

MATHEMATICAL CONSIDERATIONS ON RANDOMIZED ORTHOGONAL DECOMPOSITION METHOD FOR DEVELOPING TWIN DATA MODELS

DIANA A. BISTRIAN

ABSTRACT. This paper introduces the approach of Randomized Orthogonal Decomposition (ROD) for producing twin data models in order to overcome the drawbacks of existing reduced order modelling techniques. When compared to Fourier empirical decomposition, ROD provides orthonormal shape modes that maximize their projection on the data space, which is a significant benefit. A shock wave event described by the viscous Burgers equation model is used to illustrate and evaluate the novel method. The new twin data model is thoroughly evaluated using certain criteria of numerical accuracy and computational performance.

1. INTRODUCTION

The discovery of a realistic approximation of the complicated response of raw data by models of low complexity, i.e. reduced order models (ROM), has drawn the attention of researchers in recent years. This is conceivable given that these complex systems are dominated by a variety of underlying patterns with varying contributions in the reconstitution of the data [1]. ROM models have different computational properties based on the mathematical strategy used to create them. A twin data model is a substitute model whose primary function is to replicate the behavior of the original process. The key benefit of coupling the dynamical process with a simplified twin data model is to accurately map the dynamics to timeframes where it suffers from considerable changes and is therefore challenging to examine.

The most well-known methods for creating ROM models are currently proper orthogonal decomposition (POD), which is based on Fourier empirical decomposition and dynamic mode decomposition (DMD). The principles and disadvantages of these two methods are briefly presented below.

For purposes of reduced order modelling, numerous practitioners have embraced the POD technique, for exemplification see [2, 3, 4, 5, 6, 7, 8]. The intrusive model order reduction is usually derived by combining POD with Galerkin projection methods [9, 10]. Because the Galerkin projection is theoretically conducted by arduous computation and requires stabilizing procedures in the course of numerical implementation, this methodology has problems with efficiency and lacks stability [11, 12, 13, 14].

Understanding and refining the DMD technique has received a lot of attention, and various DMD variations have been made available, see [15, 16, 17, 18, 19, 20, 21]. Practitioners of modal decomposition frequently debate the choice of DMD modes to be employed for the flow reconstruction [22, 23]. The offline part of the procedure takes extra care and the CPU time is increased by specifying a DMD modes' selection criterion [24].

2010 *Mathematics Subject Classification.* 93A30, 70K75, 65C20.

Key words and phrases. Twin data model, Randomized orthogonal decomposition, Burgers model, Hopf-Cole-Transformation.

Another disadvantage of the DMD method is that it doesn't create orthogonal modes, necessitating a significantly large number of modes.

To overcome the aforementioned problems with the existing approaches, this research presents a novel technique for developing fluid dynamics twin data models. This paper introduces the algorithm of Randomized Orthogonal Decomposition (ROD) for creating twin data models having by definition the smallest error and the highest correlation in relation to the original data. It will be shown that a key advantage of ROD is the availability of orthonormal shape modes that maximize their projection on the data space in comparison with Fourier empirical decomposition.

The viscous Burgers equation model's description of a shock wave event is used to demonstrate and analyze the new algorithm's performance. The numerical results are presented for a detailed evaluation of the new twin data model utilizing particular criteria such as numerical accuracy and computational efficiency.

The remainder of the article is organized as follows. Section 2 discusses the mathematical aspects of the randomized orthogonal decomposition approach. Section 3 presents the test example with the exact mathematical solution. Section 4 presents ROD's twin data model and conducts a qualitative study of the model. A summary and conclusions are given in Section 5.

2. RANDOMIZED ORTHOGONAL DECOMPOSITION

Suppose that $D = [0, L] \subset \mathbb{R}$ represents the computational domain and let the Hilbert space $L^2(D)$ of square integrable functions on D

$$L^2(D) = \left\{ \phi : D \rightarrow \mathbb{R} \left| \int_D |\phi|^2 dx < \infty \right. \right\} \quad (1)$$

to be endowed with the inner product

$$\langle \phi_i(x), \phi_j(x) \rangle_{L^2(D)} = \int_D \phi_i(x) \phi_j(x) dx \quad \text{for } \phi_i, \phi_j \in L^2(D) \quad (2)$$

and the induced norm $\|\phi\|_{L^2(D)} = \sqrt{\langle \phi, \phi \rangle_{L^2(D)}}$ for $\phi \in L^2(D)$.

The data $u_i(x, t) = u(x, t_i)$, $t_i = i\Delta t$, $i = 0, \dots, N_t$, represent measurements at the constant sampling time Δt , x representing the Cartesian spatial coordinate.

The data matrix whose columns represent the individual data samples is called *the snapshot matrix*

$$V = [u_0 \quad u_1 \quad \dots \quad u_{N_t}] \in \mathbb{R}^{N_x \times (N_t+1)}. \quad (3)$$

Each column $u_i \in \mathbb{R}^{N_x}$ is a vector with N_x components, representing the spatial measurements corresponding to the $N_t + 1$ time instances.

Proposition 1. (Fourier Empirical Orthogonal Decomposition) *Let*

$$V = [u_0 \quad u_1 \quad \dots \quad u_{N_t}] \in \mathbb{R}^{N_x \times (N_t+1)}$$

be a real-valued data matrix of rank $r \leq \min(N_x, N_t + 1)$, whose columns $u_j \in \mathbb{R}^{N_x}$, $j = 1, \dots, N_t + 1$ are data snapshots. The Singular Value Decomposition (SVD) yields the factorization

$$V = \Psi \Sigma \Phi^T \quad (4)$$

of the matrix V , where $\Psi = [\psi_1, \dots, \psi_{N_x}] \in \mathbb{R}^{N_x \times N_x}$ and $\Phi = [\varphi_1, \dots, \varphi_{N_t+1}] \in \mathbb{R}^{(N_t+1) \times (N_t+1)}$ are orthogonal matrices,

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{N_x \times (N_t+1)},$$

with $D = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r \times r}$ and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$. Then $\{\psi_i\}_{i=1}^r$ and $\{\varphi_i\}_{i=1}^r$ represent the eigenvectors of VV^T and V^TV , respectively, with eigenvalues $\lambda_i = \sigma_i^2 > 0$ for $i = 1, \dots, r$, due to the relations

$$V\varphi_i = \sigma_i\psi_i \quad \text{and} \quad V^T\psi_i = \sigma_i\varphi_i \quad \text{for } i = 1, \dots, r.$$

It follows that

$$\begin{aligned} u_j &= \sum_{i=1}^r [D\Phi^T]_{ij}\psi_i = \sum_{i=1}^r \left[\begin{array}{c} \overbrace{\Psi^T\Psi}^{=I \in \mathbb{R}^{r \times r}} \quad D\Phi^T \\ \hline \end{array} \right]_{ij} \psi_i \\ &= \sum_{i=1}^r [\Psi^TV]_{ij}\psi_i = \sum_{i=1}^r \langle u_j, \psi_i \rangle_{L^2(D)} \psi_i. \end{aligned} \quad (5)$$

In the case of linearly-independent snapshots, the empirical orthogonal decomposition involves a number of terms equal to $r = \min(N_x, N_t + 1)$, which can be a large number. By convention, let suppose that $r = \min(N_x, N_t + 1) = N_x$. The following model

$$u^{\text{Fourier}}(x, t) = \sum_{i=1}^{N_x} a_i(t) \psi_i(x), \quad a_i(t) = \langle u, \psi_i \rangle_{L^2(D)} \quad (6)$$

is called *Fourier empirical orthogonal decomposition*.

It is more convenient to seek an orthonormal base of functions

$$\Phi = \{\phi_1, \phi_2, \dots\}, \quad \langle \phi_i(x), \phi_j(x) \rangle_{L^2(D)} = \delta_{ij}, \quad \|\phi\|_{L^2(D)} = 1, \quad (7)$$

where δ_{ij} is the Kronecker delta symbol, consisting of a minimum number of functions $\phi_i(x)$, such that the approximation of $u(x, t)$ through this base is as good as possible, in order to create a twin data model of reduced complexity.

The twin data model (DTM) at every time step $\{t_1, \dots, t_{N_t}\}$ is written according to the following relation:

$$u^{\text{DTM}}(x, t_i) = \sum_{j=1}^{N_{\text{DTM}}} \underbrace{a_j(t_i)}_{\text{Modal amplitudes}} \underbrace{\phi_j(x)}_{\text{Leading shape modes}}, \quad t_i \in \{t_1, \dots, t_{N_t}\}, \quad (8)$$

where $\phi_j \in \mathbb{C}$ represent the Φ base functions, which we call *the leading shape modes*, $N_{\text{DTM}} \ll \min(N_x, N_t + 1)$ represents the number of terms in the representation (8) which we impose to be minimal and $a_j(t_i)$ represent the modal growing amplitudes.

Determination of the optimal decomposition (8) then amounts to finding the solution to the following multiobjective constrained optimization problem:

$$\begin{aligned} \min_{\phi_j, a_j, N_{\text{DTM}}} & \sum_{i=1}^{N_x} \int_D \left\| u(x, t_i) - \sum_{j=1}^{N_{\text{DTM}}} a_j(t_i) \phi_j(x) \right\|_{L^2(D)}^2 dx, \\ & - \left\| u(x, t_i) \sum_{j=1}^{N_{\text{DTM}}} a_j(t_i) \phi_j(x) \right\|_{L^2(D)}^2 \\ \min_{\phi_j, a_j, N_{\text{DTM}}} & \sum_{i=1}^{N_x} \int_D \frac{\left\| u(x, t_i) \sum_{j=1}^{N_{\text{DTM}}} a_j(t_i) \phi_j(x) \right\|_{L^2(D)}^2}{\left\| u(x, t_i)^H u(x, t_i) \right\|_{L^2(D)} \left\| \left(\sum_{j=1}^{N_{\text{DTM}}} a_j(t_i) \phi_j(x) \right)^H \sum_{j=1}^{N_{\text{DTM}}} a_j(t_i) \phi_j(x) \right\|_{L^2(D)}^2} dx, \quad (9) \\ \text{s.t.} & \quad \langle \phi_i, \phi_j \rangle_{L^2(D)} = \delta_{ij}, \quad \|\phi_i\|_{L^2(D)} = 1, \quad 1 \leq i \leq j \leq N_{\text{DTM}} \end{aligned}$$

where H denotes the conjugate transpose of the snapshot containing the data.

Definition 1. (The Projection Operator)

Let

$$V_0 = [u_0 \quad u_1 \quad \dots \quad u_{N_t-1}] \equiv \{u_i^0\}_{i=0}^{N_t-1} \in \mathbb{R}^{N_x \times N_t}$$

be a real-valued data matrix, whose columns are data snapshots.

Let

$$\Phi = \{\phi_1, \phi_2, \dots, \phi_{N_{DTM}}\}$$

be the shape modes base.

We define the bounded projection operator $P_{V_0} \Phi$, that maps every shape mode $\{\phi_i\}_{i=1}^{N_{DTM}}$ onto its projection on the data vectors $\{u_i^0\}_{i=0}^{N_t-1}$ along the computational domain direction D :

$$P_{V_0} \Phi(\phi_i, u_j^0) \equiv P_{u_j^0} \phi_i = \left(\langle \phi_i, u_j^0 \rangle_{L^2(D)} / \langle u_j^0, u_j^0 \rangle_{L^2(D)} \right) u_j^0. \quad (10)$$

Proposition 2. (Randomized Singular Value Decomposition of rank k)

Let

$$V_0 = [u_0 \quad u_1 \quad \dots \quad u_{N_t-1}] \in \mathbb{R}^{N_x \times N_t}$$

be a real-valued data matrix, whose columns are data snapshots. If we impose a target rank $k < \min(N_x, N_t)$, the Randomized Singular Value Decomposition of rank k (k -RSVD) produces k left singular vectors of V_0 and has the following steps:

1. Generate a Gaussian random test matrix M of size $N_t \times k$.
2. Compute a compressed sampling matrix by multiplication of data matrix with random matrix $Q = V_0 M$.
3. Project the data matrix to the smaller space $P = Q^H V_0$, where H denotes the conjugate transpose.
4. Produce the economy-size singular value decomposition of low-dimensional data matrix $[T, \Sigma, W] = \text{svd}(P)$.
5. Compute the right singular vectors $U = QT$, $U \in \mathbb{R}^{N_x \times k}$, $\Sigma \in \mathbb{R}^{k \times k}$, $W \in \mathbb{R}^{N_t \times k}$.

Theorem 1. (Randomized Orthogonal Decomposition: ROD)

Let

$$\begin{aligned} V_0 &= [u_0 \quad u_1 \quad \dots \quad u_{N_t-1}] \equiv \{u_i^0\}_{i=0}^{N_t-1} \in \mathbb{R}^{N_x \times N_t}, \\ V_1 &= [u_1 \quad u_2 \quad \dots \quad u_{N_t}] \equiv \{u_i^1\}_{i=1}^{N_t} \in \mathbb{R}^{N_x \times N_t}, \end{aligned} \quad (11)$$

two time-shifted data matrices with rank $r \leq \min(N_x, N_t)$, whose columns are data snapshots. Then for $1 \leq N_{DTM} \leq r$, the optimization problem

$$\begin{aligned} \max_{\phi_1, \dots, \phi_{N_{DTM}}} & \sum_{i=1}^{N_{DTM}} \sum_{j=1}^{N_t} \left\| P_{u_j^0} \phi_i \right\|_{L^2(D)}^2 \\ \text{s.t.} & \langle \phi_i, \phi_j \rangle_{L^2(D)} = \delta_{ij}, \quad \|\phi_i\|_{L^2(D)} = 1, \quad 1 \leq i \leq j \leq N_{DTM} \end{aligned} \quad (12)$$

is being solved by the subspace span $\{\phi_1, \phi_2, \dots, \phi_{N_{DTM}}\}$ spanned by the sequence of orthonormal functions

$$\{\phi_i\}_{i=1}^{N_{DTM}} = (U, X_{\cdot, i})_{L^2(D)} / \|(U, X_{\cdot, i})\|_{L^2(D)} \quad (13)$$

where U represents the matrix of left singular vectors produced by Randomized Singular Value Decomposition of rank N_{DTM} of data matrix V_0 and X denotes the eigenvectors to the Koopman propagator operator \mathcal{A} , i.e. $u_{N_t} = \mathcal{A}^{N_t} u_0$.

Proof. Following the Koopman decomposition assumption [25], we consider that a propagator operator \mathcal{A} exists, that maps every column vector onto the next one, i.e.

$$\{u_0, u_1 = \mathcal{A}u_0, u_2 = \mathcal{A}u_1 = \mathcal{A}^2 u_0, \dots, u_{N_t} = \mathcal{A}u_{N_t-1} = \mathcal{A}^{N_t} u_0\}. \quad (14)$$

For a sufficiently long sequence of the snapshots, suppose that the last snapshot u_{N_t} can be written as a linear combination of previous N_t vectors, such that

$$u_{N_t} = c_0 u_0 + c_1 u_1 + \dots + c_{N_t-1} u_{N_t-1} + \mathcal{R}, \quad (15)$$

in which $c_i \in \mathbb{R}, i = 0, \dots, N_t - 1$ and \mathcal{R} is the residual vector. The following relations are true:

$$\{u_1, u_2, \dots, u_{N_t}\} = \mathcal{A}\{u_0, u_1, \dots, u_{N_t-1}\} = \{u_1, u_2, \dots, V_0 c\} + \mathcal{R}, \quad (16)$$

where $c = (c_0 \ c_1 \ \dots \ c_{N_t-1})^T$ is the unknown column vector.

Thus, the aim is to solve the following eigenvalue problem:

$$V_1 = \mathcal{A}V_0 = V_0 \mathcal{S} + \mathcal{R}, \quad (17)$$

where \mathcal{S} approximates the eigenvalues of \mathcal{A} when $\|\mathcal{R}\|_2 \rightarrow 0$. This is equivalent to solve the minimization problem:

$$\min_{\mathcal{S}} \mathcal{R} = \|V_1 - V_0 \mathcal{S}\|_2, \quad (18)$$

where $\|\cdot\|_2$ is the L_2 -norm of \mathbb{R}^{N_x} .

The solution to the minimization problem (18) is found in the following manner. Suppose that $r \leq \min(N_x, N_t)$. Then for $1 \leq N_{DTM} \leq r$, we identify the N_{DTM} -RSVD of V_0 , that yields the factorization:

$$V_0 = U \Sigma W^H, \quad (19)$$

of the snapshot matrix V_0 , where $U = [u_1, \dots, u_{N_{DTM}}] \in \mathbb{R}^{N_x \times N_{DTM}}$ and $W = [w_1, \dots, w_{N_{DTM}}] \in \mathbb{R}^{N_t \times N_{DTM}}$ are orthogonal matrices that contain the eigenvectors of $V_0 V_0^H$ and $V_0^H V_0$, respectively, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_{N_{DTM}}) \in \mathbb{R}^{N_{DTM} \times N_{DTM}}$ is a square diagonal matrix containing the singular values of V_0 and H means the conjugate transpose.

Relations $\mathcal{A}V_0 = V_1 = V_0 \mathcal{S} + \mathcal{R}, \|\mathcal{R}\|_2 \rightarrow 0$ and $V_0 = U \Sigma W^H$ yield:

$$\mathcal{A}U \Sigma W^H = V_1 = U \Sigma W^H \mathcal{S} \Rightarrow U^H \mathcal{A}U \Sigma W^H = U^H U \Sigma W^H \mathcal{S} \Rightarrow \mathcal{S} = U^H \mathcal{A}U.$$

From $\mathcal{A}U \Sigma W^H = V_1$ it follows that $\mathcal{A}U = V_1 W \Sigma^{-1}$ and hence $\mathcal{S} = U^H (V_1 W \Sigma^{-1})$.

As a consequence, the solution to the minimization problem (18) is the matrix operator

$$\mathcal{S} = U^H (V_1 W \Sigma^{-1}). \quad (20)$$

The eigenvalues and eigenvectors of \mathcal{S} will converge toward the eigenvalues and eigenvectors of the Koopman propagator operator \mathcal{A} as a direct result of solving the minimization problem (18), which improves overall convergence.

Let $X \in \mathbb{R}^{N_{DTM} \times N_{DTM}}$, $\Lambda \in \mathbb{R}^{N_{DTM} \times N_{DTM}}$ be the eigenvectors, respectively the eigenvalues of the data propagator matrix \mathcal{S} :

$$\mathcal{S}X = X\Lambda. \quad (21)$$

Let $\Phi = \text{span}\{\phi_1, \phi_2, \dots, \phi_{N_{DTM}}\}$ be the subspace spanned by the sequence of functions

$$\{\phi_i\}_{i=1}^{N_{DTM}} = \langle U, X_{:,i} \rangle_{L^2(D)} \quad (22)$$

where U represents the matrix of left singular vectors produced by Randomized Singular Value Decomposition of rank N_{DTM} of data matrix V_0 . It follows that

$$\langle \phi_i, \phi_j \rangle_{L^2(D)} = \delta_{ij}, \quad 1 \leq i \leq j \leq N_{DTM} \quad (23)$$

i.e., Φ forms an orthogonal base to the data space. The Φ base vectors maximize their projection on the data space and they represent the solution to the constrained optimization problem (9), therefore they produce the twin data model with the expression given by Eq.(8). □

Corollary 1. *The base vectors $\{\phi_i\}_{i=1}^{N_{DTM}}$ defined by Eq.(13), respectively their corresponding modal coefficients $\{a_i\}_{i=1}^{N_{DTM}} = \langle U, X_{:,i} \rangle_{L^2(D)}$, solve the multiobjective constrained minimization problem (9).*

Corollary 2. *If*

$$\|P_{V_0}\Phi\|_{L^2(D)}^2 = \frac{1}{N_{DTM}} \sum_{i=1}^{N_{DTM}} \sum_{j=1}^{N_t} \|P_{u_j^0}\phi_i\|_{L^2(D)}^2 \quad (24)$$

represents the mean squared sum of the norms of projections of shape modes $\{\phi_i\}_{i=1}^{N_{DTM}}$ produced by Randomized Orthogonal Decomposition on the data space V_0 , and

$$\|P_{V_0}\Psi\|_{L^2(D)}^2 = \frac{1}{N_x} \sum_{i=1}^{N_x} \sum_{j=1}^{N_t} \|P_{u_j^0}\psi_i\|_{L^2(D)}^2 \quad (25)$$

represents the mean squared sum of the norms of projections in the case of the Fourier empirical orthogonal modes $\{\psi_i\}_{i=1}^{N_x}$ on the same data space, then

$$\|P_{V_0}\Phi\|_{L^2(D)}^2 > \|P_{V_0}\Psi\|_{L^2(D)}^2. \quad (26)$$

This means that, for the purpose of twin modelling, the ROD shape modes defined by Eq.(13) are qualitatively superior to Fourier empirical modes.

3. MATHEMATICAL MODEL AND THE EXACT SOLUTION

The experimental data are provided by the simulation of the nonlinear viscous Burgers equation model:

$$\begin{cases} \frac{\partial}{\partial t} u(x, t) + \frac{\partial}{\partial x} \left(\frac{u(x, t)^2}{2} \right) = \nu \frac{\partial^2}{\partial x^2} u(x, t), & t > 0, \quad \nu > 0, \\ u(x, 0) = u_0(x), & x \in \mathbb{R}, \end{cases} \quad (27)$$

where $u(x, t)$ is the unknown function of time t , ν is the viscosity parameter. The initial condition of the following form is considered:

$$u_0(x) = -\sin(\pi x), \quad (28)$$

The homogeneous Dirichlet boundary conditions of the form

$$u(0, t) = u(L, t) = 0 \quad (29)$$

are also applied to the model.

The nonlinear evolution governed by the Burgers equation is obtained with the help of the Cole–Hopf transformation defined by:

$$u = -2\nu \frac{1}{\varphi} \frac{\partial \varphi}{\partial x}. \quad (30)$$

Through an analytical handling it is found that:

$$\frac{\partial u}{\partial t} = \frac{2\nu}{\varphi^2} \left(\frac{\partial \varphi}{\partial t} \frac{\partial \varphi}{\partial x} - \varphi \frac{\partial^2 \varphi}{\partial x \partial t} \right), \quad u \frac{\partial u}{\partial x} = \frac{4\nu^2}{\varphi^3} \frac{\partial \varphi}{\partial x} \left(\varphi \frac{\partial^2 \varphi}{\partial x^2} - \frac{\partial \varphi}{\partial x} \frac{\partial \varphi}{\partial x} \right), \quad (31)$$

$$\nu \frac{\partial^2 u}{\partial x^2} = -\frac{2\nu^2}{\varphi^3} \left(2 \left(\frac{\partial \varphi}{\partial x} \right)^3 - 3\varphi \frac{\partial^2 \varphi}{\partial x^2} \frac{\partial \varphi}{\partial x} + \varphi^2 \frac{\partial^3 \varphi}{\partial x^3} \right). \quad (32)$$

Substituting these expressions into (27) it follows that

$$\frac{\partial \varphi}{\partial x} \left(\frac{\partial \varphi}{\partial t} - \nu \frac{\partial^2 \varphi}{\partial x^2} \right) = \varphi \left(\frac{\partial^2 \varphi}{\partial x \partial t} - \nu \frac{\partial^3 \varphi}{\partial x^3} \right) = \varphi \frac{\partial}{\partial x} \left(\frac{\partial \varphi}{\partial t} - \nu \frac{\partial^2 \varphi}{\partial x^2} \right). \quad (33)$$

Relation (33) indicates that if φ solves the heat equation, then $u(x, t)$ given by the Cole-Hopf transformation (30) solves the viscid Burgers equation (27). Thus the viscid Burgers equation (27) is reduced to the following one:

$$\begin{cases} \frac{\partial \varphi}{\partial t} - \nu \frac{\partial^2 \varphi}{\partial x^2} = 0, & x \in \mathbb{R}, t > 0, \nu > 0, \\ \varphi(x, 0) = \varphi_0(x) = e^{-\int_0^x \frac{u_0(\xi)}{2\nu} d\xi}, & x \in \mathbb{R}. \end{cases} \quad (34)$$

Taking the Fourier transform with respect to x for both heat equation and the initial condition (34), the analytic solution is obtained in the following form:

$$\varphi(x, t) = \frac{1}{2\sqrt{\pi\nu t}} \int_{-\infty}^{\infty} \varphi_0(\xi) e^{-\frac{(x-\xi)^2}{4\nu t}} d\xi. \quad (35)$$

From the Cole-Hopf transformation (30) we obtain the analytic solution to the problem (27) in the following form:

$$u(x, t) = \frac{\int_{-\infty}^{\infty} \frac{x-\xi}{t} \varphi_0(\xi) e^{-\frac{(x-\xi)^2}{4\nu t}} d\xi}{\int_{-\infty}^{\infty} \varphi_0(\xi) e^{-\frac{(x-\xi)^2}{4\nu t}} d\xi}. \quad (36)$$

The exact solution (36) with the initial condition (28) is computed using the Gauss-Hermite Quadrature [26]. Gauss-Hermite quadrature approximates the value of integrals of the following kind:

$$\int_{-\infty}^{\infty} f(z) e^{-z^2} dz \approx \sum_{i=1}^n w_i f(x_i), \quad (37)$$

where n represents the number of sample points used, x_i are the roots of the Hermite polynomial $H_n(x)$ and the associated weights w_i are given by

$$w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 (H_{n-1}(x_i))^2}, \quad i = 1, \dots, n. \quad (38)$$

The initial condition (28) leads to

$$\varphi_0(x) = e^{-\frac{1}{2\nu} \int_0^x u_0(\xi) d\xi} = e^{-\frac{1}{2\nu} \int_0^x -\sin(\pi\xi) d\xi} = e^{\frac{1}{2\nu\pi}} \cdot e^{-\frac{\cos(\pi x)}{2\nu\pi}}. \quad (39)$$

The exact solution (36) is written in the following form:

$$u(x, t) = \frac{\int_{-\infty}^{\infty} \frac{x-\xi}{t} e^{-\frac{\cos(\pi\xi)}{2\nu\pi}} \cdot e^{-\left(\frac{x-\xi}{\sqrt{4\nu t}}\right)^2} d\xi}{\int_{-\infty}^{\infty} e^{-\frac{\cos(\pi\xi)}{2\nu\pi}} \cdot e^{-\left(\frac{x-\xi}{\sqrt{4\nu t}}\right)^2} d\xi}. \quad (40)$$

Introducing the variable change

$$z = \frac{x-\xi}{\sqrt{4\nu t}}, \quad (41)$$

the exact solution to the Burgers equation model (27) with the initial condition (28) is

$$u(x, t) = \frac{\int_{-\infty}^{\infty} 4\nu z e^{-\frac{1}{2\nu\pi} \cos[\pi(x-z\sqrt{4\nu t})]} e^{-z^2} dz}{\int_{-\infty}^{\infty} \sqrt{4\nu t} e^{-\frac{1}{2\nu\pi} \cos[\pi(x-z\sqrt{4\nu t})]} e^{-z^2} dz}. \quad (42)$$

The computational domain considered is $[0, L]$, where $L = 2$, the computational time is $[0, T]$, where $T = 3$, viscosity parameter in the Burgers equation model is $\nu = 10^{-2}$. The computational domain is uniformly discretized by using $N = 100$ grid points, which yields a mesh-size $\Delta x = 0.02$.

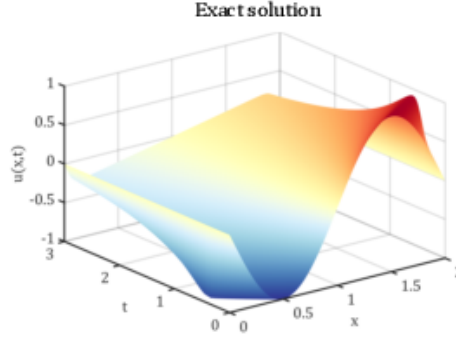


FIGURE 1. The exact solution of the viscid Burgers equation model (27), with initial condition (28)

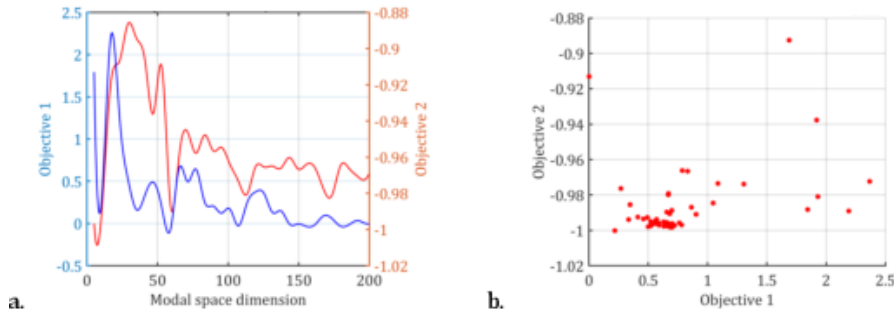


FIGURE 2. The objectives of the optimization problem (9) and the Pareto front solution obtained by genetic algorithm

Eq. (28), representing the initial condition, yields a sinusoidal pulse with an abrupt change of slope at the extremities of the domain. Figure 1 illustrates the exact solution computed with the technique of the Gauss-Hermite quadrature with $n = 100$ nodes.

4. THE TWIN DATA MODEL. QUALITATIVE ANALYSIS

As a high fidelity substitute model for the precise solution of the Burgers equation model under investigation here, the twin data model (8) is developed utilizing the randomized orthogonal decomposition (ROD).

The training data comprises of $N_t = 300$ total number of snapshots taken in time at regularly spaced time intervals $\Delta t = 0.01$, $N_x = 101$ number of spatial measurements per time snapshot.

The optimal dimension $N_{DTM} = 10$ of the leading shape modes space is determined as the solution to the multiobjective optimization problem with nonlinear constraints (9). ROD algorithm solves this problem and finds Pareto front of the two fitness functions using a genetic algorithm. Figure 2 illustrates the objectives of the optimization problem (9) and the Pareto front solution, for the considered test case.

Figure 3 illustrates the twin data model produced with ROD. The figure presents also the leading shape modes and the corresponding modal amplitudes that contribute to the assembly of the model.

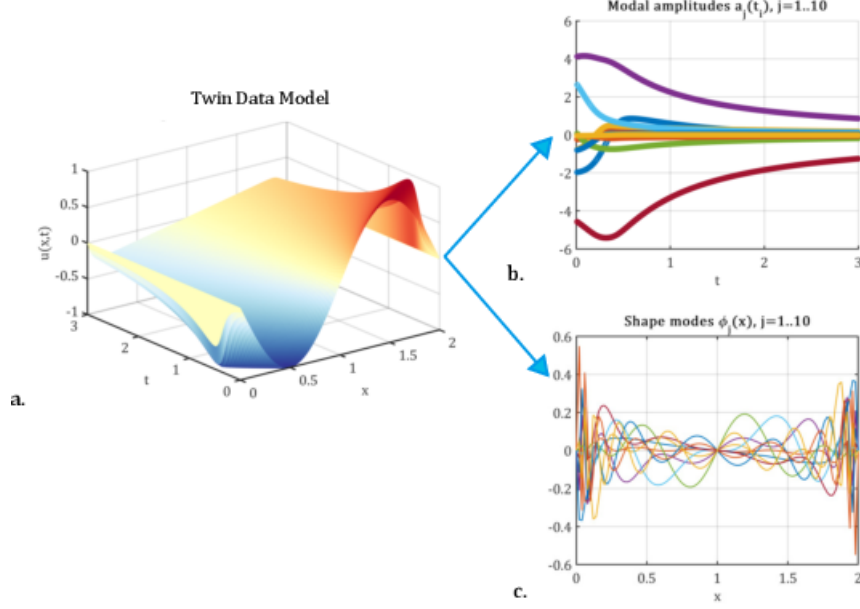


FIGURE 3. a.The twin data model as the solution of ROD algorithm; b.The modal amplitudes; c.The corresponding leading shape modes

We introduce $\langle \cdot \rangle_T$ as a time average operator over $[t_1, T]$ corresponding to the arithmetic time-average of equally spaced elements of the interval $[t_1, T]$:

$$\langle f(t) \rangle_T = \frac{1}{N_t} \sum_{i=1}^{N_t} f(t_i), \quad t_i \in \{t_1, t_2, \dots, t_{N_t} = T\}. \quad (43)$$

The absolute error between the exact solution and the twin data model is defined by relation:

$$Error^{DTM} = \langle \|u(x, t) - u^{DTM}(x, t)\|_2 \rangle_T, \quad t \in [t_1, T]. \quad (44)$$

The correlation coefficient is used to validate the quality of the twin data model over the exact solution and has the following expression:

$$Corr^{DTM} = \left\langle \frac{\|u(x, t) u^{DTM}(x, t)\|_2^2}{\|u(x, t)\|_2^2 \|u^{DTM}(x, t)\|_2^2} \right\rangle_T, \quad t \in [t_1, T]. \quad (45)$$

In order to perform a qualitative analysis of the twin data model, Table 1 presents the space dimension of the leading shape modes computed by the ROD, the absolute error given by Eq.(44) and the correlation coefficient given by Eq.(45) between the exact solution and the twin data model. Analysing the results presented in Table 1, it is obvious that the ROD algorithm creates a model that is perfectly correlated with the original data (i.e. $Corr^{DTM} = 1$), having the absolute error of order $\mathcal{O}(10^{-7})$.

Table 2 presents the mean squared sum of the projection norms and highlights that the ROD shape modes maximize their projection compared to the Fourier modes. It was demonstrated that the leading modes computed by Randomized Orthogonal Decomposition are qualitatively superior to the Fourier modes.

TABLE 1. Qualitative analysis of the twin data model

DTM complexity	DTM Absolute error	DTM Correlation
$N_{DTM} = 10$	8.7264×10^{-7}	1.0000

TABLE 2. Qualitative analysis of the shape modes projection

ROD Shape Modes Projection Norm	Fourier Modes Projection Norm
$\ P_{V_0} \Phi\ _{L^2(D)}^2 = 56.6294$	$\ P_{V_0} \Psi\ _{L^2(D)}^2 = 3.5029$

5. CONCLUSIONS

A topic of significant interest to data scientists has been the focus of the current study. The technique of Randomized Orthogonal Decomposition (ROD) was presented in this study in order to produce twin data models of lower complexity that accurately reflect the dynamics of fluid flows. Details of the mathematical framework were provided.

It has been demonstrated that when a multiobjective optimization problem is formulated, randomized orthogonal decomposition outperforms Fourier techniques and reduces the projection error. It was established that ROD-generated leading shape modes are qualitatively superior than Fourier empirical modes in the sense that ROD modes maximize their projection on the data space. The twin data model produced using the current method has an exact correlation to the original data. Traditional approaches that are substantially more computationally expensive and do not necessarily have high precision, such as adjoint model reduction, POD-DEIM, or Galerkin projection methods, can be effectively avoided by Randomized Orthogonal Decomposition. The suggested approach will be tested on two-dimensional datasets with applications from other domains in a subsequent study.

REFERENCES

- [1] Holmes, P., Lumley, J., Berkooz, G., *Turbulence, coherent structures, dynamical systems and symmetry*, Cambridge University Press, 1996.
- [2] Kaiser, E., Kutz, J., Brunton, S., *Sparse identification of nonlinear dynamics for model predictive control in the low-data limit*, Proceedings Of The Royal Society A, 474, 2018.
- [3] Xiao, D., Heaney, C., Fang, F., Mottet, L., Hu, R., Bistran, D., Aristodemou, D., Navon, I., Pain, C., *A domain decomposition non-intrusive reduced order model for turbulent flows*, Computers & Fluids, 182, 15-27, 2019.
- [4] Stefanescu, R., Hite, J., Cook, J., Smith, R., Mattingly, J., *Surrogate-based robust design for a non-smooth radiation source detection problem*, Algorithms, 12, 113, 2019.
- [5] Xiao, D., Fang, F., Heaney, C., Navon, I., Pain, C., *A domain decomposition method for the non-intrusive reduced order modelling of fluid flow*, Computer Methods In Applied Mechanics And Engineering, 354, 307-330, 2019.
- [6] Li, S., Kaiser, E., Laima, S., Li, H., Brunton, S., Kutz, J., *Discovering time-varying aerodynamics of a prototype bridge by sparse identification of nonlinear dynamical systems*, Physical Review E, 100, 2019.
- [7] Wang, Y., Ding, X., Hu, K., Fang, F., Navon, I., Lin, G., *Feasibility of DEIM for retrieving the initial field via dimensionality reduction*, Journal Of Computational Physics, 429, 110005, 2021.
- [8] Brunton, S., Kutz, J., *Data-driven science and engineering: Machine learning, dynamical systems and control*, Cambridge University Press, 2022.
- [9] Bistran, D., Navon, I., *An improved algorithm for the shallow water equations model reduction: Dynamic Mode Decomposition vs POD*, International Journal For Numerical Methods In Fluids, 78, 552-580, 2015.
- [10] San, O., Iliescu, T., *A stabilized proper orthogonal decomposition reduced-order model for large scale quasigeostrophic ocean circulation*, Advances In Computational Mathematics, 41, 1289-1319, 2014.
- [11] Champion, K., Brunton, S., Kutz, J., *Discovery of nonlinear multiscale systems: Sampling strategies and embeddings*, SIAM Journal On Applied Dynamical Systems, 18, 312-333, 2019.

- [12] Mauroy, A., Sootla, A., Mezic, I., *The Koopman Operator in Systems and Control: Theory, Numerics, and Applications*, Springer, 2019.
- [13] Ahmed, S., San, O., Bistrián, D., Navon, I., *Sampling and resolution characteristics in reduced order models of shallow water equations: Intrusive vs nonintrusive*, International Journal For Numerical Methods In Fluids, 92, 992-1036, 2020.
- [14] Iliescu, T., *ROM Closures and Stabilizations for Under-Resolved Turbulent Flows*, 2022 Spring Central Sectional Meeting, 2022.
- [15] Chen, K., Tu, J., Rowley, C., *Variants of dynamic mode decomposition: boundary condition, Koopman and Fourier analyses*, Nonlinear Science, 22, 887-915, 2012.
- [16] Tu, J., Rowley, C., Luchtenburg, D., Brunton, S., Kutz, J., *On dynamic mode decomposition: Theory and applications*, Journal Of Computational Dynamics, 1, 391-421, 2014.
- [17] Kutz, J., Fu, X., Brunton, S., *Multiresolution dynamic mode decomposition*, SIAM Journal On Applied Dynamical Systems, 15, 713-735, 2016.
- [18] Noack, B., Stankiewicz, W., Morzynski, M., Schmid, P., *Recursive dynamic mode decomposition of transient and post-transient wake flows*, Journal Of Fluid Mechanics, 809, 843-872, 2016.
- [19] Erichson, N., Brunton, S., Kutz, J., *Compressed dynamic mode decomposition for background modeling*, Journal Of Real-Time Image Processing, 16, 1479-1492, 2019.
- [20] Bistrián, D., Navon, I., *Randomized dynamic mode decomposition for nonintrusive reduced order modelling*, International Journal For Numerical Methods In Engineering, 112, 3-25, 2017.
- [21] Ahmed, S., Dabaghian, P., San, O., Bistrián, D., Navon, I., *Dynamic mode decomposition with core sketch*, Physics Of Fluids, 34, 066603, 2022.
- [22] Noack, B., Morzynski, M., Tadmor, G., *Reduced-Order Modelling for Flow Control*, Springer, 2011.
- [23] Tissot, G., Cordier, L., Benard, N., Noack, B., *Model reduction using Dynamic Mode Decomposition*, Comptes Rendus Mecanique, 342, 410-416, 2014.
- [24] Bistrián, D., Navon, I., *The method of dynamic mode decomposition in shallow water and a swirling flow problem*, International Journal For Numerical Methods In Fluids, 83, 73-89, 2017.
- [25] Koopman, B., *Hamiltonian systems and transformations in Hilbert space*, Proc. Nat. Acad. Sci., 17, 315-318, 1931.
- [26] Brass, H., Petras, K., *Quadrature Theory: The Theory of Numerical Integration on a Compact Interval*, American Mathematical Soc., 2011.

UNIVERSITY POLITEHNICA OF TIMISOARA
DEPARTMENT OF ELECTRICAL ENGINEERING AND INDUSTRIAL INFORMATICS
REVOLUTIEI NR.5, 331128 HUNEDOARA, ROMANIA
E-mail address: diana.bistrián@upt.ro