Using recurrent neural networks for nonlinear component computation in advection-dominated reduced-order models

Romit Maulik
Argonne Leadership Computing Facility
Argonne National Laboratory
Lemont, IL 60439
rmaulik@anl.gov

Vishwas Rao
Mathematics and Computer Science Division
Argonne National Laboratory
Lemont, IL 60439
vhebbur@anl.gov

Sandeep Madireddy
Mathematics and Computer Science Division
Argonne National Laboratory
Lemont, IL-60439
smadireddy@anl.gov

Bethany Lusch
Argonne Leadership Computing Facility
Argonne National Laboratory
Lemont, IL 60439
blusch@anl.gov

Prasanna Balapraakash
Mathematics and Computer Science Division &
Leadership Computing Facility
Argonne National Laboratory
Lemont, IL 60439
pbalapra@anl.gov

Abstract

Rapid simulations of advection-dominated problems are vital for multiple engineering and geophysical applications. In this paper, we present a long short-term memory neural network to approximate the nonlinear component of the reduced-order model (ROM) of an advection-dominated partial differential equation. This is motivated by the fact that the nonlinear term is the most expensive component of a successful ROM. For our approach, we utilize a Galerkin projection to isolate the linear and the transient components of the dynamical system and then use discrete empirical interpolation to generate training data for supervised learning. We note that the numerical time-advancement and linear-term computation of the system ensure a greater preservation of physics than does a process that is fully modeled. Our results show that the proposed framework recovers transient dynamics accurately without nonlinear term computations in full-order space and represents a cost-effective alternative to solely equation-based ROMs.

1 Introduction

High-fidelity simulations of systems characterized by nonlinear partial differential equations (PDEs) are computationally prohibitive for decision-making tasks in multiple engineering and geophysical applications. To address this issue, researchers have devoted significant effort to developing reduced-order models (ROMs) of such systems with the aim of reducing the degrees of freedom of the forward problem to manageable magnitudes. Therefore, ROMs find extensive application in uncertainty quantification, control, and multifidelity optimization. The interested reader is directed to [1] and
We denote this approach as POD-ML. The time advancement of this hybrid system is then performed which we denote as the POD-Galerkin projection method (POD-GP). We remark that the calculation where \( \nu = 1/Re \) is the viscosity and \( \mathcal{N} \) and \( \mathcal{L} \) correspond respectively to the quadratic nonlinearity and the linear operator in the viscous Burgers equation. In this study, we fix \( \nu = 1e3 \), \( L = 1 \), and the final time as \( t_f = 2.0 \) and seek to accelerate ROMs by performing a nonintrusive calculation of the nonlinear term \( \mathcal{N} \), which leads to lower memory and compute cost requirements compared with traditional numerical techniques.

2 POD-Galerkin projection and discrete empirical interpolation method

The orthogonal nature of the POD bases may be leveraged for a Galerkin projection onto a linear subspace that hierarchically embeds information. We start by revisiting Equation (1) written in the form of a full-order evolution equation for fluctuation components using \( \mathcal{N}_f \) degrees of freedom:

\[
\dot{u}(x,t) + \mathcal{N}[u(x,t);\nu] + \mathcal{L}[u(x,t);\nu] = 0, \quad (x,t,\nu) \in \Omega \times \mathcal{T} \times \mathcal{P},
\]

\[
u \frac{\partial u(x,t)}{\partial t} = \frac{x}{t+1} \sqrt{\frac{1+\nu}{10}} \exp\left(Re \frac{\nu^2}{4t+4}\right),
\]

\[
\frac{u(x,0)}{1+\sqrt{\frac{1+\nu}{10}} \exp\left(Re \frac{\nu^2}{4}\right)}, \quad u(0,t) = 0, \quad u(L,t) = 0,
\]

where \( \nu = 1/Re \) is the viscosity and \( \mathcal{N} \) and \( \mathcal{L} \) correspond respectively to the quadratic nonlinearity and the linear operator in the viscous Burgers equation. In this study, we fix \( \nu = 1e3 \), \( L = 1 \), and the final time as \( t_f = 2.0 \) and seek to accelerate ROMs by performing a nonintrusive calculation of the nonlinear term \( \mathcal{N} \), which leads to lower memory and compute cost requirements compared with traditional numerical techniques.

2 POD-Galerkin projection and discrete empirical interpolation method

The orthogonal nature of the POD bases may be leveraged for a Galerkin projection onto a linear subspace that hierarchically embeds information. We start by revisiting Equation (1) written in the form of a full-order evolution equation for fluctuation components using \( \mathcal{N}_f \) degrees of freedom:

\[
\dot{u}_f(x,t) + \mathcal{N}_f[u_f(x,t);\nu] + \mathcal{L}_f[u_f(x,t);\nu] = 0.
\]

It can be expressed in the reduced basis as

\[
\psi a_r(t,\nu) + \mathcal{N}_f[\psi a_r(t,\nu)] + \mathcal{L}_f[\psi a_r(t,\nu);\nu] = 0,
\]

where \( a_r : \mathcal{T} \times \mathcal{P} \rightarrow \mathbb{R}^{N_f} \), \( a_r \in \alpha_r \), corresponds to the temporal coefficients at one time instant of the system evolution and \( \psi \in \mathbb{R}^{N_f \times N_f} \) represents the truncated bases of the POD modes obtained from \( N_s \) snapshots of the full-order solution having \( \mathbb{R}^{N_f} \) degrees of freedom. The orthogonal nature of the reduced basis and the commutative property of the linear term can be leveraged to obtain

\[
\dot{a}_r(t,\nu) + \psi^T \mathcal{N}_f[\psi a_r(t,\nu)] + \mathcal{L}_r[a_r(t,\nu);\nu] = 0,
\]

which we denote as the POD-Galerkin projection method (POD-GP). We remark that the calculation of \( \mathcal{N}_f \) necessitates a re-projection of the reduced-order solution to full-order space and subsequent nonlinear term calculation throughout the domain. This process can be extremely expensive for advection-dominated problems such as ours. For comparison and for the generation of training data, we use the discrete empirical interpolation method (DEIM) [4], which reduces the number of nonlinear term calculations significantly. Our goal is to completely preclude nonlinear term computation in full-order space and to utilize DEIM as a bridge between POD-GP and our formalism. We denote this approach as POD-ML. The time advancement of this hybrid system is then performed by using a simple first-order Euler integrator (although this may easily be extended to higher-order methods such as Runge-Kutta or Adams-Bashforth techniques).

The DEIM procedure is outlined in the following. Let \( \varphi \in \mathbb{R}^{N_f \times N_m} \) be the truncated POD basis matrix for snapshots of the nonlinear term \( \mathcal{N}_f \), where \( N_m \) is the number of vectors in the truncated basis. DEIM calculates a unit-vector matrix \( \mathcal{P} \in \mathbb{R}^{N_f \times N_m} \), specifying \( N_m \) locations (out of the \( N_f \) locations in the full-order space), where \( \mathcal{N} \) may be computed to construct an approximation for \( \mathcal{N}_f \). Then we define a matrix \( \mathcal{P} \in \mathbb{R}^{N_f \times N_f} \) as

\[
\mathcal{P} = \varphi (\varphi^T \varphi)^{-1} \varphi^T
\]
and approximate Equation (4) as
\[ \dot{a}_r(t, \nu) + L_r[a_r(t, \nu); \nu] + \psi^T \mathcal{P}N_f[\psi a_r(t, \nu)] = 0. \] (6)

Note that the linear operator and \( \phi(\mathcal{P}^T \phi)^{-1} \) are precomputed, leading to a reduction in the cost of calculating the nonlinear term \( \mathcal{P}^T \mathcal{N}_f \). Depending on \( N_m \), this approach costs significantly less than POD-GP. The exact algorithm utilized to obtain \( \mathcal{P} \) is a variant of least squares and is given in [3]. Figure 1 shows a validation of our numerical method where both POD-GP and POD-DEIM are approximately equal in accuracy for our problem.

### 3 Nonlinear surrogate using long short-term memory

We utilized a long short-term memory (LSTM) neural network to further reduce the computational complexity of DEIM by calculating the nonlinear term, namely, \( \psi^T \mathcal{P}N_f(t, \psi a_r) \) in Equation (6) nonintrusively. From this point, we denoted the DEIM approximation for the nonlinear term as \( N_r \). Our training data was generated by calculating DEIM coefficients from the analytical solution (keeping \( N_r = 12 \) and \( N_m = 24 \)) for 300 snapshots of the nonlinear term in time.

We utilized a standard LSTM network [5] with one cell, \( N_h = 30 \) hidden neurons, and a fully connected output layer to do many-to-one predictions for the nonlinear term \( N_r \) in a recursive fashion. In other words, the outputs of the network are the DEIM nonlinear terms, which are reutilized as inputs for the next prediction. We used an input window of 10 to give an output of the nonlinear term at the next timestep, a batch size of 32, a learning rate of 0.001, and the ADAM optimizer. We observed that 60 epochs were sufficient for training the network under this setting. The training utilized a random 20% validation split and saved the best model according to this data. Testing was performed through deployment within the hybrid numerical and ML framework as explained below. Training and assessment were performed by using an Intel core-i7 CPU with a basic build of TensorFlow 1.13.
Figure 2a outlines the collected DEIM coefficients \( (N_r) \) from the analytical solution showing oscillations. To predict these coefficients appropriately, we used a Savitsky-Golay low-pass spatial filter to mimic the effect of numerical smoothing in spatial and temporal discretizations. Figure 2b shows the learning of the evolution of these coefficients (as an \textit{a priori} assessment). The LSTM predictions then were added to the linear component of the governing equation to obtain an accurate state advancement, as shown in Figure 2c. Preprocessing of the oscillatory DEIM coefficients was crucial to the stability of the hybrid machine learning and numerical method. \( L_2 \)-errors in modal evolution for each ROM technique (when compared with the truth) were 0.064 for POD-GP, 0.059 for POD-DEIM, and 0.044 for POD-ML, indicating similar fidelity of the final solution. Assessments of computational cost are precluded here because for simple systems such as the Burgers equation, the cost of LSTM inference dominates the nonlinear computation. Benefits of the proposed method may be observed for larger systems with much greater degrees of freedom. For our problem, without any reduction, the per step complexity is \( O(N_f^2) \) FLOPS + \( O(N_f) \) nonlinear evaluations. POD-GP needs \( O(N_r^2) + O(N_f N_r) \) FLOPS + \( O(N_f) \) evaluations, where \( N_r \ll N_f \). POD-DEIM requires \( O(N_r^2) + O(N_f N_r + N_f N_m) \) FLOPS + \( O(N_m) \) nonlinear evaluations, where \( N_m \ll N_f \). In contrast, our proposed method utilizes \( O(N_h N_r + N_h N_m + N_r^2 + N_f N_r) \) FLOPS and no online nonlinear evaluations. For larger problems where typically \( N_h, N_m, \) and \( N_r \) are of the same order of magnitude, our method will yield exceptional gains.

4 Conclusions

We have developed a method that precludes the reprojection of the ROM into full-order space for nonlinear term computation and, instead, utilizes prior knowledge of its evolution through the learning of precalculated DEIM coefficients. However, smoothing is required for the training data to mimic the stability-preserving nature of numerical methods. This requirement may also be true when the data comes from noisy physical or numerical experiments. The results here suggest that ROMs can be improved significantly for quick evaluation of extremely expensive physical systems with excellent retention of accuracy through the adoption of sequence learning techniques.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. This research was funded in part and used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. This paper describes objective technical results and analysis.

References

Figure 2: Preprocessing of DEIM coefficients (top), \textit{a priori} LSTM predictions for DEIM coefficients (middle) and relative errors between POD-ML and POD-DEIM (bottom) showing good agreement between the surrogate and equation based models.
The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan.

http://energy.gov/downloads/doe-public-access-plan