Reduced Order Modeling of 2-D Reaction-Diffusion System Based on POD-DEIM and k-means Clustering

Enrico Alberto Cutillo[1], Gianmarco Petito[1], Katarzyna Bizon[2], Gaetano Continillo[1]
[1]Dipartimento di Ingegneria, Università del Sannio Benevento, Italy
[2]Department of Engineering, Politechnika Krakowska, Kraków, Poland

1 Introduction

Advances in computing power now allow detailed numerical simulation of a wide variety of chemically reactive flows in complex pathways, yet there is a strong need of development of model order reduction techniques. In fact, fast and accurate models are required in real-time control applications or in the detailed study of systems characterized by a large number of state variables and/or operating parameters. A widespread and powerful technique, especially when it comes to the reduction of nonlinear infinite dimensional systems described by evolutionary partial differential equations (PDE), is Proper Orthogonal Decomposition (POD) [1]. Either from experimental or numerically derived observations, POD delivers a set of empirical basis functions which can be introduced into classical subspace projection approaches for model order reduction such as the Galërkin method. POD-Galërkin has been widely used for model reduction in many fields, however there are still issues that need further investigation. Among the others, the policy of the collection of a representative set of experimental or simulation data is crucial for generating a global basis suitable for the determination of accurate Reduced Order Models (ROM). While the exploration of the parameter space is a well-established policy [2], there is no clear indication in the literature about the influence of the total number of snapshots and their distribution on the ROM performance. Recently, a methodology based on the mutual information for the selection of the most uncorrelated snapshots was proposed [3]. An alternative approach, based on a similar principle, was proposed later [4]; it comes down to selecting the most informative snapshots based on clustering methodology. Moreover, it is well known that POD reduced models do not generate efficient numerical codes when dealing with nonlinear equations in the primitive variables, in that it is still necessary to reevaluate matrix coefficients in the new conditions at each simulation time step, and this can make the whole approach highly inefficient. To overcome this limitation, Discrete Empirical Interpolation Methods (DEIM) have been proposed and employed, requiring reevaluation of the nonlinear terms in correspondence of a given set of optimally located state space points [5].

In this work, the method of POD is applied to a problem of self-ignition of a coal stockpile, described by two-dimensional reaction-diffusion equations [6]. Aiming at the improvement of the ROM performance, different sampling strategies, clustering and DEIM nodes collocation are studied. The performances of various POD-DEIM ROMs determined from various time sampled sets of representative solutions are compared among each other, and conclusions are drawn.

Correspondence to: katarzyna.bizon@pk.edu.pl

2 Mathematical model

The analyzed geometry for a coal stockpile is the infinitely long slab with rectangular cross section shown in Fig. 1a. The lower side of the domain lays on the ground, the upper and the right sides are exposed to ambient conditions, whereas the left side corresponds to the longitudinal symmetry plane of the pile. The gaseous reactant diffuses through the porous medium and reacts according to a first-order one-step exothermic chemical reaction. The classic Arrhenius exponential expression is used to describe the temperature dependence of the reaction rate. Consumption of the solid reactant is neglected, thus, only one mass balance equation for the gas reactant is formulated. Gas and solid temperature are assumed to be equal, resulting in only one energy balance equation. Based on the above assumptions the model equations are:

$$\frac{\partial c}{\partial t} = \mathrm{Le}\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2}\right) - \phi^2 c \exp\left(-\frac{\gamma}{T}\right) \;\; ; \quad \frac{\partial T}{\partial t} = \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right) + \beta\phi^2 c \exp\left(-\frac{\gamma}{T}\right) \tag{1}$$

where c is the dimensionless concentration of the gas reactant, T is the dimensionless temperature, Le, the Lewis number, is the ratio between mass and heat diffusivities, β is a dimensionless heat of reaction, ϕ is the thermal Thiele modulus, and γ is a dimensionless activation energy. Boundary conditions are:

$$\frac{\partial c}{\partial x} = \frac{\partial T}{\partial x} = 0 \quad \text{at} \quad x = 0\,, \qquad c\left(x_w, y, 0\right) = 1,\; T\left(x_w, y, 0\right) = 1 \;\; \forall\, y \in [0,1] \tag{2}$$

$$\frac{\partial c}{\partial y} = 0 \;\text{ and }\; T = 1 \quad \text{at} \quad y = 0\,, \qquad c\left(x, y_w, 0\right) = 1,\; T\left(x, y_w, 0\right) = 1 \;\; \forall\, x \in [0,1]\,. \tag{3}$$

The problem is supplemented with the following initial conditions: $c\left(x, y, 0\right) = 1,\; T\left(x, y, 0\right) = 1$.

3 Proper Orthogonal Decomposition, Discrete Empirical Interpolation and *k*-Means Clustering

For a given set of snapshots $\mathbf{y}\left(\mathbf{x}, t_i\right),\; \mathbf{x} = [x_1, x_2, ...x_N],\; i = 1, ..., M$, collected from Full Order Model (FOM) simulations, the POD basis $\mathbf{\Phi} = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ...\boldsymbol{\phi}_N\}$ can be obtained by solving the eigenvalue problem given by $\mathbf{C\Phi = \Lambda C}$, where $\mathbf{C}$ is the autocorrelation matrix defined as $\mathbf{C} = 1/M \cdot \mathbf{Y}\mathbf{Y}^T$ [3,4]. In the latter formula $\mathbf{Y} \in R^{N \times M}$ denotes the snapshots represented in matrix form. Employing the determined POD basis, the state variable can be expressed in a truncated form as $\mathbf{y} \approx \mathbf{\Phi}_K \mathbf{c}_K$, where $K \leq N$ is an approximation order, $\mathbf{\Phi}_K \in R^{N \times K}$ is a matrix composed of the first K columns of the matrix $\mathbf{\Phi}$, whereas $\mathbf{c}_K$ are time-dependent coefficients of the POD modes.

Since the FOM is constructed by applying the method of lines for the discretization of the original PDE problem, it can be written in the following matrix form of N ordinary differential equations (ODEs):

$$\frac{d}{dt}\mathbf{y} = \mathbf{A}\mathbf{y} + \mathbf{F}(\mathbf{y})\,, \quad \mathbf{y}(0) = \mathbf{y}_{init} \tag{4}$$

Introduction of the truncated state variable in Eq. (4) followed by the projection of the resulting system onto the POD basis yields the following ROM:

$$\frac{d}{dt}\mathbf{c}_K = \mathbf{\Phi}_K^{\mathrm{T}} \mathbf{A} \mathbf{\Phi}_K \, \mathbf{c}_K + \mathbf{\Phi}_K^{\mathrm{T}} \mathbf{F}\left(\mathbf{\Phi}_K \mathbf{c}_K\right), \quad \mathbf{c}_K(0) = \mathbf{\Phi}_K^{\mathrm{T}} \mathbf{y}_{init} \tag{5}$$

The computational burden resulting from the need of evaluating of the non-linear term $\mathbf{F}\left(\mathbf{\Phi}_K \mathbf{c}_K\right)$ in the original *N*-dimensional space can be further reduced by applying DEIM [5]. This method involves the determination of an additional POD basis from the snapshots of a nonlinear expression, denoted here by $\mathbf{\Psi} = \{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, ...\boldsymbol{\psi}_N\}$, with evaluation of that expression only at optimally selected grid points, the so-called interpolation points, collected within matrix **P**. Application of the DEIM technique described in detail in [5] yields the following final set of ODEs:

$$\frac{d}{dt}\mathbf{c}_K = \mathbf{\Phi}_K^{\mathrm{T}} \mathbf{A} \mathbf{\Phi}_K \mathbf{c}_K + \mathbf{\Phi}_K^{\mathrm{T}} \mathbf{\Psi}_J \left(\mathbf{P}^{\mathrm{T}} \mathbf{\Psi}_J\right)^{-1} \mathbf{F}\left(\mathbf{P}^{\mathrm{T}} \mathbf{\Phi}_K \mathbf{c}_K\right), \quad \mathbf{c}_K(0) = \mathbf{\Phi}_K^{\mathrm{T}} \mathbf{y}_{init} \tag{6}$$

where $\mathbf{\Psi}_J \in R^{N \times J}$ and $\mathbf{P} \in R^{N \times J}$, $J \leq N$

Usually, the snapshots employed for the determination of the POD basis, both from the snapshots of the state variable and of the nonlinear term, are collected at uniform time-sampling rate. Here, in order to increase the amount of information contained in a given number of POD modes, the procedure based on the so-called *k*-means clustering is employed [4]. Namely, the solution profiles are first grouped into clusters and then the POD modes are determined from the centroids of the clusters. More precisely, the *k*-means algorithm groups together subsets of 'nearby' solutions, based on a mutual distance, in this case defined as a squared Euclidean distance.

4 Results and discussion

The parameter set chosen for the simulation is the following: Le = 0.233, ϕ^2 = 70000, β = 4.287 and $\gamma \in [11.5, 14]$. All of the time series and solution diagrams reported within this study refer to a point of the spatial domain marked with the red dot in Fig. 1a. Prior to the application of POD-Galërkin, the problem is resolved using the FOM, derived here by finite difference approximation of the spatial derivatives at coordinate nodes uniformly distributed along *x* and *y* with $\Delta x = \Delta y = 0.01$, yielding a mesh of 10^4 nodes. This results in the transformation of the infinite dimensional model consisting of 2 PDEs into a large system of $2 \cdot 10^4$ ODEs. The system of ODEs was then integrated using the variable time step solver *ode15s* of MATLAB in the time interval $t \in [0, 5]$, and the solution was sampled with a uniform timestep equal to $\Delta t = 2 \cdot 10^{-4}$, resulting in 25001 snapshots. Figure 1b shows the solution diagram for dimensionless temperature and concentration, obtained using the FOM upon varying the parameter γ.

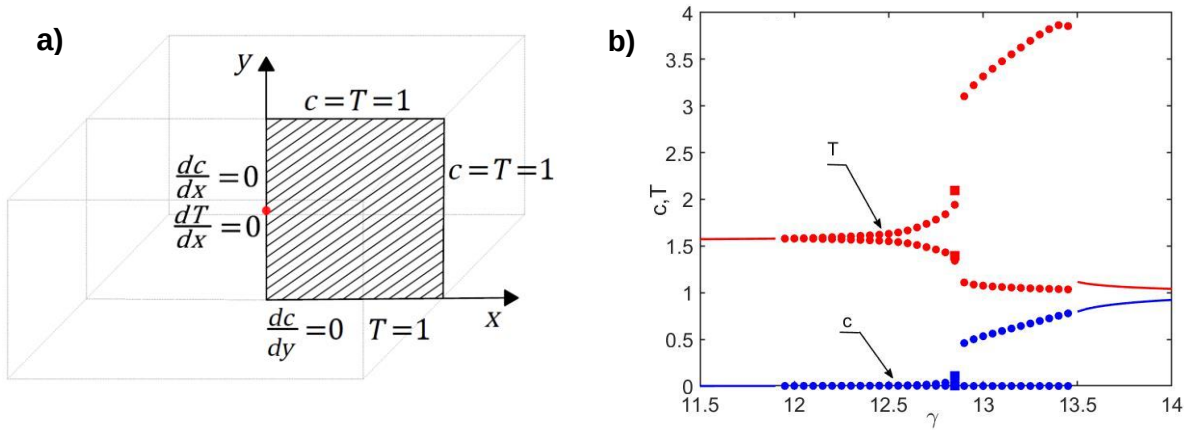

Figure 1. Geometric representation of the problem with the relevant boundary conditions (a); and solution diagram obtained with the FOM by varying the parameter γ (b). In (b) solid lines, dots and squares denote, respectively, steady-state solution, maxima and minima of period-1, and maxima and minima of period-2 solutions.

Table 1: Computational times and errors in the approximation of the dimensionless temperature for various values of γ. POD and POD-DEIM constructed using the *k*-means clustering algorithm on two FOM sets of snapshots collected at $\gamma = 12.6$ and $\gamma = 13$.

	$\gamma = 12.6$		$\gamma = 12.85$		$\gamma = 13$	
	CPU time	error	CPU time	error	CPU time	error
FOM	$1.10 \cdot 10^4$	-	$1.04 \cdot 10^4$	-	$6.97 \cdot 10^3$	-
POD	559.31	$1.98 \cdot 10^{-4}$	503.84	$5.88 \cdot 10^{-4}$	479.99	$1.64 \cdot 10^{-4}$
POD-DEIM	10.83	$2.03 \cdot 10^{-4}$	9.53	$1.49 \cdot 10^{-4}$	10.59	$1.45 \cdot 10^{-4}$

The solutions can be classified in steady-state and oscillatory solutions. Three bifurcation points are shown, one at $\gamma = 11.95$ where the first oscillatory solution is found indicating a Hopf bifurcation point. Then for $\gamma = 12.85$ a period-2 oscillation is detected, and finally at $\gamma = 13.5$ the system returns back to steady-state. Chaos is hypothesized to occur in a narrow window around $\gamma = 12.85$ but has not been detected.

Two objectives were addressed in this study. The first one was to achieve a substantial reduction in the computational time while maintaining a high level of accuracy of the ROM, i.e. the aim was to build a ROM capable of accurately predicting the behavior of the system. For this scope, the classical POD-Galërkin technique has been first applied using $K = 200$ modes for the approximation of the state variables. The computational cost of model simulation has been reduced by two orders of magnitude as it can be seen in Table 1; further speed up has been achieved using, in addition, the DEIM algorithm, to address the bottleneck operation of evaluating the nonlinear term in the 10000 points of the FOM domain.

The POD-DEIM constructed using again $K = 200$ POD modes, and $J = 500$ interpolation nodes within the DEIM procedure, resulted to be up to 1020 times faster than the FOM, and up to 55 times faster than the reduced model built with the classical POD approach.

Attention has also been given to snapshot selection strategies by performing tests with and without the *k*-means clustering algorithm. To find the most suitable dataset for the construction of reduced models, two sampling strategies were compared. In the first strategy, all snapshots obtained from the FOM solution with $\gamma = 13$ were utilized; in the second one, a set of 2000 centroids of clusters determined using the *k*-means clustering algorithm applied to a mix of FOM solutions obtained for $\gamma = 13$ and $\gamma = 12.6$ were used to build POD basis. These two values of γ correspond to the neighborhood of the onset of the period doubling bifurcation cascade ($\gamma = 12.85$) where chaotic behavior of the system is hypothesized.

Two things can be noted in Table 1, where the error of the approximation of the ROM determined using the second strategy is reported. Firstly, the error for POD-DEIM is roughly the same as the error of the standard POD. Moreover, for various values of γ, the error remains of the same order of magnitude, confirming the global character of both POD and POD-DEIM derived bases.

Figure 2 shows the set of DEIM interpolation points obtained using two snapshot selection strategies, plotted on a contour diagram showing the value of the nonlinear term in the energy balance at $t = 0.34$. This instant has been chosen to underline the differences between the two results. It is quite evident that, when the interpolation nodes are determined based on the snapshots coming only from a single FOM solution (Fig. 2a), the interpolation nodes are concentrated in specific regions of the domain. On the other hand, when the DEIM points are determined based on clustered snapshots coming from the combination of two different FOM solutions, they spread more uniformly over the domain (Fig. 2b).

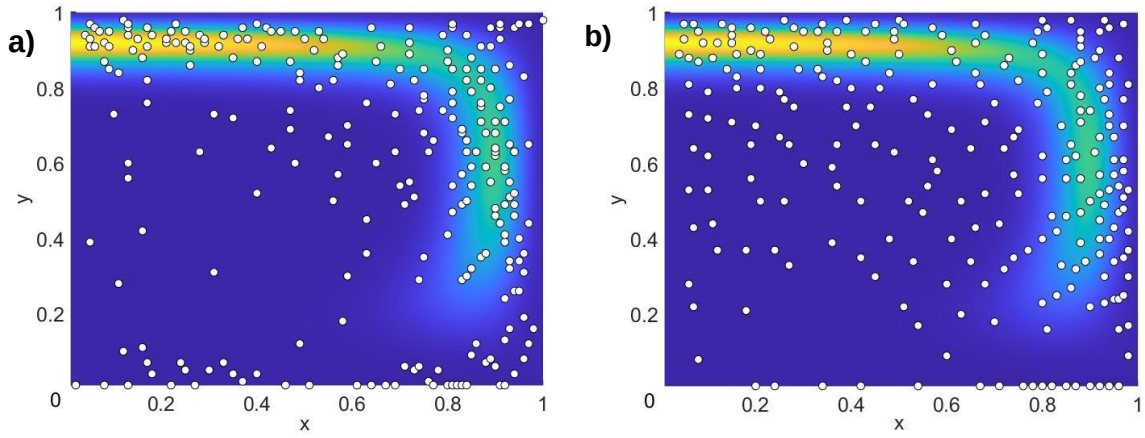

Figure 2. DEIM interpolation nodes determined from all nonlinear term snapshots obtained for $\gamma = 13$ (a), and from a set of 2000 snapshots extracted by *k*-means clustering algorithm run on FOM solutions obtained for $\gamma = 13$ and $\gamma = 12.6$ (b). The frames in figure are taken at $t = 0.34$ AUT (arbitrary units of time).

Figure 3a shows the time series of temperature and concentration calculated using FOM and POD-DEIM ($K = 200$, $J = 500$) determined with the use of *k*-means clustering with 2000 centroids. The solutions obtained using both methods practically overlap, confirming the very high accuracy of the ROM. Figure 3b reports the solution diagram calculated using POD-DEIM. It appears that the reduced model predicts the bifurcations points and maxima and minima of oscillatory solutions, just as the FOM (Fig. 1b) except for the bifurcation point at $\gamma = 11.95$, as the first oscillatory solution predicted by the POD-DEIM model is observed at $\gamma = 12.15$.

Finally, Fig. 4 reports the temperature solutions obtained from FOM and from POD-DEIM, in bi-dimensional representation and with reference to three different values of time.

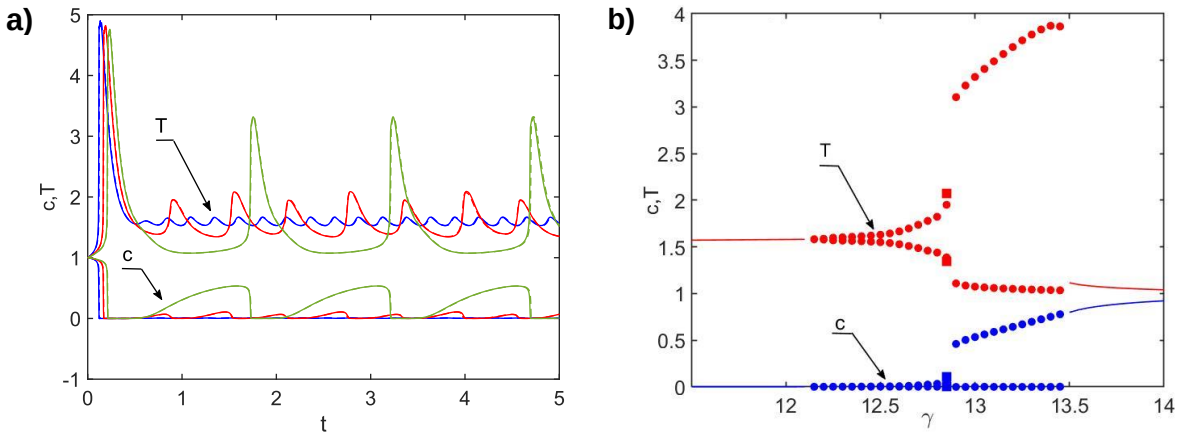

Figure 3: Comparison of temperature and concentration obtained using FOM (solid line) and POD-DEIM (dashed lines) for different values of γ (a), and solution diagram obtained using POD-DEIM upon varying γ (b). In (b) solid lines, dots and squares denote, respectively, steady-state solution, maxima and minima of period-1, and maxima and minima of period-2 solutions.

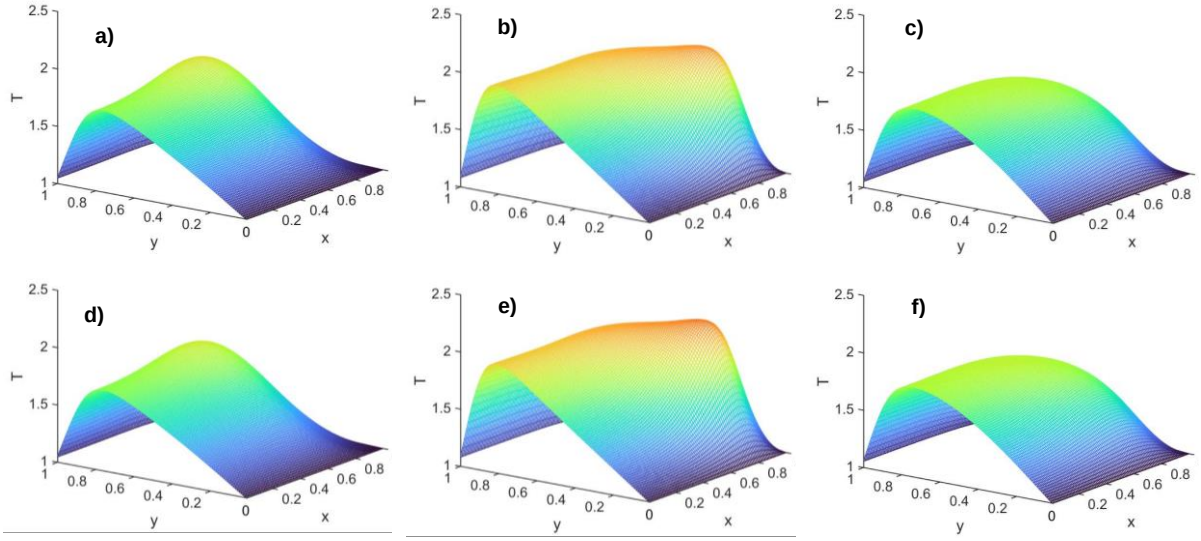

Figure 4: Comparison of temperature solutions for γ =12.6 obtained from FOM (first row) and POD-DEIM based ROM (second row) at three different time instants: t = 1.26 (a) and (d), t = 1.3198 (b) and (e), t = 1.36 (c) and (f).

5 Conclusions

In this study a bi-dimensional model of a reaction-diffusion system has been solved. The solution diagram of the FOM has been reported. A POD-DEIM model has been constructed to reproduce the FOM solutions with high accuracy and much lower computational cost. The influence of two different sampling strategies have been investigated which resulted in a different distribution of DEIM interpolation nodes inside the analyzed domain. The innovative approach based on *k*-means clustering has allowed the construction of a global basis for the reproduction of the FOM solution diagram. The further adoption of DEIM algorithm in combination with *k*-means clustering not only reduced up to 1020 times the computational time for the calculation of one solution, but also reproduced all of the bifurcation points found with the FOM, thus demonstrating quantitative and qualitative agreement.

References

[1] Holmes P, Lumley JL, Berkooz G. (1996). Turbulence, coherent structures, dynamical systems and symmetry, Cambridge University Press.

[2] Bizon K, Continillo G, Russo L, Smuła J. (2008). On POD reduced models of tubular reactor with periodic regimes. Comput. Chem. Eng. 32: 105.

[3] Kostka S, Lynch AC, Huelskamp BC, Kiel BV, Gord JR, Roy S. (2012). Characterization on flame-sheeding behavior behind a bluff-body using proper orthogonal decomposition. Combust. Flame 159: 2872.

[4] Bizon K. Continillo G. (2021). Efficient optimization of a multifunctional catalytic fixed-bed reactor via reduced-order modeling approach. Chem. Eng. Res. Des. 165:214–229.

[5] Bizon K. (2017). Assessment of a POD method for the dynamical analysis of a catalyst pellet with simultaneous chemical reaction, adsorption and diffusion: Uniform temperature case. Comput. Chem. Eng. 97: 259-270.

[6] Continillo G, Faraoni V, Maffettone PL, Crescitelli S (2000). Non-linear dynamics of a self-ignition reaction system. Chem. Eng. Sci. 55:303-309.