Non-intrusive reduced order modelling with least squares fitting on a sparse grid

Z. Lin[a], D. Xiao[b,c,*], F. Fang[b], C.C. Pain[b], I.M. Navon[d]

[a]*School of Mathematical Sciences, Zhejiang University*
No. 38 Zheda Road, Hangzhou, China, 310027.
[b]*Novel Reservoir Modelling and Simulation Group,*
Department of Earth Science and Engineering, Imperial College London,
Prince Consort Road, London, UK, SW7 2AZ.
[c]*China University of Geosciences, Wuhan, 430074, China*
[d]*Department of Scientific Computing, Florida State University, Tallahassee, FL, 32306-4120, USA*

Abstract

This article presents a non-intrusive reduced order model (NIROM) for general, dynamic partial differential equations. Based upon proper orthogonal decomposition (POD) and Smolyak sparse grid collocation, the method first projects the unknowns with full space and time coordinates onto a reduced POD basis. Then we introduce a new least squares fitting procedure to approximate the dynamical transition of the POD coefficients between subsequent time steps, taking only a set of full model solution snapshots as the input. Thus, the physics and numerics of the original PDE model is fully transparent to this methodology and its level of non-intrusiveness is improved compared to existing ROMs. Furthermore, we take adaptive measures to address the instability issue arising from reduced order iterations of the POD coefficients.

This model can be applied to a wide range of physical and engineering scenarios and we test it on a couple problems in fluid dynamics. It is demonstrated that this reduced order approach captures the dominant features of the high fidelity models with reasonable accuracy while the computation complexity is reduced by several orders of magnitude.

Keywords: Non-intrusive, Least Squares Fitting, POD, Smolyak sparse grid

*Corresponding author
Email address: dh.xiao@imperial.ac.uk (D. Xiao)

1. Introduction

In many areas of science and engineering, iterative computations and data assimilations for large-scale dynamical systems are often required to understand, to predict and/or to control various phenomena. Straightforward simulations of such problems may be very inefficient: Extensive resources are exhausted to produce intermediate results that have little significance due to different sources of parameter variations, errors and noise. Therefore, reduced order models (ROMs) have become prevalent thanks to their potential to achieve major speedup for standard numerical procedures. Their applicability relies on the presumption that the predominant physical mechanisms operate on a much lower dimensional space. A particular class of implementations are based on the proper orthogonal decomposition (POD) method whose variants have been successfully applied in a wide spectrum of research fields: air pollution dispersion[1], shallow water equations[2, 3], convective flows [4], ocean modelling [5, 3, 6, 7, 8, 9], 4-D variational data assimilation [10, 11, 12], neutron transport [13], fluid and structure interaction[14, 15], inversion [16] and molecular simulation [17].

Traditionally, a reduced order model implies re-derive a simpler physical model with additional restrictions such as homogenisation and parametrisation, or the substantial rewriting of the numerical schemes for the physical model which is often termed as "intrusiveness". The intrusive reduced order model (ROM) has been suffering from instability and non-linear inefficiency problems [18, 19, 7, 20, 21, 8, 22]. In addition, the dependency of the source code results in that the ROM is difficult to modify and implement[23].

To circumvent or avoid these issues, researchers found another way of solving the reduced order system: non-intrusive method. The non-intrusive method is popular in recent years since it is independent of the complex original dynamic system and it is therefore easy to implement even when the source code is not available. A number of non-intrusive reduced order modelling (NIROM) methods have been proposed. Walton et al. and Xiao et al. proposed a NIROM using radial basis function (RBF) method and proper othogonal decomposition (POD)[24, 25]. Noack [26] and Noori [27] used neural network to construct a NIROM. Xiao et. al. used a Smolyak sparse grid collocation method in which multidimensional, vector-valued Smolyak functions are used to replace the differential equations and to evolve the state variables [28]. This approach not only avoids the necessity of code modification but also circumvents the so-called curse of dimensionality where computational complexity grows exponentially as the dimension of the problem increases.

In this paper, we presented a new non-intrusive reduced order model method based on least square fitting and Smolyak sparse grid. The advantage of this method lies in it uses smolyak sparse grid method to lower the order of polynomial fit. As we know, high order polynomials may cause a poorer fitting since they are oscillatory between sample data points. This combination of Smolyak sparse grid and least square fitting induces a great potential in problems with high number of sample data points(many thousands of data points). Instead of traversing each nodal point in a Smolyak sparse grid with its coordinates as the input to a full model simulation and then interpolating with the collective output, we compute the coefficients of the Smolyak polynomials by solving a least squares fitting problem whose data are recycled from the generation of

the POD basis. We first properly decompose the temporally equidistant snapshots of the solution variables generated by the CFD software Fluidity [29]. Then we assemble the input data for the least square fitting by projecting the snapshots onto the reduced basis. Finally, a linear system of normal equations is solved producing the desired fit defined by its function values on a sparse grid, with which one can approximate the infinite dimensional evolution of physical variables with finite dimensional map for POD coefficients. Essentially, the physics and numerics of the original, full model are transparent to the reduced order model and it can be readily made into a universal black box that is compatible with arbitrary POD framework.

The structure of the paper is as follows. Section 2 reviews the general methodology of POD followed by previous results on non-intrusive ROMs with the introduction of a Smolyak sparse grid. Section 3 derives the least squares fitting problem and its solution for the non-intrusive modelling. Section 4 presents the stabilisation techniques we develops to address the instability issue we encounter with the direct application of the least squares approximation. Section 5 demonstrates the methods capabilities by solving two test problems in fluid dynamics. Finally in Section 6, a summary of conclusions and discussions of future work are presented.

2. POD-Based Reduced Order Modelling and Smolyak Sparse Grids

We consider a physical, dynamical system such as the Navier-Stokes equations which takes the form

$$\frac{\partial \boldsymbol{\psi}}{\partial t} = F(\boldsymbol{\psi}) \tag{1}$$

with appropriate initial and boundary conditions, where the vector-valued function $\boldsymbol{\psi}$, defined on a region in space, say $\Omega \subset \mathbb{R}^3$, and on the time interval $[0, T]$, is connected to a prescribed forcing, F, that is in general a complex, integro-differential operator derived from some physical laws.

2.1. Reduction via Proper Othogonal Decomposition

The solution $\boldsymbol{\psi}$ to the master equation (1) lives in an infinite-dimensional function space and the first step of our reduced order modeling is to project it onto a finite-dimensional subspace via proper orthogonal decomposition (POD). This procedure generates a set of basis functions that is constructed from a collection of snapshots that are taken at a number of time instances of the full model simulation. This basis serves as the axes of a coordinate system that one can represent and reconstruct functions in physical space with coefficient vectors.

In the formulation presented here snapshots of each component of the solution vector is individually recorded. Without loss of generality, we use the Navier-Stokes equations to illustrate where $\boldsymbol{\psi}$ is consisted of 3 velocity components, (u_x, u_y, u_z), and a pressure component, p. Given a computational mesh of $\mathcal{N}$ nodes, the s^{th} snapshot of $\boldsymbol{\psi}$ then contains four $\mathcal{N}-$vectors, denoted as Ψ_s^x, Ψ_s^y, Ψ_s^z, and Ψ_s^p respectively, with which we assemble four separate $\mathcal{N} \times S$ matrices as

$$\Psi^x = (\Psi_1^x, \Psi_2^x, \ldots, \Psi_S^x), \quad \Psi^y = (\Psi_1^y, \Psi_2^y, \ldots, \Psi_S^y) \tag{2}$$

and so on for the other two components, where S is the number of snapshots available. For the purpose of simplicity we will omit the superscripts denoting unknown components from here on since the following procedure is applied to each of the four matrices in an independent and identical manner.

Next we construct a normalised snapshot matrix $\widetilde{\Psi}$ with columns

$$\widetilde{\Psi}_k = \Psi_k - \frac{1}{S}\sum_{k=1}^{S}\Psi_k, \quad k = 1, 2, \ldots, S \tag{3}$$

to which we then apply a singular value decomposition (SVD) as

$$\widetilde{\Psi} = U\Sigma V^T. \tag{4}$$

Here the matrices $U \in R^{\mathcal{N}\times\mathcal{N}}$ and $V \in R^{S\times S}$ consist of the orthogonal vectors for $\widetilde{\Psi}\widetilde{\Psi}^T$ and $\widetilde{\Psi}^T\widetilde{\Psi}$, respectively and Σ is a diagonal matrix of size $\mathcal{N} \times S$ whose nonzero (positive) entries, arranged in a decreasing order, are the singular values of $\widetilde{\Psi}$ and we denote them as λ_k, $k = 1, 2, \cdots, S$.

Eventually, the proper orthogonal decomposition of $\widetilde{\Psi}$, namely, a reduced-order, orthonormal set of basis functions $\{\phi_k\}_{k=1}^{S}$ are retrieved by renormalising the projections of $\widetilde{\Psi}$ onto the column vectors of V as

$$\phi_k = \frac{\widetilde{\Psi}V_k}{\sqrt{\lambda_k}}, \qquad k = 1, 2, \ldots, S \tag{5}$$

Moreover, this basis can be optimised by keeping only the first P members which correspond to the largest P singular values, respectively. Note that these vectors are optimal in the sense that no other rank-P set of basis vectors can be closer to the snapshot matrix $\widetilde{\Psi}$ measured in the Frobenius norm by the Eckart–Young Theorem. This further reduction is especially accurate when there is a "scale separation" among the singular values at λ_P, namely,

$$\frac{\sum_{i=P+1}^{S}\lambda_i^2}{\sum_{i=1}^{S}\lambda_i^2} \ll 1 \tag{6}$$

With an optimal POD basis $\{\phi_k\}_{k=1}^{P}$ any variable ψ on a $\mathcal{N}$-node mesh can be represented by

$$\psi = \overline{\psi} + \sum_{j=1}^{P}\alpha_j\phi_j, \tag{7}$$

where $\alpha_j \equiv \alpha_j(t)$ denote the time-varying coefficients of the POD expansion and $\overline{\psi}$ is the time-independent mean of the ensemble of snapshots for the variable ψ which is analogous to the average in (3). This series expansion transforms the full physical space to the reduced order space and vice versa.

A standard Galerkin procedure can then be applied for the series expansion (7). Substituting the series into the master equation (1) and following by a first-order finite difference discretisation in time, we finally arrive at a ROM, namely, an iteration scheme for the reduced order POD coefficient at arbitrary time step in the form of

$$\alpha_k^{n+1} = f_k(\boldsymbol{\alpha}^n = (\alpha_1^n, \alpha_2^n, \cdots, \alpha_P^n)), \quad n = 1, 2, \cdots. \tag{8}$$

plus the starting values α_k^0 for $k = 1, 2, \cdots, P$ and the superscripts denote time discretisation. The central subject in reduced order modelling is therefore to find an efficient way to estimate the $\mathbb{R}^P \to \mathbb{R}$ functions f_k so one can compute the POD coefficients at arbitrary times and reconstruct the physical variables from them using the expansion (7).

2.2. Non-Intrusive Reduced Order Modelling (NIROM) on a Smolyak Sparse Grid

Xiao *et. al.* developed a non-intrusive algorithm to interpolate the transitional functions, i.e., $f_k's$ in (8), for the POD coefficients from their values on a Smolyak sparse grid [28]. In particular, instead of the standard, intrusive approach that formulates a linear system as a projected version of the master equation (1) that requires a major re-derivation of the full model, they approximated each f_k with a multidimensional polynomial interpolant. Additionally, to avoid the curse of dimensionality, which is the exponential growth of interpolation points entailing the increase of the POD basis size P, a Smolyak sparse grid was introduced whose number of nodes is only a polynomial function of its dimension size.

The Smolyak interpolant $\hat{f}_k^{P,\mu}(\boldsymbol{x})$ for any P-dimensional point $\boldsymbol{x} = (x_1, x_2, \cdots, x_P)$ with approximation level μ can expanded as a weighted sum of tensor product operators

$$\hat{f}_k^{P,\mu}(\boldsymbol{x}) = \sum_{|\boldsymbol{\ell}|=\max(\mu-P+1,0)}^{\mu} (-1)^{\mu-|\boldsymbol{\ell}|} \times \binom{P-1}{\mu-|\boldsymbol{\ell}|} \times (U^{\ell_1} \otimes \cdots \otimes U^{\ell_P})(\hat{f}_k^{P,\mu})(\boldsymbol{x}) \tag{9}$$

where the summation index $\boldsymbol{\ell} = (\ell_1, \cdots, \ell_P) \in \mathbb{N}^P$, $|\boldsymbol{\ell}| = \sum_{n=1}^{P} |\ell_n|$ traverses through the μ different levels of the sparse grid in each dimension and the operator

$$(U^{\ell_1} \otimes \cdots \otimes U^{\ell_P})(f_k)(\boldsymbol{x}) = \sum_{i_1=1}^{M(\ell_1)} \cdots \sum_{i_P=1}^{M(\ell_P)} \left(\hat{f}_k^{P,\mu}(x_{i_1}^{\ell_1}, ..., x_{i_P}^{\ell_P}) \times \prod_{j=1}^{P} \psi_{i_j}^{\ell_j}(x_j) \right) \tag{10}$$

is also a multidimensional sum weighted by the nodal values of the interpolant. Here the ℓ_j^{th} level 1D grid in the j^{th} dimension, $j = 1, 2, \cdots, P$, has the size

$$M(\ell_j) = \begin{cases} 2^{\ell_j} + 1, & \ell_j = 1, 2, \cdots, \mu; \\ 1, & \ell_j = 0, \end{cases} \tag{11}$$

and its i_j^{th} node is denoted by $x_{i_j}^{\ell_j}$. Finally, the last term on the right hand side of (10) is the product of one-dimensional, basis interpolating polynomials. For example, in the standard Lagrange form,

$$\psi_i^{\ell}(x) = \prod_{\substack{n=1 \\ n \neq i}}^{M(\ell)} \frac{x - x_n^{\ell}}{x_i^{\ell} - x_n^{\ell}}. \tag{12}$$

It is worth noting that the sparse grid, as the collection of all nodes, can be written as a union of μ nested product grids

$$\mathcal{H}^{P,\mu} = \bigcup_{|\boldsymbol{l}|=\max(\mu-P+1,0)}^{\mu} \mathcal{G}_{\boldsymbol{\ell}} = \bigcup_{|\boldsymbol{\ell}|=\max(\mu-P+1,0)}^{\mu} \left(\bigotimes_{i=1}^{P} \{x_k^{\ell_i}\}_{k=1}^{M(\ell_i)} \right). \tag{13}$$

The interpolation conditions simply implies

$$\hat{f}_k^{P,\mu}(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{\ell}} = (x_{i_1}^{\ell_1}, ..., x_{i_P}^{\ell_P})) = f_k(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{\ell}}), \quad k = 1, 2, 3, \cdots, P. \tag{14}$$

with which Xiao *et. al.* applied the following algorithm to obtain a non-intrusive reduced order model:

1. Perform a POD given snapshots of the solution to equation (1);
2. Generate a Smolyak sparse grid that covers the POD projections of solution snapshots onto the reduced order space;
3. For every node on the sparse grid, reconstruct an initial condition in physical space using the transform (7);
4. Evolve the full model for a short time, Δt, with the reconstructed initial conditions and transform the results back to the reduced coefficient space with (7);
5. Find the Smolyak interpolants for the reduced order model (8) using the formulas (9), (10) and (14) with the nodal data computed in the previous step.

3. Least Squares Fitting on a Smolyak Sparse Grid

The POD-based non-intrusive procedure circumvents the need to manipulate the full model (1) or its numerical implementation and introduces impressive computation economy with the use of Smolyak sparse grids. In this section, we will propose an alternative way to approximate $f_k(\boldsymbol{x})$, $k = 1, 2, \cdots, P$ using a least squares fitting approach that is independent from the full model except for an input set of solution snapshots. Effectively, Step 3 through 5 of the algorithm outlined at the end of Section 2.2 will be replaced and the reduced order model will take form of a least squares fit instead of a Smolyak interpolant.

Now we seek the optimal fit, denoted by $F_k^{P,\mu}$, $k = 1, \cdots, P$, from the family of admissible functions, Ω, namely, all $\mathbb{R}^P \to \mathbb{R}$ polynomials defined on the same Smolyak sparse grid as in Section 2.2. The least squares condition for data fitting demands that $F_k^{P,\mu}$ minimises the standard L^2 distance in $P-$dimensional space, that is,

$$\sum_{n=1}^{S-1} \left\| F_k^{P,\mu}(\boldsymbol{\alpha}^n) - \boldsymbol{\alpha}^{n+1} \right\|_2^2 = \min_{g \in \Omega} \sum_{n=1}^{S-1} \left\| g(\boldsymbol{\alpha}^n) - \boldsymbol{\alpha}^{n+1} \right\|_2^2. \tag{15}$$

Here the data set

$$\left\{ \left(\boldsymbol{\alpha}^n = (\alpha_1^n, \cdots, \alpha_P^n), \alpha_k^{n+1} = f_k(\boldsymbol{\alpha}^n) \right), n = 1, 2, \cdots, S-1 \right\} \tag{16}$$

contains duple pairs of POD projections of the solution at two subsequent time steps calculated by the full model and they are assumed to be connected by the transition relation (8).

The formulas (9) through (12) demonstrates how any polynomial function $g(\boldsymbol{x})$ on the sparse grid $\mathcal{H}^{P,\mu}$ is uniquely specified by its values on grid nodes. Therefore, we define a quadratic cost function of these nodal values as

$$J_k\left(\{ g(\boldsymbol{x}), \boldsymbol{x}^s \in \mathcal{H}^{P,\mu} \} \right) = \sum_{n=1}^{S-1} \left[g(\boldsymbol{\alpha}^n) - \alpha_k^{n+1} \right]^2. \tag{17}$$

Substituting the nodal expansion for $g(\boldsymbol{\alpha}^n)$ into the right hand side and set all the partial derivatives to be zero lead to a linear system, namely, the *normal equations*, whose solution consists of the nodal values of $F_k^{P,\mu}(\boldsymbol{x})$. Essentially, the series (9) and (10) can be re-organised into a single sum

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{x}^s \in \mathcal{H}^{P,\mu}} g(\boldsymbol{x}^s)\, w_{\boldsymbol{x}^s}(\boldsymbol{x}) \tag{18}$$

where $w_{\boldsymbol{x}^s}$ is the weighting function associated with the node $\boldsymbol{x}^s$. Consequently,

$$\frac{1}{2}\frac{\partial J_k}{\partial g(\boldsymbol{x}^*)}\bigg|_{g\equiv F_k^{P,\mu}} = \sum_{\boldsymbol{x}^s \in \mathcal{H}^{P,\mu}} F_k^{P,\mu}(\boldsymbol{x}^s) \times \left(\sum_{n=1}^{S-1} w_{\boldsymbol{x}^*}(\boldsymbol{\alpha}^n)\, w_{\boldsymbol{x}^s}(\boldsymbol{\alpha}^n) \right) - \sum_{n=1}^{S-1} w_{\boldsymbol{x}^*}(\boldsymbol{\alpha}^n)\, \alpha_k^{n+1} = 0 \tag{19}$$

for any node $\boldsymbol{x}^*$ on the grid $\mathcal{H}^{P,\mu}$. Equivalently, in matrix notation,

$$\begin{bmatrix} \langle w_{\boldsymbol{x}^1}, w_{\boldsymbol{x}^1}\rangle_s & \langle w_{\boldsymbol{x}^1}, w_{\boldsymbol{x}^2}\rangle_s & \langle w_{\boldsymbol{x}^1}, w_{\boldsymbol{x}^3}\rangle_s & \cdots \\ \langle w_{\boldsymbol{x}^2}, w_{\boldsymbol{x}^1}\rangle_s & \langle w_{\boldsymbol{x}^2}, w_{\boldsymbol{x}^2}\rangle_s & \langle w_{\boldsymbol{x}^2}, w_{\boldsymbol{x}^3}\rangle_s & \cdots \\ \langle w_{\boldsymbol{x}^3}, w_{\boldsymbol{x}^1}\rangle_s & \langle w_{\boldsymbol{x}^3}, w_{\boldsymbol{x}^2}\rangle_s & \langle w_{\boldsymbol{x}^3}, w_{\boldsymbol{x}^3}\rangle_s & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \cdot \begin{pmatrix} F_k^{P,\mu}(\boldsymbol{x}^1) \\ F_k^{P,\mu}(\boldsymbol{x}^2) \\ F_k^{P,\mu}(\boldsymbol{x}^3) \\ \vdots \end{pmatrix} = \begin{pmatrix} \langle w_{\boldsymbol{x}^1}, \mathbb{P}_k^+\rangle_s \\ \langle w_{\boldsymbol{x}^2}, \mathbb{P}_k^+\rangle_s \\ \langle w_{\boldsymbol{x}^3}, \mathbb{P}_k^+\rangle_s \\ \vdots \end{pmatrix} \tag{20}$$

in which $\{\boldsymbol{x}^1, \boldsymbol{x}^2, \boldsymbol{x}^3, \dots\}$ is an arbitrary ordering of the nodes of $\mathcal{H}^{P,\mu}$, the inner product

$$\langle f, g\rangle_s = \sum_{n=1}^{S-1} f(\boldsymbol{\alpha}^n)\, g(\boldsymbol{\alpha}^n) \tag{21}$$

and the projector $\mathbb{P}_k^+$ returns the k^{th} component of $\boldsymbol{\alpha}^{n+1}$.

Clearly, the coefficient matrix of the linear system (20), given that it is rank-full as long as the number of grid nodes is less than that of the snapshots, is symmetric and positive-definite so it can be solved numerically with standard techniques such as QR factorisation, successive over-relaxation and etc.

Repeating the above procedure for $k = 1, 2, \cdots, P$ we obtain the least square fitting approximations to the reduced order model (8). The fundamental advantage of this approach over the interpolation model is that the only thing we only need the decomposed solution snapshots, $\{\boldsymbol{\alpha}^n, n = 1, 2, \cdots, S\}$ from the full model without the necessity to run it. This separates the ROM from detailed derivation and simulation of the physics and the level of non-intrusiveness is improved, as well as the enormous gain in computation economy.

4. Iteration Stabilisation with Coefficient Damping

As we found out in preliminary results, straightforward iterations of formulas (8) using its least squares approximation reconstructed with the solution to the system (20) suffers from instability issues that often arise in reduced order modelling. In particular, depending on specific problems and parameters, some POD coefficients diverges to

infinity after a random number of iterations when we use the reduced order modelling to evolve the dynamical system.

While the numerical and physical origins of this instability seem to depend on the specifics of the full model and are therefore beyond the scope of this paper, we next offer a simple and effective remedy by supplementing the fitting algorithm with a *prediction-correction* step: Once we detect that an output component of the model (8) diverges from the Smolyak grid that covers the dataset, we enforce a damping to this component. That is, with the tilde denoting the fitting approximations and the starting value $\tilde{\boldsymbol{\alpha}}^1 = \boldsymbol{\alpha}^1$, one iteration of this ROM consists of two parts:

$$\text{Prediction:} \qquad \hat{\alpha}_k^{n+1} = F_k^{P,\mu}(\tilde{\boldsymbol{\alpha}}^n); \tag{22a}$$

$$\text{Correction:} \qquad \tilde{\alpha}_k^{n+1} = \begin{cases} \hat{\alpha}_k^{n+1}, & \alpha_k^{\min} < \hat{\alpha}_k^{n+1} < \alpha_k^{\max}; \\ \delta\, \alpha_k^{\min}, & \hat{\alpha}_k^{n+1} \le \alpha_k^{\min}; \\ \delta\, \alpha_k^{\max}, & \hat{\alpha}_k^{n+1} \ge \alpha_k^{\max} \end{cases} \tag{22b}$$

for $n = 1, 2, \cdots$, $k = 1, 2, \cdots, P$ and the admissible range of $\tilde{\alpha}_k^{n+1}$ is pre-determined from the data as

$$\alpha_k^{\min} = \min_{i=1}^{S} \alpha_k^i, \qquad \alpha_k^{\max} = \max_{i=1}^{S} \alpha_k^s. \tag{23}$$

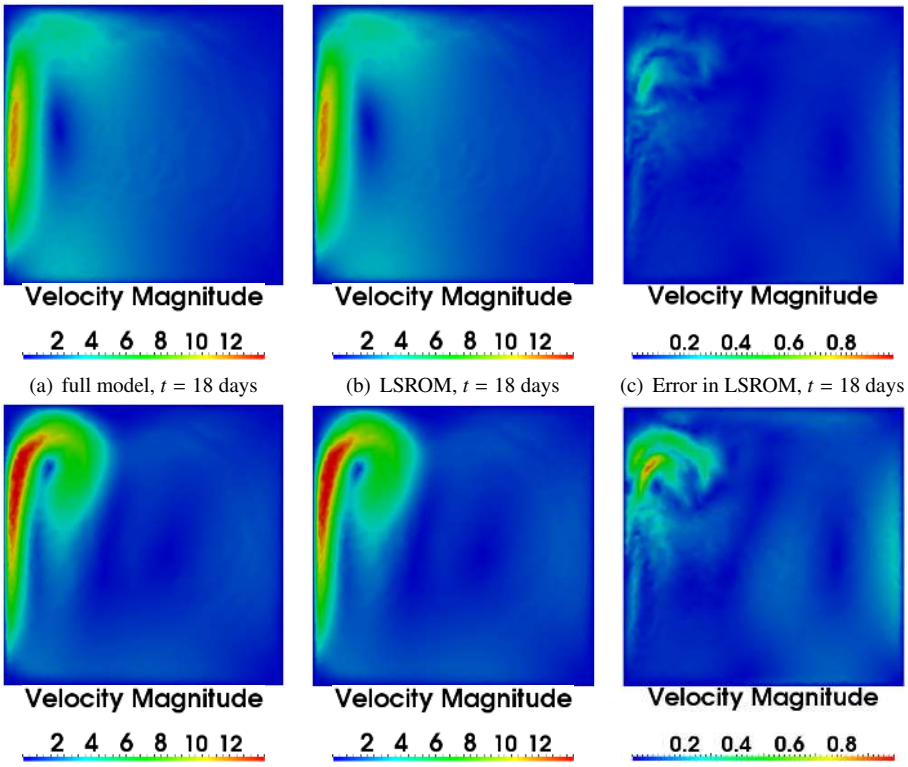
Here we have introduced a variable damping parameter, $\delta \in (0, 1]$ and we will see in the numerical example that an empirical choice of $\delta \in (0.8, 1.0]$ usually yields results with reasonable accuracy.

Notice that this filter only affects how the reduced order model substitutes the full model, while the derivation of the model itself, and its non-intrusive features, remained unaltered. There are various alternative options that may achieve regularity such as limiting the scope of fitting data (e.g., Finite Impulse Response filter), putting more weights on the data with small perturbations / gradients (e.g., weighted least squares) and etc. However, for the sake of simplicity we elect to postpone the exploration toward these directions to future work.

5. Numerical Examples

In this section we demonstrate the performance of the POD-based, non-intrusive reduced order scheme with least squares fitting by modelling a gyre flow and the flow past a cylinder that produces a von Kármán vortex street with two different Reynolds numbers. For both test cases, the solutions from the fidelity full model, which is simulated at the platform of Fluidity, serve as the reference, exact values for model comparisons as well as the snapshots for the POD basis generation. The error analysis was carried out with the same metrics used by Xiao *et. al.* [28], namely, the root mean square error (RMSE) and correlation coefficient on the $\mathcal{N}$-node finite element, physical mesh. For example, the RMSE at the n^{th} time step is defined by

$$\text{RMSE}^n = \frac{\|\psi^n_{\text{NIROM}} - \psi^n_f\|_2}{\sqrt{\mathcal{N}}}. \tag{24}$$



(a) full model, $t = 18$ days (b) LSROM, $t = 18$ days (c) Error in LSROM, $t = 18$ days

(d) full model, $t = 36$ days (e) LSROM, $t = 36$ days (f) Error in LSROM, $t = 36$ days

Figure 1: The gyre flow computed at time instances 18 (left) and 36 days (right). The panels (a) and (d) are full model simulations. The panels (b) and (e) are LSROM approximations with 3 POD basis function for each variable. The panels (c) and (f) are the differences between the full and LSROM models.

where ψ^n_{ROM} and ψ^n_f are the nodal snapshots of the NIROM solution and of the full solution on the full mesh, respectively and $\|\cdot\|_2$ is the standard vector 2-norm. In all simulations, the ROM operates on a one level Smolyak sparse grid, that is, $\mu = 1$ in Eq. (15), as increasing the number of grid level is found to provide negligible accuracy gain while the growing grid size demands doubling or quadrupling the snapshot data to guarantee the wellposedness of the linear system (20).

5.1. Problem 1: Gyre flow

We first test the least squares NIROM (LSROM) on the simulation of a shallow-water gyre flow. The geophysical scenario depicts a fluid circulating within a rectangular domain of the size $L_x \times L_y = 1000\text{km} \times 1000\text{km}$. The flow is driven by the uni-directional surface wind stress prescribed by

$$\tau_y = \tau_0 \cos(\pi y / L_y) \quad \text{and} \quad \tau_x = 0 \tag{25}$$

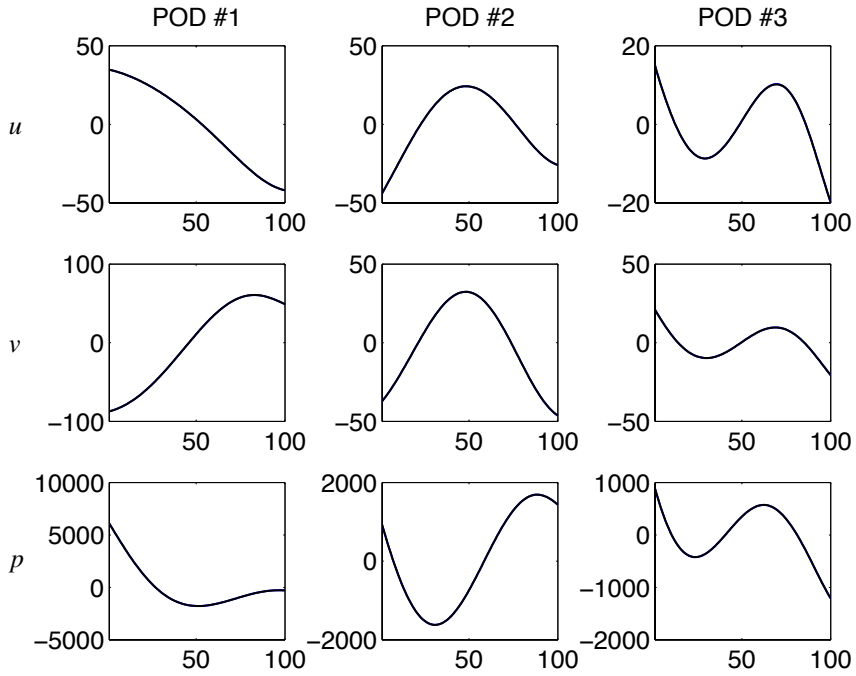


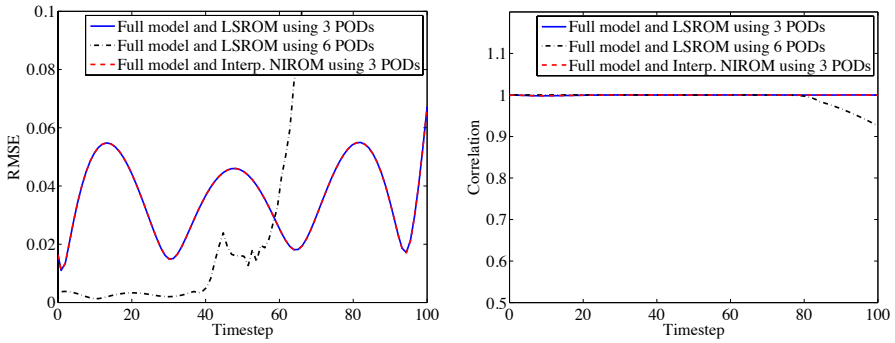
Figure 2: Gyre flow: The evolutions of POD coefficients for 3×3 POD basis functions. In each of the figures above, the horizontal axis is the timestep and the vertical axis is the value of the POD coefficients; the black, solid line is the projected POD coefficient of the full model solution and the blue, dotted line is its LSROM approximation. In these figures two curves are indistinguishable from each other.

where the maximum zonal stress $\tau_0 = 0.1$ N m^{-1}. In the full model, we address the Coriolis effect with the beta-plane approximation by setting $\beta = 1.8 \times 10^{-11}$ m^{-1} s^{-1} and we adopt the baseline fluid density $\rho_0 = 10^3$ kg m^{-3}.

In this case, the high-fidelity full model was generated on a 2823-node finite element mesh for a duration of $T = 36.5$ days using a time step of $\Delta t = 0.365$ days. Accordingly we record 100 snapshots of the solution from which only 3 POD basis functions for each variable (velocities u, v and pressure p) are extracted using the procedure described in Section 2.1. Then with the least square fitting method outlined in Section 3, we compute the nodal values of the transition functions (8) on the Smolyak grid determined by Step 2 at the end of Section 2.2. And finally, we iterate the reduced order model for each time step and reconstruct the LSROM-approximated solution on the physical mesh with expansion formula (7).

The comparison between the full model and the LSROM approximation is illustrated in Figure 1, in which we plotted the velocity magnitudes at two time instances, $t = 18$ days (the 50^{th} time step) and $t = 36$ days (the 98^{th} time step), as well as the difference between full model and LSROM at these times. It is clear that the dominant features of the flow field is accurately resolved by the reduced order model.

On the other hand, we plot the evolution of the POD coefficients in the reduced



(a) Evolution of RMSE (b) Evolution of the correlation coefficient

Figure 3: Gyre flow: The error diagnostics of the LSROM approximation with the reference of the interpolation-based NIROM.

space in Figure 2. Here the 3 POD coefficients for each of the three variables (a total of 9) are plotted for all times in which the full model projections are virtually indistinguishable from the NIROM approximations. POD #1 is the coefficient associated with the leading POD basis function for each variable, or equivalently, α_1 in the series expansion (7) and so on. This figure also confirms that new ROM is a good approximation to the full model.

Quantitatively, Figure 3 displays evolutions of the RMSE and of the correlation coefficient between the full model and the non-intrusive model, respectively. Here we plotted the results for two gyre simulations of LSROM, one with 3 POD basis functions and the other with 6, and the result by the interpolation-based NIROM with 3 POD basis functions. Together they further verify that the ROM approximation achieves an overall good agreement with full model, especially for the LSROM with only 3 POD bases for which the results are essentially equivalent to the interpolation based NIROM [28] under the same setting. It should be noted that in these results the damping parameter for stabilisation $\delta = 1.0$ applied according to Section 4 and this choice is optimal in the sense that the temporal maximum of the associated RMSE is minimised.

However, we notice an intrigue phenomenon: We are able to increase the accuracy, indicated by smaller RMSE and larger correlation coefficient, with a doubled number of POD bases (from 3 to 6). However, the improvement only lasts until the 57^{th} time step after which the RMSE soars and the correlation coefficient diverges from unity significantly after the 80^{th} time step. This demonstrates another aspect of the stability issue we encountered as more POD modes potentially promote instability. An intuition suggests that the fitting nature of the model necessitates more snapshot data (eg., more frequently sampling or equivalently, smaller Δt) to maintain a certain accuracy level as we also see in exploratory simulations. Due to the scope of this paper we elect to study the detailed statistical relationship between the sizes of the POD decomposition and of the snapshot matrix in this LSROM framework to future work. For example, one would want to determine *a priori* the optimal number of POD bases used given a fixed number of snapshots and vice versa.

Furthermore, we compare the computation costs by different models in Table 1

where we list the CPU times consumed by different stages of different models at each time step. We observed that although the previous interpolation-based NIROM and the new fitting-based LSROM share the POD projection costs and both methods achieve great speed-ups from the full model, the cost of the fitting step in LSROM is negligible compared to that of the interpolations in Xiao *et. al.*. Since the fitting solves a model-independent linear system (20) and the interpolations always run the full model for a short period of time, it is to be expected that the LSROM with improved non-intrusiveness will possess even bigger advantage for more complicated scenarios.

Model	Assembling and Solving	Projecting	Interpolation	Fitting	**Total**
Full Model	1.0070	0.0000	0.0000	0.0000	1.0070
Interp. NIROM	0.0000	0.0040	0.0020	0.0000	0.0060
LSROM	0.0000	0.0040	0.0000	0.0001	0.0041

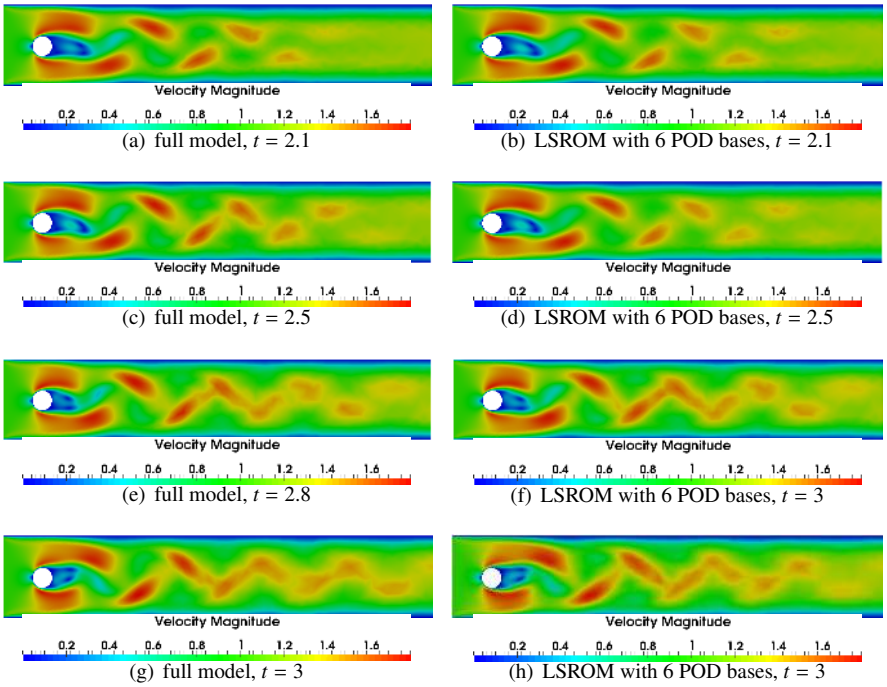
Table 1: Gyre flow: Comparison of the normalised CPU consumption by different models at each time step.

5.2. Problem 2: Flow past a cylinder

The second case on which we test the newly proposed LSROM method is also a classical example in computational fluid dynamics: An viscous inlet flow passes through a rectangular channel section that contains a cylinder and the oscillating downstream flow shed a street of von Kámán vortices. The simulation domain here is 2×0.4 in non-dimensional units and a cylinder of radius 0.12 is located at the point $(0.2, 0.2)$. The uniform upstream flow enters the domain from the left edge with unit speed rightward and it is slightly compressible. Typical outflow condition is applied at the right edge while the fluid is required to have no slip and no outward flow at the upper and lower edges of the channel. Finally, Dirichlet boundary conditions are enforced at the cylinder wall. We run the model under two settings of Reynolds number, $Re = 400$ and $Re = 3600$, to verify the general applicability of the method.

First, we run the high-fidelity full model with a prescribed Reynolds number $Re = 400$ on a 3213-node finite element mesh. Then we record the u, v and p solution variables between $t \in [2, 3]$ at regularly spaced time instances with $\Delta t = 0.01$ and a total of 100 snapshots are obtained from which we generated 6 POD basis functions for each of the variables followed by the LSROM approximation procedure.

Figure 4 through 6 illustrates how the least squares NIROM resolves the full model in this case. Visual inspection of Figure 4 confirms the satisfactory performance of the LSROM for this problem by reproducing the location, the size and the magnitude of almost all the vortices. Figure 5 displays the error distribution of the ROM and it is general an order of magnitude smaller than the exact values. Figure 6 documents the evolution of the POD coefficients of the full model and of the LSROM. Additionally, the error diagnostics for this problem are shown in Figure 7 with a comparison to the results obtained by the interpolation-based NIROM using the same number of POD bases. Here the optimal damping parameter δ for the LSROM is chosen to be 0.94. We see that for this test problem, the interpolation-based NIROM deviates from the



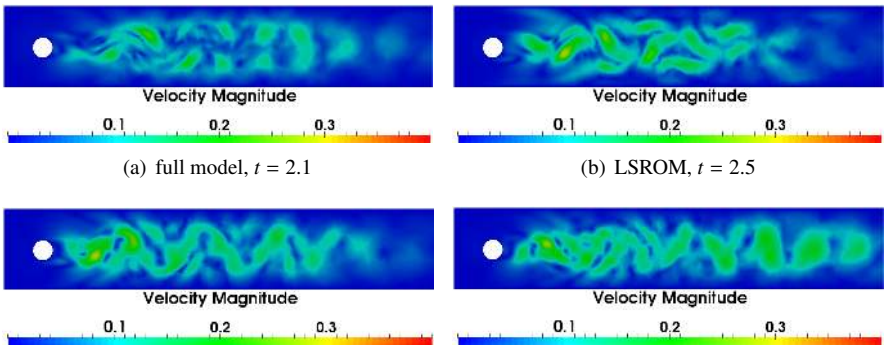
(a) full model, $t = 2.1$ (b) LSROM with 6 POD bases, $t = 2.1$

(c) full model, $t = 2.5$ (d) LSROM with 6 POD bases, $t = 2.5$

(e) full model, $t = 2.8$ (f) LSROM with 6 POD bases, $t = 3$

(g) full model, $t = 3$ (h) LSROM with 6 POD bases, $t = 3$

Figure 4: Flow past a cylinder at $Re = 400$. The figures above compare the full model (figures (a), (c), (e) and (g)) and the LSROM with 6 POD basis functions (figures (b), (d), (f) and (h)) at $t = 2.1, 2.5, 2.8$ and 3.0.



(a) full model, $t = 2.1$ (b) LSROM, $t = 2.5$

(c) full model, $t = 2.8$ (d) LSROM, $t = 3$

Figure 5: Flow past a cylinder at $Re = 400$: The difference between full model and LSROM with 6 POD basis functions. $Re = 400$.

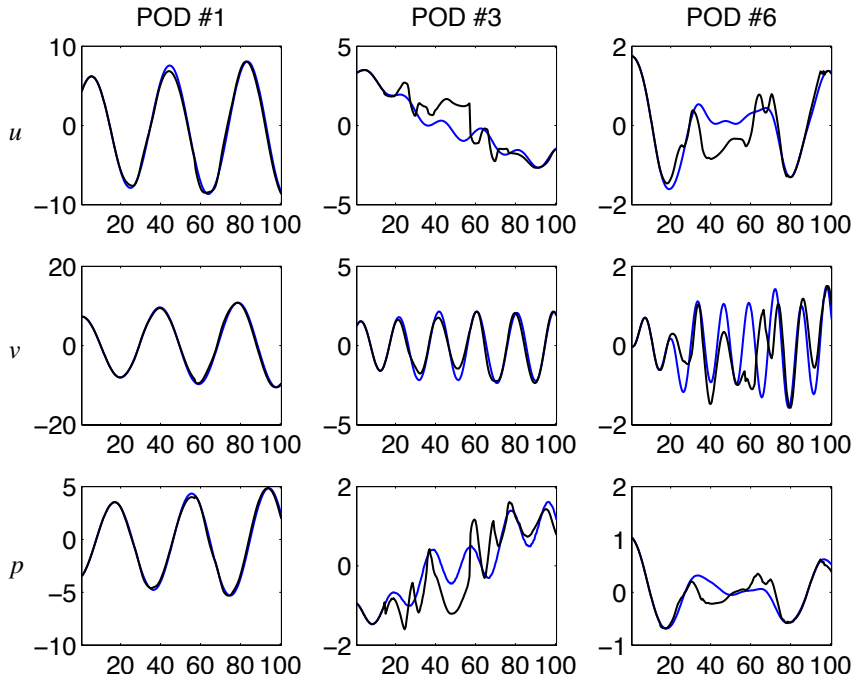
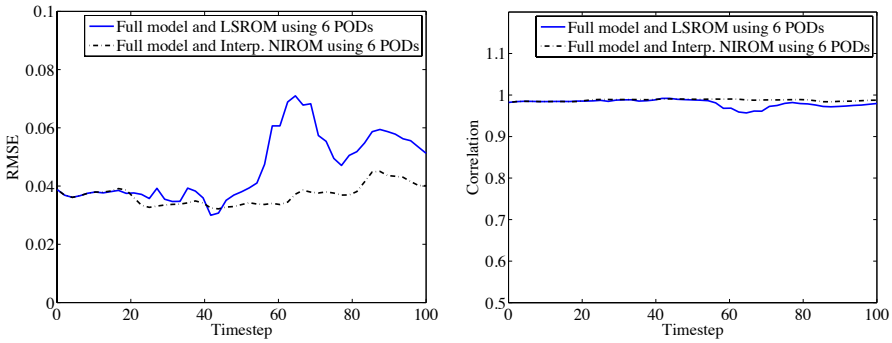


Figure 6: Flow past a cylinder at $Re = 400$: The evolutions of POD coefficients for 9 POD basis functions. In each of the figures above, the horizontal axis is the timestep and the vertical axis is the value of the POD coefficients; the black, solid line is the projected POD coefficient of the full model solution and the blue, dotted line is its LSROM approximation.

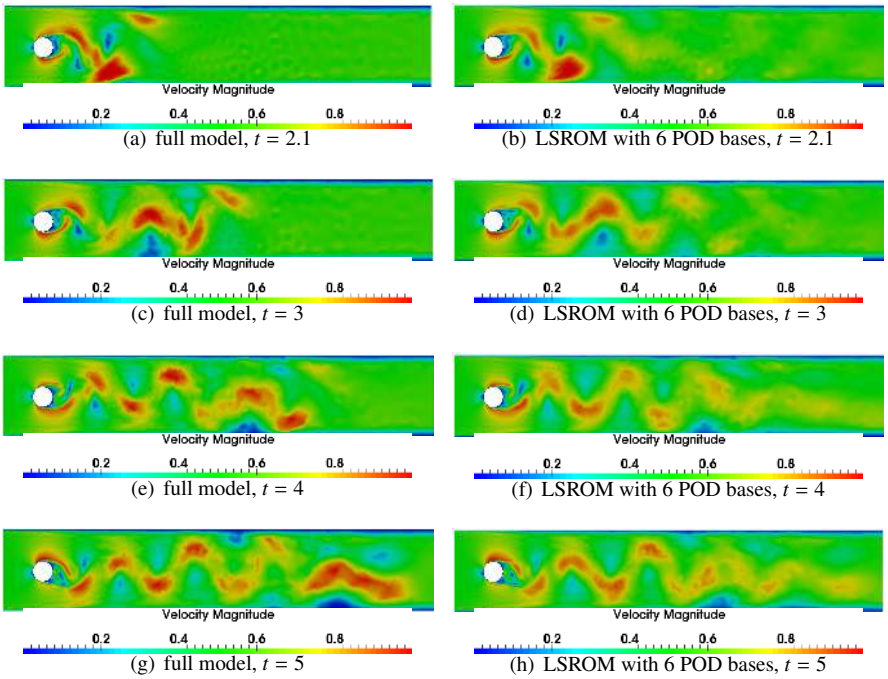
LSROM and it performs slightly better after the midpoint in the temporal axis but both methods are accurate approximations.

Next we increase the level of turbulence of the flow by setting $Re = 3600$ and we keep track of the simulation for a longer duration for $t \in [2, 6]$ with $\Delta t = 0.02$. All the other physical and numerical parameters remain the same. The corresponding results are demonstrated in Figure 8 through 10. Here the optimal damping parameter δ for the LSROM is chosen to be 0.84. Although the LSROM still qualitatively resolves the main flow structures but the accuracy is apparently reduced compared to the previous two test cases, the gyre flow and the flow past a cylinder at a lower Reynolds number. This should not be surprising as the spatial and temporal gradients of the variable are significantly magnified in this case and are thus harder to capture. These results can no doubt be improved by tuning the numerical parameters, such as the number of POD bases and of the snapshots used in the fitting, but the overall agreement achieved by this very restricted set of POD basis is still worth reporting. And as we mentioned before, we will postpone the investigation in this aspect to future study.



(a) Evolution of RMSE (b) Evolution of the correlation coefficient

Figure 7: Flow past a cylinder at $Re = 400$: The error diagnostics of LSROM with 6 POD bases in comparison with the interpolation-based NIROM with the same number of POD bases.



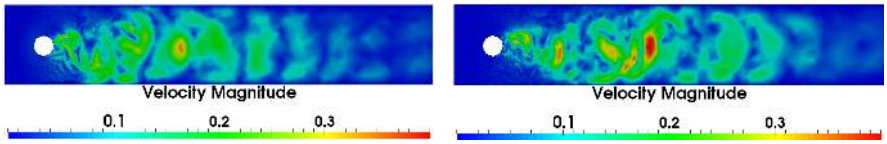
(a) full model, $t = 2.1$ (b) LSROM with 6 POD bases, $t = 2.1$

(c) full model, $t = 3$ (d) LSROM with 6 POD bases, $t = 3$

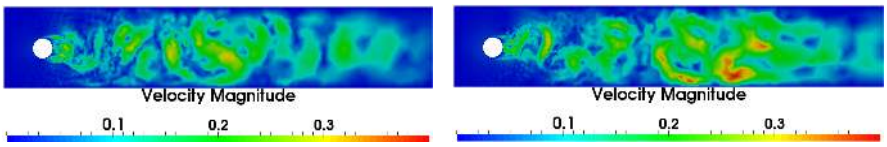
(e) full model, $t = 4$ (f) LSROM with 6 POD bases, $t = 4$

(g) full model, $t = 5$ (h) LSROM with 6 POD bases, $t = 5$

Figure 8: Flow past a cylinder at $Re = 3600$. The figures above compare the full model (figures (a), (c), (e) and (g)) and the LSROM with 6 POD basis functions (figures (b), (d), (f) and (h)) at $t = 2.1, 2.5, 2.8$ and 3.0.

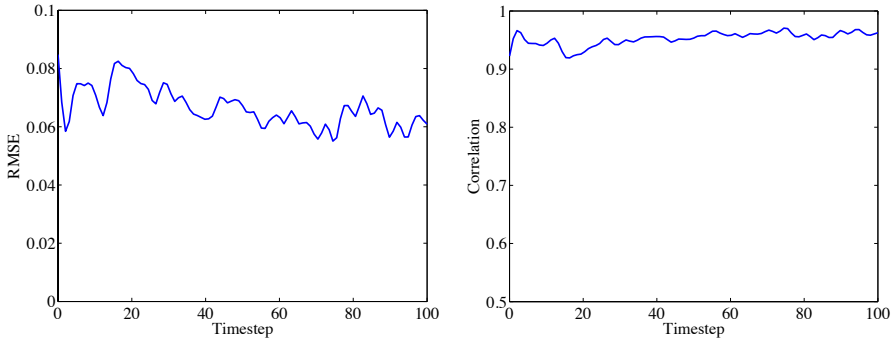


(a) full model, $t = 2.1$ (b) LSROM, $t = 2.5$



(c) full model, $t = 2.8$ (d) LSROM, $t = 3$

Figure 9: Flow past a cylinder at $Re = 3600$: The error diagnostics of LSROM with 6 POD bases.



(a) Evolution of RMSE (b) Evolution of the correlation coefficient

Figure 10: Flow past a cylinder at $Re = 3600$: The error diagnostics of LSROM with 6 POD bases.

6. Conclusions and Discussions

In this article a new non-intrusive reduced order model is introduced. The key ingredient here is a least squares fitting procedure to approximate the reduced order model formulated by standard POD routines that project the full model solution onto a finite-dimensional space. The new method also represent the reduced order model by polynomials defined on a Smolyak sparse grid whose coefficients are now obtained by data fitting. Compared to other intrusive ROMs and a previous interpolation-based non-intrusive ROM, this method only requires a number of solution snapshots from the full model and does not need to modify nor further run the original model in the process of ROM generation. In other words, the detailed physics and numerics of the original model are completely transparent to this least squares NIROM (LSROM). Another necessary module of the method involves the stabilisation of the reduced order iteration in the place of the governing differential equations. A numerical damping parameter is chosen to avoid the divergence of the POD coefficients.

The least square NIROM was tested on two CFD test cases against a high-performance fluid solver (Fluidity). In the simulations of a shallow water gyre and of a Kármán vortex street downstream of a cylinder at $Re = 400$ the non-intrusive model resolve the flow structures with great precision, measured by root mean squared errors and correlation coefficents, with only 3 or 6 POD basis functions. For the flow past a cylinder at $Re = 3600$, the accuracy of LSROM with 6 POD bases is decreased but it still captures the main features of the flow field. Preliminary results show that using more POD bases could increase the accuracy for a limited time beyond which the instability emerges. Future work will investigate the error and stability dependence on the number of POD bases and on the number of solution snapshots adopted into the fitting. Overall, the least squares fitting model with improved non-intrusiveness is a good approximation to the full model with enormous gain in computation and modelling economy.

The future work will address the stability issues of the ROM with greater details, such as optimising the data set for the least squares fitting by windowing or weighting, and error analysis to determine the relationship between the optimal number of POD bases used and the number of full model solution snapshots available. Morover, the LSROM will be applied to more complicated problems in sciences and engineering to fulfill its potential.

Acknowledgments

This work was carried out under partial funding from Imperial College-Zhejiang University Joint Applied Data Science Lab. Authors would like to acknowledge the support of the UK's Natural Environment Research Council projects(NER/A/S/2003/00595, NE/C52101X/1 and NE/C51829X/1), the Engineering and Physical Sciences Research Council (GR/R60898, EP/I00405X/1 and EP/J002011/1), and the Imperial College High Performance Computing Service. Prof. I.M. Navon acknowledges the support of NSF/CMG grant ATM-0931198. Xiao acknowledges the support of NSFC grant 11502241 and China postdoctoral science foundation grant (2014M562087). Lin acknowledges the support of NSFC grant 11201419. Prof. Pain and Fang are grateful for the support provided by BP Exploration. Prof. Pain is grateful for the support of the EPSRC MEMPHIS multi-phase flow programme grant.

References

[1] F. Fang, T. Zhang, D. Pavlidis, C. C. Pain, A. G. Buchan, and I. M. Navon. Reduced order modelling of an unstructured mesh air pollution model and application in 2D/3D urban street canyons. *Atmospheric Environment*, 96:96–106, 2014.

[2] R. Ştefănescu and I. M. Navon. POD/DEIM nonlinear model order reduction of an ADI implicit shallow water equations model. *Journal of Computational Physics*, 237:95–114, 2013.

[3] R. Ştefănescu, A. Sandu, and I. M. Navon. Comparison of POD reduced order strategies for the nonlinear 2D shallow water equations. *International Journal for Numerical Methods in Fluids*, 76(8):497–521, 2014.

[4] O. San and J. Borggaard. Principal interval decomposition framework for POD reduced-order modeling of convective Boussinesq flows. *International Journal for Numerical Methods in Fluids*, 78(1):37–62, 2015.

[5] D. A. Bistrian and I. M. Navon. An improved algorithm for the shallow water equations model reduction: Dynamic Mode Decomposition vs POD. *International Journal for Numerical Methods in Fluids*, 2015.

[6] F. Fang, C. C. Pain, I. M. Navon, A. H. Elsheikh, J. Du, and D. Xiao. Non-linear Petrov-Galerkin methods for reduced order hyperbolic equations and discontinuous finite element methods. *Journal of Computational Physics*, 234:540–559, 2013.

[7] D. Xiao, F. Fang, J. Du, C. C. Pain, I. M. Navon, A. G. Buchan, A. H. Elsheikh, and G. Hu. Nonlinear Petrov-Galerkin methods for reduced order modelling of the Navier-Stokes equations using a mixed finite element pair. *Computer Methods In Applied Mechanics and Engineering*, 255:147–157, 2013.

[8] D. Xiao, F. Fang, A. G. Buchan, C. C. Pain, I. M. Navon, J. Du, , and G. Hu. Nonlinear model reduction for the Navier-Stokes equations using Residual DEIM method. *Journal of Computational Physics*, 263:1–18, 2014.

[9] D. Xiao, F. Fang, C. C. Pain, and I. M. Navon. Non-intrusive reduced order 3D free surface modelling. *submitted to Ocean Modelling*, 2015.

[10] D. N. Daescu and I. M. Navon. A dual-weighted approach to order reduction in 4D-Var data assimilation. *Monthly Weather Review*, 136(3):1026–1041, 2008.

[11] Y. Cao, I. M. Navon, and Z. Luo. A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids*, 53:1571–1583, 2007.

[12] R. Ştefănescu, A. Sandu, and I. M. Navon. POD/DEIM reduced-order strategies for efficient four dimensional variational data assimilation. *Journal of Computational Physics*, 295:569–595, 2015.

[13] A. G. Buchan, A. A. Calloo, M. G. Goffin, S. Dargaville, F. Fang, C. C. Pain, and I. M. Navon. A POD reduced order model for resolving angular direction in neutron/photon transport problems. *Journal of Computational Physics*, 296:138–157, 2015.

[14] D. Xiao, P. Yang, F. Fang, J. Xiang, C. C. Pain, and I. M. Navon. Non-intrusive reduced order modeling of fluid-structure interactions. *Computer Methods in Applied Mechanics and Engineering*, 303:35–54, 2016.

[15] M. F. Barone, I. Kalashnikova, M. R. Brake, and D. J. Segalman. Reduced order modeling of fluid/structure interaction. *Sandia National Laboratories Report, SAND No*, 7189, 2009.

[16] M. Wang, D. Dutta, K. Kim, and J. C. Brigham. A computationally efficient approach for inverse material characterization combining gappy POD with direct inversion. *Computer Methods in Applied Mechanics and Engineering*, 286:373 – 393, 2015.

[17] K. C. Hoang, Y. Fu, and J. H. Song. An hp-proper orthogonal decomposition-moving least squares approach for molecular dynamics simulation. *Computer Methods in Applied Mechanics and Engineering*, 298:548 – 575, 2016.

[18] M. Schlegel and B. R. Noack. On long-term boundedness of Galerkin models. *Journal of Fluid Mechanics*, 765:325–352, 2 2015.

[19] J. Östh, B. R. Noack, S. Krajnović, D. Barros, and J. Bore. On the need for a nonlinear subscale turbulence term in POD models as exemplified for a high-Reynolds-number flow over an Ahmed body. *Journal of Fluid Mechanics*, 747:518–544, 5 2014.

[20] M. Barrault, Y. Maday, N.C. Nguyen, and A.T. Patera. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *C. R. Acad. Sci. Paris, Ser*, 339:667–672, 2004.

[21] S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal of Scientific Computation*, 32:2737–2764, 2010.

[22] M. J. Rewienski. *A trajectory piecewise-linear approach to model order reduction of nonlinear dynamical systems*. PhD thesis, Citeseer, 2003.

[23] C. Han. Blackbox stencil interpolation method for model reduction. Master's thesis, Massachusetts Institute of Technology, 2012.

[24] S. Walton, O. Hassan, and K. Morgan. Reduced order modelling for unsteady fluid flow using proper orthogonal decomposition and radial basis functions. *Applied Mathematical Modelling*, 37(20):8930–8945, 2013.

[25] D. Xiao, F. Fang, C. Pain, and G. Hu. Non-intrusive reduced order modelling of the Navier-Stokes equations based on RBF interpolation. *International Journal for Numerical Methods in Fluids*, 79(11):580–595, 2015.

[26] B. R. Noack, M. Morzynski, and G. Tadmor. *Reduced-order modelling for flow control*, volume 528. Springer, 2011.

[27] R. Noori, A. R. Karbassi, K. Ashrafi, M. Ardestani, and N. Mehrdadi. Development and application of reduced-order neural network model based on proper orthogonal decomposition for BOD5 monitoring: Active and online prediction. *Environmental Progress and Sustainable Energy*, 32(1):120–127, 2013.

[28] D. Xiao, F. Fang, A. G. Buchan, C. C. Pain, I. M. Navon, and A. Muggeridge. Non-intrusive reduced order modelling of the Navier–Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 293:522–541, 2015.

[29] C. C. Pain, M. D. Piggott, A. J. H. Goddard, F. Fang, G. J. Gorman, D. P. Marshall, M. D. Eaton, P. W. Power, and C. R. E. De Oliveira. Three-dimensional unstructured mesh ocean modelling. *Ocean Modelling*, 10(1):5–33, 2005.