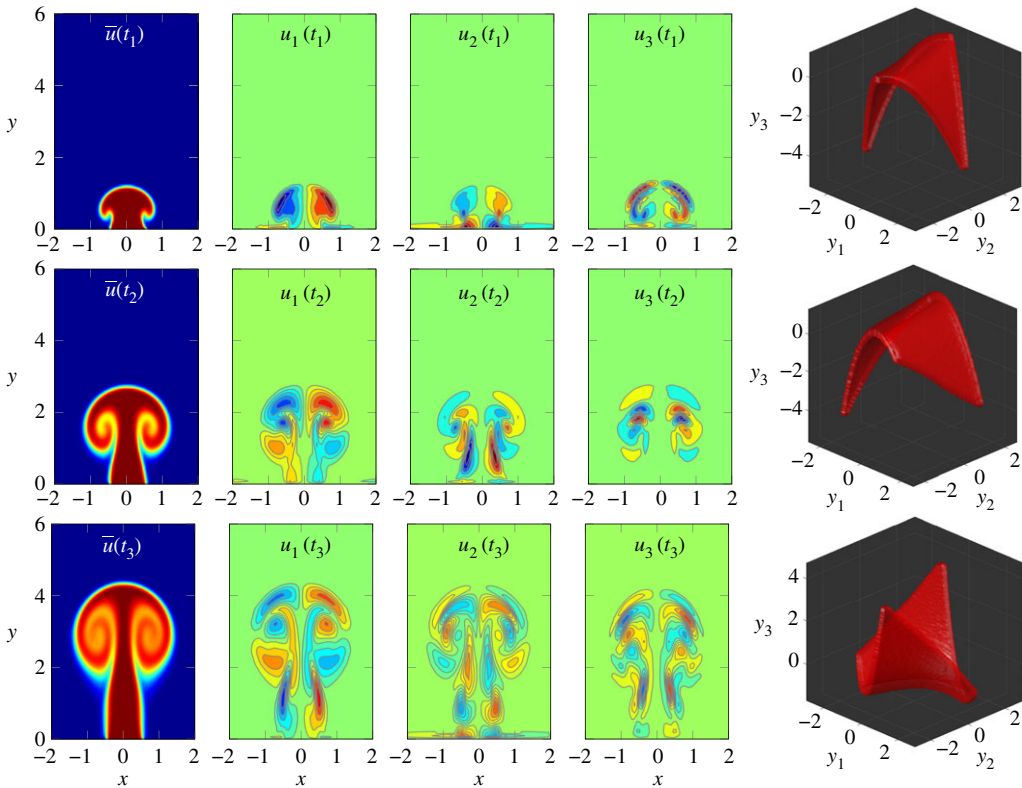


Figure 9. Dynamic basis reduction of a transient vertical jet in quiescent flow. Left: Flow mean visualized by a passive scalar. Middle: The three most dominant dynamics modes are visualized by vorticity ω_z . Right: The iso-surface of the joint probability density function $\pi_{y_1 y_2 y_3}$ of the first three dynamic modes. Each row shows the state of the reduction at a fixed time with first row: $t_1 = 2$; second row: $t_2 = 6$; and third row: $t_3 = 10$. (Online version in colour.)

the corresponding times is shown in the right column of figure 10. The low-dimensional attractor exhibits a different structure from that of the full domain shown in figure 9. Applying the DB reduction to a subset of observations enables the extraction of local low-dimensional structures by ‘zooming in’ to a ROI. Moreover, this enables the employment of the DB reduction in a domain decomposition scheme [30], where each domain has a set of locally orthonormal DBs. This is in contrast to the DO/BO decompositions, where neither of the decompositions can be applied to a subset of the full domain. This is because both DO and BO are model driven and as such they would require the specification of the boundary conditions for the mean and the spatial modes at the boundaries of the ROI. Clearly, these time-dependent values at the arbitrary boundaries are not known.

In figure 11*a*, we make a comparison between POD and DB reconstruction errors, given by equation (3.5), in which the POD mode \mathbf{U} represents the first r dominant POD modes and \mathbf{Y} is obtained via the orthogonal projection of the data onto \mathbf{U} , i.e. $\mathbf{Y}(t) = \langle \mathbf{U}, \mathbf{T}(t) \rangle$. As shown in figure 11, the reduction of the DB with $r=2$ is comparable to that of POD with $r=20$ for long time, and in the early evolution (i.e. $t < 4$) the error of DB ($r=2$) is smaller than that of POD with $r=50$. A similar behaviour is observed for DB $r=3, 4$ and 5, where a 10-fold increase in POD reduction size is required to have a comparable reduction error with DB. We also observe that the DB error increases with time. This is the result of two effects: (i) as the jet advects upward, the uncertainty grows and more modes are required to better estimate the observations; and (ii) the effect of unresolved modes in the DB reduction. In the case of drastic reduction, for example $r=2$, discarding the unresolved modes will lead to an increase

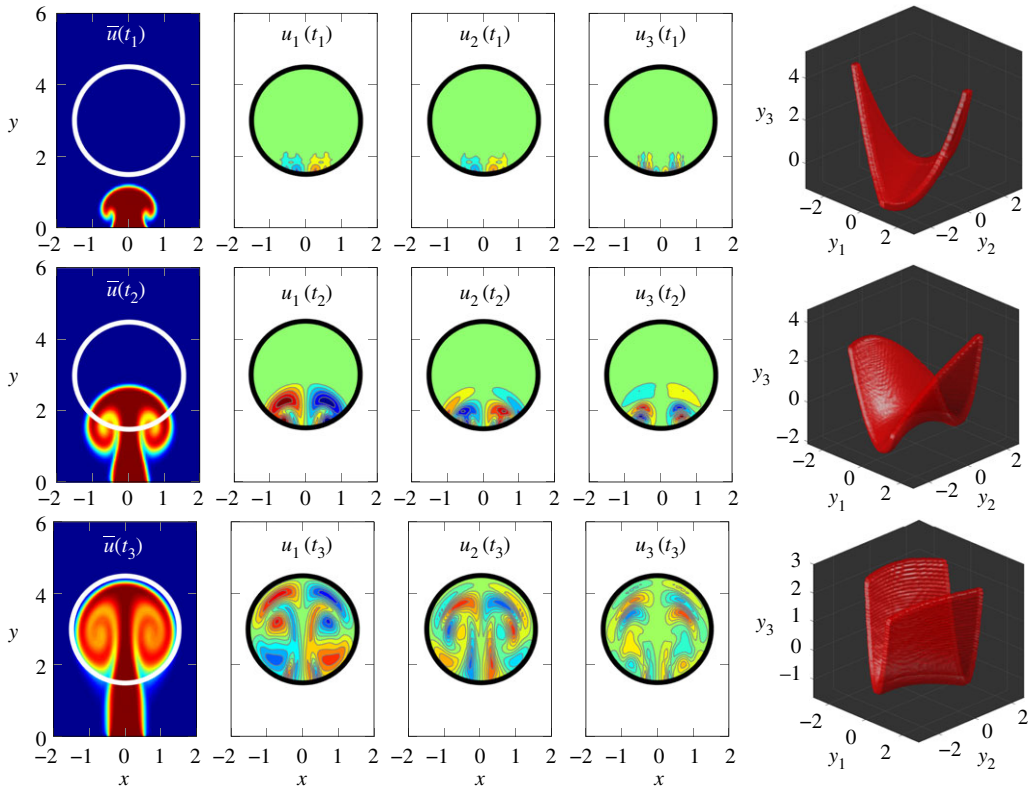


Figure 10. Dynamic basis reduction of a transient vertical jet in quiescent flow with a clipped domain. Only data points that lie inside the circle are used in the dynamic basis reduction. Left: Flow mean visualized by a passive scalar. Middle: the three most dominant dynamics modes are visualized by vorticity ω_z . Right: The iso-surface of the joint probability density function $\pi_{y_1 y_2 y_3}$ of the first three dynamic modes. Each row shows the state of the reduction at a fixed time with first row: $t_1 = 2$; second row: $t_1 = 6$; and third row: $t_3 = 10$. (Online version in colour.)

in error over time. The effect of unresolved modes can be properly studied in the Mori–Zwanzig formalism [31]. As can be seen in figure 11*a* increasing the number of DBs significantly decreases the error and the growth rate of the error.

In figure 11*b*, we show the results of different time discretization schemes for: (i) computing $\dot{\mathbf{T}}$ with two different schemes of first-order Euler explicit (EE1) and fourth-order central finite difference (FD4) and (ii) the time advancement of $\mathbf{U}(t)$ and $\mathbf{Y}(t)$ via equations (2.11) and (2.12) with EE1 and fourth-order Runge–Kutta (RK4). All four combinations of the above schemes are considered with the identifiers in the form of X–Y, where the first part of the identifier (X) shows the discretization scheme used for estimating $\dot{\mathbf{T}}$ and the second part of the identifier (Y) shows the time-advancement scheme for solving equations (2.11) and (2.12). In particular we considered these schemes: EE1–EE1, EE1–RK4, FD4–EE1 and FD4–RK4. We used DB with a size of $r = 5$ for all these schemes. The errors of schemes EE1–EE1 and FD4–EE1 are very close—suggesting that the time-advancement error of solving equations (2.11) and (2.12) is the primary source of error and increasing the accuracy of estimating $\dot{\mathbf{T}}$ does not improve the overall accuracy. Comparing the errors of EE1–EE1 with those of EE1–RK4 confirms that the error can be reduced by increasing the accuracy of the time integrator of the evolution equations for \mathbf{U} and \mathbf{Y} . However, comparing the errors of EE1–RK4 with those of FD4–RK4 demonstrates that the primary contributor to the error for the scheme EE1–RK4 is the estimation of $\dot{\mathbf{T}}$. The totality of these observations shows that both the accuracy of estimating $\dot{\mathbf{T}}$ and solving the evolution equations of the DB affect the reduction error potentially significantly. Obviously, if Δt is decreased, the contribution of the time-discretization error will become less significant than that of the reduction error.

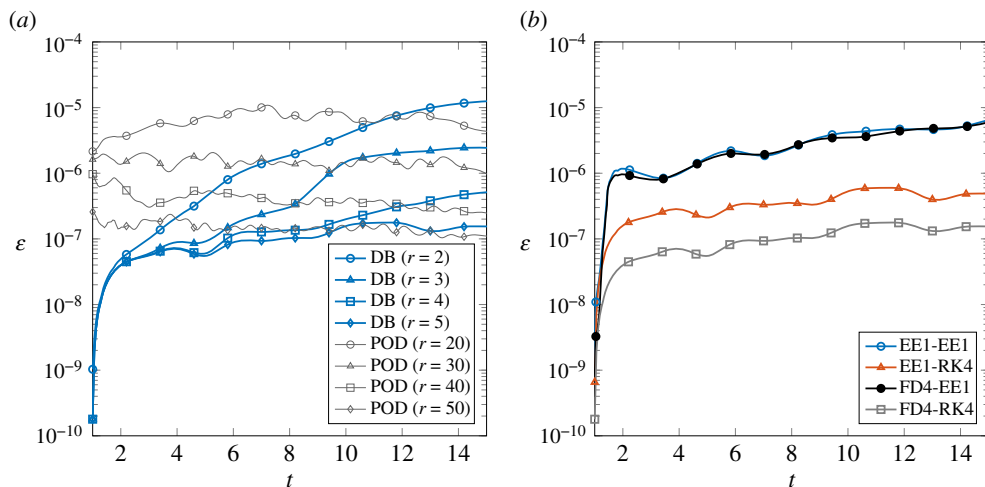


Figure 11. (a) Comparison of the reduction error between the dynamic basis (DB) and proper orthogonal decomposition (POD). (b) Effect of the discretization scheme for computing $\dot{\mathbf{T}}$ and evolution of the DB. First-order Euler explicit (EE1), fourth-order central finite difference (FD4) and Runge–Kutta (RK4). Identifiers are shown in the form of X–Y: X shows the discretization scheme used for estimating $\dot{\mathbf{T}}$ and Y shows the time-advancement scheme for solving the evolution equation for the DB. (Online version in colour.)

4. Conclusion

The objective of this paper is to develop a reduced description of the transient dynamics that relies on the observables of the dynamics. To this end, we present a variational principle whose optimality conditions lead to a closed-form evolution equation for a set of time-dependent orthonormal modes and their associated coefficients, referred to as the dynamic basis. The DB is driven by the time derivative of the random observations, and it does not require the underlying dynamical system that generates the observables. In that sense, the proposed reduction is purely observation driven and may apply to systems whose models are not known.

The application of the DB was demonstrated for a variety of examples. (i) The advection equation with random wave speed for which the exact solution for the DB was derived and was used for validation purposes. (ii) Stochastic Burgers equation, for which DB reduction was compared with KL and DO decompositions. (iii) Kuramoto–Sivashinsky equation subject to random initial conditions where it was shown that the DB captures the transient intrinsic instability—even in the presence of a high-dimensional random initial condition. The performance of the DB reduction was compared against the DMD. (iv) Navier–Stokes equations subject to random boundary conditions for which we compared the effectiveness of the DB versus POD, as an example of static-basis reductions.

The DB evolution equation has a linear computational complexity with respect to the number of observables and the number of random realizations. This is a favourable feature if applications of large dimension are of interest, for example weather forecast, ocean models and turbulent flows.

Data accessibility. The code and data used in this paper can be obtained from: <https://github.com/hbabae/Dynamic-Basis>. The package Chebfun must be installed to run the codes.

Competing interests. I declare I have no competing interests.

Funding. The research has been supported by the NASA Transformational Tools and Technologies (TTT) Project, grant no. 80NSSC18M0150, and an award by the American Chemical Society, Petroleum Research Fund, grant no. 59081-DN19.

Acknowledgements. The author would like to thank Prof. Karniadakis for useful comments. This research was also supported in part by the University of Pittsburgh Center for Research Computing through the resources

Appendix A. Optimality condition of the variational principle

In the following we show that the first-order optimality condition of the variational principle (2.6) leads to closed-form evolution equations for \mathbf{U} and \mathbf{Y} . For the sake of brevity we drop the explicit dependence on (t) in the following derivation. We first observe that

$$\begin{aligned} \mathcal{G}(\dot{\mathbf{U}}, \dot{\mathbf{Y}}) &= \left(\dot{\mathbf{u}}_i^T \mathbf{W}_x \dot{\mathbf{u}}_j \right) \left(\mathbf{y}_i^T \mathbf{W}_\xi \mathbf{y}_j \right) + \left(\mathbf{u}_i^T \mathbf{W}_x \mathbf{u}_j \right) \left(\dot{\mathbf{y}}_i^T \mathbf{W}_\xi \dot{\mathbf{y}}_j \right) + 2 \left(\dot{\mathbf{u}}_i^T \mathbf{W}_x \mathbf{u}_j \right) \left(\mathbf{y}_i^T \mathbf{W}_\xi \dot{\mathbf{y}}_j \right) \\ &\quad - 2 \dot{\mathbf{u}}_i^T \mathbf{W}_x \dot{\mathbf{T}} \mathbf{W}_\xi \mathbf{y}_i - 2 \mathbf{u}_i^T \mathbf{W}_x \dot{\mathbf{T}} \mathbf{W}_\xi \dot{\mathbf{y}}_i + \|\mathbb{E}[\dot{\mathbf{T}}]\|_F + \lambda_{ij} \left(\mathbf{u}_i^T \mathbf{W}_x \dot{\mathbf{u}}_j - \phi_{ij} \right). \end{aligned}$$

The first-order optimality condition requires $\frac{\partial \mathcal{G}}{\partial \dot{\mathbf{u}}_k} = 0$ and $\frac{\partial \mathcal{G}}{\partial \dot{\mathbf{y}}_k} = 0$ for $k = 1, 2, \dots, r$. We first observe that the gradient of \mathcal{G} with respect to $\dot{\mathbf{u}}_k$ can be expressed as

$$\frac{\partial \mathcal{G}}{\partial \dot{\mathbf{u}}_k} = 2 \mathbf{W}_x \dot{\mathbf{u}}_j \left(\mathbf{y}_k^T \mathbf{W}_\xi \mathbf{y}_j \right) + 2 \mathbf{W}_x \mathbf{u}_j \left(\dot{\mathbf{y}}_k^T \mathbf{W}_\xi \dot{\mathbf{y}}_j \right) - 2 \mathbf{W}_x \dot{\mathbf{T}} \mathbf{W}_\xi \mathbf{y}_k + \lambda_{jk} \mathbf{W}_x \mathbf{u}_j = \mathbf{0}. \quad (\text{A } 1)$$

To eliminate λ_{kj} , we take the inner product of the above equation by \mathbf{u}_l

$$\mathbf{u}_l^T \frac{\partial \mathcal{G}}{\partial \dot{\mathbf{u}}_k} = 2 \phi_{lj} \left(\mathbf{y}_k^T \mathbf{W}_\xi \mathbf{y}_j \right) + 2 \delta_{lj} \left(\dot{\mathbf{y}}_k^T \mathbf{W}_\xi \dot{\mathbf{y}}_j \right) - 2 \mathbf{u}_l^T \mathbf{W}_x \dot{\mathbf{T}} \mathbf{W}_\xi \mathbf{y}_k + \lambda_{jk} \delta_{lj} = 0,$$

where we have made use of the definition of $\phi_{ij}(t) := \langle \mathbf{u}_i, \dot{\mathbf{u}}_j \rangle$. Rearranging the above equation results in

$$\lambda_{lk} = -2 \phi_{lj} \left(\mathbf{y}_k^T \mathbf{W}_\xi \mathbf{y}_j \right) - 2 \left(\dot{\mathbf{y}}_k^T \mathbf{W}_\xi \dot{\mathbf{y}}_j \right) + 2 \mathbf{u}_l^T \mathbf{W}_x \dot{\mathbf{T}} \mathbf{W}_\xi \mathbf{y}_k.$$

After multiplying equation (A 1) from the left by $\mathbf{W}_x^{-1}/2$ and substituting λ_{jk} from the above equation we obtain

$$\dot{\mathbf{u}}_j \left(\mathbf{y}_k^T \mathbf{W}_\xi \mathbf{y}_j \right) = \left(\mathbf{I} - \mathbf{u}_j \mathbf{u}_j^T \mathbf{W}_x \right) \dot{\mathbf{T}} \mathbf{W}_\xi \mathbf{y}_k + \mathbf{u}_j \phi_{jl} \left(\dot{\mathbf{y}}_k^T \mathbf{W}_\xi \dot{\mathbf{y}}_l \right).$$

In the vectorized form the above equation can be written as

$$\dot{\mathbf{U}} = \left(\mathbf{I} - \mathbf{U} \mathbf{U}^T \mathbf{W}_x \right) \dot{\mathbf{T}} \mathbf{W}_\xi \mathbf{Y} \mathbf{C}^{-1} + \mathbf{U} \boldsymbol{\phi},$$

where $\mathbf{C}_{ij} = \mathbf{y}_i^T \mathbf{W}_\xi \mathbf{y}_j$ is the covariance matrix. Similarly,

$$\frac{\partial \mathcal{G}}{\partial \dot{\mathbf{y}}_k} = 2 \mathbf{W}_\xi \dot{\mathbf{y}}_j \left(\mathbf{u}_k^T \mathbf{W}_x \mathbf{u}_j \right) + 2 \mathbf{W}_\xi \mathbf{y}_j \left(\dot{\mathbf{u}}_k^T \mathbf{W}_x \dot{\mathbf{u}}_j \right) - 2 \mathbf{W}_\xi \dot{\mathbf{T}}^T \mathbf{W}_x \mathbf{u}_k = \mathbf{0}. \quad (\text{A } 2)$$

In the above expression we have made use of $(\dot{\mathbf{u}}_i^T \mathbf{W}_x \mathbf{u}_j)(\mathbf{y}_i^T \mathbf{W}_\xi \dot{\mathbf{y}}_j) = (\mathbf{y}_i^T \mathbf{W}_\xi \dot{\mathbf{y}}_j)(\dot{\mathbf{u}}_i^T \mathbf{W}_x \mathbf{u}_j)$ since both $(\dot{\mathbf{u}}_i^T \mathbf{W}_x \mathbf{u}_j)$ and $(\mathbf{y}_i^T \mathbf{W}_\xi \dot{\mathbf{y}}_j)$ are scalars. By multiplying the above equation (A 2) by $\mathbf{W}_\xi^{-1}/2$ from the left and replacing $\mathbf{u}_k^T \mathbf{W}_x \mathbf{u}_j = \delta_{kj}$ and $\dot{\mathbf{u}}_j^T \mathbf{W}_x \mathbf{u}_k = \phi_{kj} = -\phi_{jk}$, we obtain

$$\dot{\mathbf{y}}_k = \dot{\mathbf{T}}^T \mathbf{W}_x \mathbf{u}_k + \mathbf{y}_j \phi_{jk}.$$

In the vectorized form the above equation can be expressed as

$$\dot{\mathbf{Y}} = \dot{\mathbf{T}}^T \mathbf{W}_x \mathbf{U} + \mathbf{Y} \boldsymbol{\phi}.$$

Appendix B. Proof of equivalence of reductions

Theorem B.1. *Suppose that $\{\mathbf{Y}, \mathbf{U}\}$ and $\{\tilde{\mathbf{Y}}, \tilde{\mathbf{U}}\}$ satisfy the evolution equations (2.7) and (2.8) with different choices of $\boldsymbol{\phi}$ functions denoted by $\boldsymbol{\phi} \in \mathbb{R}^{r \times r}$ and $\tilde{\boldsymbol{\phi}} \in \mathbb{R}^{r \times r}$, respectively. We also assume that the two reductions are initially equivalent, i.e. $\mathbf{U}(0) = \tilde{\mathbf{U}}(0) \mathbf{R}_0$ and $\mathbf{Y}(0) = \tilde{\mathbf{Y}}(0) \mathbf{R}_0$, where $\mathbf{R}_0 \in \mathbb{R}^{r \times r}$ is an orthogonal rotation matrix. Then $\{\mathbf{Y}, \mathbf{U}\}$ and $\{\tilde{\mathbf{Y}}, \tilde{\mathbf{U}}\}$ are equivalent for $t > 0$, with a rotation matrix $\mathbf{R}(t) \in \mathbb{R}^{r \times r}$ governed by the matrix differential equation $\dot{\mathbf{R}} = \mathbf{R} \boldsymbol{\phi} - \tilde{\boldsymbol{\phi}} \mathbf{R}$.*

Proof. We replace $\mathbf{U} = \tilde{\mathbf{U}}\mathbf{R}$ and $\mathbf{Y} = \tilde{\mathbf{Y}}\mathbf{R}$ into the evolution equations (2.7) and (2.8). The evolution of the DB can be written as

$$\begin{aligned}\dot{\mathbf{U}} &= \dot{\tilde{\mathbf{U}}}\mathbf{R} + \tilde{\mathbf{U}}\dot{\mathbf{R}} \\ &= (\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{W}_x)\dot{\mathbf{T}}\mathbf{W}_\xi\tilde{\mathbf{Y}}\mathbf{R}\mathbf{C}^{-1} + \tilde{\mathbf{U}}\mathbf{R}\dot{\phi}.\end{aligned}$$

Therefore,

$$\begin{aligned}\dot{\tilde{\mathbf{U}}} &= (\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{W}_x)\dot{\mathbf{T}}\mathbf{W}_\xi\tilde{\mathbf{Y}}\mathbf{R}\mathbf{C}^{-1}\mathbf{R}^T + \tilde{\mathbf{U}}(\mathbf{R}\dot{\phi}\mathbf{R}^T - \dot{\mathbf{R}}\mathbf{R}^T) \\ &= (\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{W}_x)\dot{\mathbf{T}}\mathbf{W}_\xi\tilde{\mathbf{Y}}\tilde{\mathbf{C}}^{-1} + \tilde{\mathbf{U}}\tilde{\phi},\end{aligned}$$

where $\tilde{\mathbf{C}} = \mathbb{E}[\tilde{\mathbf{Y}}^T\tilde{\mathbf{Y}}] = \mathbf{R}\mathbf{C}\mathbf{R}^T$ and, therefore, $\tilde{\mathbf{C}}^{-1} = \mathbf{R}\mathbf{C}^{-1}\mathbf{R}^T$, where \mathbf{C} and $\tilde{\mathbf{C}}$ are similar matrices and have the same eigenvalues. In the last equation we have used $\dot{\mathbf{R}} = \mathbf{R}\dot{\phi} - \dot{\phi}\mathbf{R}$ and made use of the identity $\mathbf{R}^{-1} = \mathbf{R}^T$. The evolution of the stochastic coefficients can be written as

$$\begin{aligned}\dot{\mathbf{Y}} &= \dot{\tilde{\mathbf{Y}}}\mathbf{R} + \tilde{\mathbf{Y}}\dot{\mathbf{R}} \\ &= \dot{\mathbf{T}}^T\mathbf{W}_x\dot{\tilde{\mathbf{U}}} + \tilde{\mathbf{Y}}\mathbf{R}\dot{\phi}.\end{aligned}$$

Therefore,

$$\begin{aligned}\dot{\tilde{\mathbf{Y}}} &= \dot{\mathbf{T}}^T\mathbf{W}_x\dot{\tilde{\mathbf{U}}} + \tilde{\mathbf{Y}}(\mathbf{R}\dot{\phi}\mathbf{R}^T - \dot{\mathbf{R}}\mathbf{R}^T) \\ &= \dot{\mathbf{T}}^T\mathbf{W}_x\dot{\tilde{\mathbf{U}}} + \tilde{\mathbf{Y}}\tilde{\phi}.\end{aligned}$$

■

References

- Hey T, Tansley S, Tolle K. 2009 *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research.
- van der Maaten LJP, Postma EO, van den Herik HJ. 2009 Dimensionality reduction: a comparative review. See https://lvdmaaten.github.io/publications/papers/TR_Dimensionality_Reduction_Review_2009.pdf.
- Noack BR. 2016 From snapshots to modal expansions—bridging low residuals and pure frequencies. *J. Fluid Mech.* **802**, 1–4. (doi:10.1017/jfm.2016.416)
- Marzouk Y, Xiu D. 2009 A stochastic collocation approach to Bayesian inference in inverse problems. *Commun. Comput. Phys.* **6**, 826–847. (doi:10.4208/cicp.2009.v6.p826)
- Xiu D, Karniadakis GE. 2002 The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* **24**, 619–644. (doi:10.1137/S1064827501387826)
- Branicki M, Majda AJ. 2012 Fundamental limitations of polynomial chaos for uncertainty quantification in systems with intermittent instabilities. *Commun. Math. Sci.* **11**, 55–103. (doi:10.4310/CMS.2013.v11.n1.a3)
- Wan X, Karniadakis GE. 2006 Long-term behavior of polynomial chaos in stochastic flow simulations. *Comput. Methods Appl. Mech. Eng.* **195**, 5582–5596. (doi:10.1016/j.cma.2005.10.016)
- Schmid PJ. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28. (doi:10.1017/S0022112010001217)
- Kutz J, Brunton S, Brunton B, Proctor J. *Dynamic mode decomposition: data-driven modeling of complex systems*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Crommelin DT, Majda AJ. 2004 Strategies for model reduction: comparing different optimal bases. *J. Atmos. Sci.* **61**, 2206–2217. (doi:10.1175/1520-0469(2004)061<2206:SFMRCDD>2.0.CO;2)
- Sapsis TP, Lermusiaux PFJ. 2009 Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D* **238**, 2347–2360. (doi:10.1016/j.physd.2009.09.017)
- Cheng M, Hou TY, Zhang Z. 2013 A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations I: derivation and algorithms. *J. Comput. Phys.* **242**, 843–868. (doi:10.1016/j.jcp.2013.02.033)

13. Choi M, Sapsis TP, Karniadakis GE. 2014 On the equivalence of dynamically orthogonal and bi-orthogonal methods: theory and numerical simulations. *J. Comput. Phys.* **270**, 1–20. (doi:10.1016/j.jcp.2014.03.050)
14. Babae H, Choi M, Sapsis TP, Karniadakis GE. 2017 A robust bi-orthogonal/dynamically-orthogonal method using the covariance pseudo-inverse with application to stochastic flow problems. *J. Comput. Phys.* **344**, 303–319. (doi:10.1016/j.jcp.2017.04.057)
15. Beck MH, Jäckle A, Worth GA, Meyer HD. 2000 The multiconfiguration time-dependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets. *Phys. Rep.* **324**, 1–105. (doi:10.1016/S0370-1573(99)00047-2)
16. Bardos C, Golse F, Gottlieb AD, Mauser NJ. 2003 Mean field dynamics of fermions and the time-dependent Hartree–Fock equation. *J. Math. Pures Appl.* **82**, 665–683. (doi:10.1016/S0021-7824(03)00023-0)
17. Koch O, Lubich C. 2007 Dynamical low-rank approximation. *SIAM J. Matrix Anal. Appl.* **29**, 434–454. (doi:10.1137/050639703)
18. Babae H, Sapsis TP. 2016 A minimization principle for the description of modes associated with finite-time instabilities. *Proc. R. Soc. A* **472**, 20150779. (doi:10.1098/rspa.2015.0779)
19. Babae H, Farazmand M, Haller G, Sapsis TP. 2017 Reduced-order description of transient instabilities and computation of finite-time Lyapunov exponents. *Chaos* **27**, 063103. (doi:10.1063/1.4984627)
20. Musharbash E, Nobile F. 2018 Dual dynamically orthogonal approximation of incompressible Navier Stokes equations with random boundary conditions. *J. Comput. Phys.* **354**, 135–162. (doi:10.1016/j.jcp.2017.09.061)
21. Exascale Mathematics Working Group. 2014 *Applied mathematics research for exascale computing*. Oakridge, TN: US Department of Energy. See <https://collab.cels.anl.gov/display/examath/Exascale+Mathematics+Home>.
22. Kuramoto Y, Tsuzuki T. 1976 Persistent propagation of concentration waves in dissipative media far from thermal equilibrium. *Prog. Theor. Phys.* **55**, 356–369. (doi:10.1143/PTP.55.356)
23. Sivashinsky GI, Michelson DM. 1980 On irregular wavy flow of a liquid film down a vertical plane. *Prog. Theor. Phys.* **63**, 2112–2114. (doi:10.1143/PTP.63.2112)
24. Sivashinsky GI. 1977 Nonlinear analysis of hydrodynamic instability in laminar flames—I. Derivation of basic equations. *Acta Astronaut.* **4**, 1177–1206. (doi:10.1016/0094-5765(77)90096-0)
25. Driscoll TA, Hale N, Trefethen LN. 2014 *Chebfun guide*. Oxford, UK: Pafnuty Publications.
26. Blanchard A, Mowlavi S, Sapsis T. 2018 Control of linear instabilities by dynamically consistent order reduction on optimally time-dependent modes. *Nonlinear Dyn.* **95**, 2745–2764. (doi:10.1007/s11071-018-4720-1)
27. Babae H. 2013 Uncertainty quantification of film cooling effectiveness in gas turbines. Master’s thesis, Louisiana State University, Baton Rouge, LA, USA.
28. Bagheri S, Schlatter P, Schmid PJ, Henningson DS. 2009 Global stability of a jet in crossflow. *J. Fluid Mech.* **624**, 33–44. (doi:10.1017/S0022112009006053)
29. Karniadakis GE, Sherwin SJ. 2005 *Spectral/hp element methods for computational fluid dynamics*. Oxford, UK: Oxford University Press.
30. Zhang D, Babae H, Karniadakis G. 2018 Stochastic domain decomposition via moment minimization. *SIAM J. Sci. Comput.* **40**, A2152–A2173. (doi:10.1137/17M1160756)
31. Chorin AJ, Hald OH, Kupferman R. 2002 Optimal prediction with memory. *Physica D* **166**, 239–257. (doi:10.1016/S0167-2789(02)00446-3)