

---

# Nonlinear-manifold reduced order models with domain decomposition

---

**Alejandro N. Diaz**  
Rice University  
Houston, TX 77005  
and5@rice.edu

**Youngsoo Choi**  
Lawrence Livermore National Laboratory  
Livermore, CA 94550  
choi15@llnl.gov

**Matthias Heinkenschloss**  
Rice University  
Houston, TX 77005  
heinken@rice.edu

## Abstract

A nonlinear-manifold reduced order model (NM-ROM) is a great way of incorporating underlying physics principles into a neural network-based data-driven approach. We combine NM-ROMs with domain decomposition (DD) for efficient computation. NM-ROMs offer benefits over linear-subspace ROMs (LS-ROMs) but can be costly to train due to parameter scaling with the full-order model (FOM) size. To address this, we employ DD on the FOM, compute subdomain NM-ROMs, and then merge them into a global NM-ROM. This approach has multiple advantages: parallel training of subdomain NM-ROMs, fewer parameters than global NM-ROMs, and adaptability to subdomain-specific FOM features. Each subdomain NM-ROM uses a shallow, sparse autoencoder, enabling hyper-reduction (HR) for improved computational speed. In this paper, we detail an algebraic DD formulation for the FOM, train HR-equipped NM-ROMs for subdomains, and numerically compare them to DD LS-ROMs with HR. Results show a significant accuracy boost, on the order of magnitude, for the proposed DD NM-ROMs over DD LS-ROMs in solving the 2D steady-state Burgers' equation.

## 1 Introduction

In science and engineering, complex tasks often involve repeatedly simulating a large-scale, parameterized, nonlinear system referred to as the full-order model (FOM). Ensuring high fidelity requires a high-dimensional model, leading to significant computational costs and lengthy simulations. As a result, tasks like design optimization become impractical for large-scale problems. Model reduction offers a solution by replacing the FOM with a computationally efficient, low-dimensional model called a reduced-order model (ROM). This ROM approximates the FOM's behavior with adjustable accuracy, making it suitable for many-query applications. However, construction of accurate and computationally efficient ROMs poses challenges. To address them, we integrate the nonlinear-manifold ROM (NM-ROM) approach with an algebraic domain-decomposition (DD) framework.

Various model reduction methods have been integrated with DD, like reduced basis elements (RBE) [1, 17, 18, 26, 27, 27, 37, 38], and the alternating Schwarz method [5, 9, 28, 43]. However, they are often specialized to specific problems, dealing with the physical domain at the PDE level. In contrast, the authors in [25] take an algebraic approach by decomposing the FOM at the discrete level and computing linear-subspace ROMs (LS-ROMs) for each subdomain. While LS-ROMs work well in many cases [2–4, 7, 8, 10, 12–15, 19–22, 24, 31, 39, 42], it is well known that advection-dominated problems and problems with sharp gradients cannot be well-approximated using low-dimensional linear subspaces. These problems are said to have slowly decaying Kolmogorov  $n$ -width [40]. Recent approaches, such as nonlinear-manifold ROMs (NM-ROMs), address these problems by nonlinearly approximating the FOM in a low-dimensional nonlinear manifold. This is typically achieved through training an autoencoder on FOM snapshot data (e.g., [23, 29, 30, 32, 33]). However, training of

NM-ROMs is expensive. Indeed, in the monolithic single-domain case, the high-dimensionality of the FOM training data results in a large number of neural network (NN) parameters requiring training. In [6] this cost issue was mitigated by first computing a low dimensional proper orthogonal decomposition (POD) model, and then using a NN to train the coefficients in this POD. Instead, we integrate an autoencoder framework with DD. By coupling NM-ROM with DD, one can compute FOM training data on subdomains, thus reducing the dimensionality of subdomain NM-ROM training data, resulting in fewer parameters that need to be trained per subdomain NM-ROM.

We also note that couplings of NNs and DD for solutions of partial differential equations (PDEs) have been considered in previous work (e.g., [34–36, 44]). However, these approaches use deep learning to solve a PDE by representing its solution as a NN and minimizing a corresponding physics-informed loss function. In contrast, our work uses autoencoders to reduce the dimensionality of an existing numerical model. The autoencoders are pretrained in an *offline* stage to find low-dimensional representations of FOM snapshot data, and used in an *online* stage to significantly reduce the computational cost and runtime of numerical simulations. Our work is the first to couple autoencoders with DD in the reduced-order modeling context.

Here, we extend the work of [25] on DD LS-ROM and integrate NM-ROM with hyper-reduction (HR) using shallow, sparse autoencoders discussed in [32]. We incorporate the NM-ROM approach into this framework because of its success when applied to problems with slowly decaying Kolmogorov  $n$ -width. DD allows one to compute FOM training snapshots on subdomains, thus reducing the dimensionality of subdomain NM-ROM training data, resulting in fewer parameters that need to be trained per subdomain NM-ROM. We use wide, shallow, and sparse autoencoder architecture, which allows HR to be efficiently applied, thus reducing the complexity caused by nonlinearity and yielding computational speedup. Additionally, we modify the wide, shallow, and sparse architecture used in [32] to also include a sparsity mask for the encoder input layer as well as the decoder output layer. The proposed DD NM-ROM approach is compared with DD LS-ROM on the 2D Burgers' equation.

## 2 DD full order model

First consider the monolithic, single-domain FOM written as a residual equation

$$\mathbf{r}(\mathbf{x}; \boldsymbol{\mu}) = \mathbf{0}, \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^{N_x}$  is the state,  $\boldsymbol{\mu} \in \mathcal{D} \subset \mathbb{R}^{N_\mu}$  is a parameter, and  $\mathbf{r} : \mathbb{R}^{N_x} \times \mathbb{R}^{N_\mu} \rightarrow \mathbb{R}^{N_x}$  is the residual function. FOMs of the form (1) typically arise from discretizations of partial differential equations (PDEs). One can reformulate (1) into a DD formulation by partitioning the residual equation into  $n_\Omega$  systems of equations (so-called *algebraic* subdomains), coupling them via *compatibility constraints*, and converting the systems of equations into a least-squares problem, resulting in

$$\min_{(\mathbf{x}_i^\Omega, \mathbf{x}_i^\Gamma)_{i=1, \dots, n_\Omega}} \frac{1}{2} \sum_{i=1}^{n_\Omega} \|\mathbf{r}_i(\mathbf{x}_i^\Omega, \mathbf{x}_i^\Gamma; \boldsymbol{\mu})\|_2^2, \quad \text{s.t.} \quad \sum_{i=1}^{n_\Omega} \mathbf{A}_i \mathbf{x}_i^\Gamma = \mathbf{0}, \quad (2)$$

where  $\mathbf{x}_i^\Omega \in \mathbb{R}^{N_i^\Omega}$ ,  $\mathbf{x}_i^\Gamma \in \mathbb{R}^{N_i^\Gamma}$ ,  $\mathbf{r}_i : \mathbb{R}^{N_i^\Omega} \times \mathbb{R}^{N_i^\Gamma} \times \mathcal{D} \rightarrow \mathbb{R}^{N_i^r}$ , and  $\mathbf{A}_i \in \{-1, 0, 1\}^{N_a \times N_i^\Gamma}$  are the  $i$ -th subdomain interior-state, interface-state, residual function, and compatibility constraint matrix, respectively. The sparsity pattern of the monolithic residual function  $\mathbf{r}$  determines the structure of the subdomain residual functions  $\mathbf{r}_i$ , as well as the decomposition of the state  $\mathbf{x}$  into subdomain states  $(\mathbf{x}_i^\Omega, \mathbf{x}_i^\Gamma)$ . The interior-states  $\mathbf{x}_i^\Omega$  are those that are *only* used to compute the residual  $\mathbf{r}_i$  in the  $i$ -th subdomain, whereas the interface-states  $\mathbf{x}_i^\Gamma$  are also used in the residual computation of neighboring subdomains. The equality constraint determined by  $\mathbf{A}_i$  enforces equality on the overlapping interface states. For further details, see [16, Sec. 2] or [25, Sec. 2].

## 3 DD nonlinear-manifold reduced order model

For each subdomain  $i \in \{1, \dots, n_\Omega\}$ , let  $\mathbf{g}_i^\Omega : \mathbb{R}^{n_i^\Omega} \rightarrow \mathbb{R}^{N_i^\Omega}$ ,  $n_i^\Omega \ll N_i^\Omega$ , and  $\mathbf{g}_i^\Gamma : \mathbb{R}^{n_i^\Gamma} \rightarrow \mathbb{R}^{N_i^\Gamma}$ ,  $n_i^\Gamma \ll N_i^\Gamma$ , be decoders such that  $\mathbf{x}_i^\Omega \approx \mathbf{g}_i^\Omega(\hat{\mathbf{x}}_i^\Omega)$  and  $\mathbf{x}_i^\Gamma \approx \mathbf{g}_i^\Gamma(\hat{\mathbf{x}}_i^\Gamma)$ . Also let  $\mathbf{B}_i \in \{0, 1\}^{N_i^B \times N_i^r}$ ,  $N_i^B \leq N_i^r$ , denote a row-sampling matrix for collocation HR, and let  $\mathbf{C} \in \mathbb{R}^{n_C \times N_a}$ ,  $n_C \ll N_a$ , be a Gaussian test matrix. The DD NM-ROM is evaluated by solving

$$\min_{(\hat{\mathbf{x}}_i^\Omega, \hat{\mathbf{x}}_i^\Gamma)_{i=1, \dots, n_\Omega}} \frac{1}{2} \sum_{i=1}^{n_\Omega} \left\| \mathbf{B}_i \mathbf{r}_i \left( \mathbf{g}_i^\Omega(\hat{\mathbf{x}}_i^\Omega), \mathbf{g}_i^\Gamma(\hat{\mathbf{x}}_i^\Gamma) \right) \right\|_2^2, \quad \text{s.t.} \quad \sum_{i=1}^{n_\Omega} \mathbf{C} \mathbf{A}_i \mathbf{g}_i^\Gamma(\hat{\mathbf{x}}_i^\Gamma) = \mathbf{0}. \quad (3)$$

If HR is not applied (i.e.,  $\mathbf{B}_i = \mathbf{I}$  in (3)), the ROM's computational savings are limited because evaluation of residuals  $(\widehat{\mathbf{x}}_i^\Omega, \widehat{\mathbf{x}}_i^\Gamma) \rightarrow (\mathbf{g}_i^\Omega(\widehat{\mathbf{x}}_i^\Omega), \mathbf{g}_i^\Gamma(\widehat{\mathbf{x}}_i^\Gamma)) \rightarrow \mathbf{r}_i(\mathbf{g}_i^\Omega(\widehat{\mathbf{x}}_i^\Omega), \mathbf{g}_i^\Gamma(\widehat{\mathbf{x}}_i^\Gamma))$  scales with the size  $N_i^\Omega$  and  $N_i^\Gamma$  of the FOM. Thus, HR is applied to decrease the computational complexity caused by the nonlinearity of  $\mathbf{r}_i$ , and increase the computational speedup. We use [11, Algo. 3] to greedily compute a row sampling matrix  $\mathbf{B}_i$  for collocation HR. The application of HR to the decoders  $\mathbf{g}_i^\Omega$  and  $\mathbf{g}_i^\Gamma$  is discussed further in Sec. 3.1. Following [25], we apply a Gaussian test matrix  $\mathbf{C} \in \mathbb{R}^{n_C \times N_a}$ ,  $n_C \ll N_a$ , to convert the compatibility constraints into a so-called ‘‘weak compatibility constraint’’, which decreases the number of constraints to avoid making the DD ROM over-determined.

The DD FOM (2) and DD NM-ROM (3) are solved using an inexact Lagrange-Newton sequential quadratic programming (SQP) solver, where the Hessian of the Lagrangian is replaced with a Gauss-Newton approximation. This avoids computation of second order derivatives of residuals and constraints in (3), but still achieves good convergence for (2) and (3). For further details, see [16].

The DD NM-ROM (3) formulation has several benefits. Training, i.e., computation of the  $\mathbf{g}_i^\Omega$  and  $\mathbf{g}_i^\Gamma$  is local, involves few parameters, and can be done in parallel. The ROMs can be adjusted to localized features of the problem, which may result in smaller ROMs. Parallelization can be used to speed up ROM computation/training and ROM execution.

### 3.1 NM-ROM architecture and training

We use single-layer, wide, and sparse decoders with smooth activation functions to represent the maps  $\mathbf{g}_i^\Omega$  and  $\mathbf{g}_i^\Gamma$ . The corresponding encoders, denoted  $\mathbf{h}_i^\Omega$  and  $\mathbf{h}_i^\Gamma$ , are also single-layer, wide, and sparse. Shallow networks are used for computational efficiency; fewer layers correspond to fewer repeated matrix-vector multiplications when evaluating the decoders. The shallow depth necessitates a wide network to maintain enough expressiveness for use in NM-ROM. Smooth activations (i.e., swish) are used to ensure that  $\mathbf{g}_i^\Omega$  and  $\mathbf{g}_i^\Gamma$  are continuously differentiable. Normalization and de-normalization layers are also applied at the encoder input and decoder output layers, respectively.

Sparsity is applied at the decoder output layer so that HR can be applied. The sparsity allows one to compute a *subnet*, which only keeps track of the hidden nodes required to compute the output nodes that remain after HR. Further details can be found in [32, Sec. 3.2], [16, Sec. 5.3]. We also apply a sparsity mask to the encoder input layer so that the autoencoders are symmetric across the latent layer. The sparsity pattern has a tri-banded structure inspired by 2D finite difference stencils, where the number of nonzeros per band and the separation between bands are hyper-parameters.

To train the autoencoders, we first generate FOM snapshots in an *offline* stage by solving (2) at parameters  $\{\boldsymbol{\mu}_\ell\}_{\ell=1}^M$ , and collect interior- and interface-state snapshot datasets  $\mathbf{X}_i^\Omega \in \mathbb{R}^{N_i^\Omega \times M}$  and  $\mathbf{X}_i^\Gamma \in \mathbb{R}^{N_i^\Gamma \times M}$ . Alternatively, one can solve the monolithic FOM (1) at each  $\boldsymbol{\mu}_\ell$  and restrict the corresponding states  $\mathbf{x}(\boldsymbol{\mu}_\ell)$  to interior-states  $\mathbf{x}_i^\Omega(\boldsymbol{\mu}_\ell)$  and interface-states  $\mathbf{x}_i^\Gamma(\boldsymbol{\mu}_\ell)$  for each subdomain. We use the latter approach. The autoencoders  $(\mathbf{h}_i^\Omega, \mathbf{g}_i^\Omega)$  and  $(\mathbf{h}_i^\Gamma, \mathbf{g}_i^\Gamma)$  are then trained in parallel by minimizing the respective MSE losses

$$\mathcal{L}_i^\Omega = \frac{1}{M} \sum_{\ell=1}^M \left\| \mathbf{x}_i^\Omega(\boldsymbol{\mu}_\ell) - \mathbf{g}_i^\Omega(\mathbf{h}_i^\Omega(\mathbf{x}_i^\Omega(\boldsymbol{\mu}_\ell))) \right\|_2^2, \quad \mathcal{L}_i^\Gamma = \frac{1}{M} \sum_{\ell=1}^M \left\| \mathbf{x}_i^\Gamma(\boldsymbol{\mu}_\ell) - \mathbf{g}_i^\Gamma(\mathbf{h}_i^\Gamma(\mathbf{x}_i^\Gamma(\boldsymbol{\mu}_\ell))) \right\|_2^2 \quad (4)$$

for each subdomain  $i = 1, \dots, n_\Omega$ . The snapshots undergo a random 90-10 split for training and validation, and the MSE loss is minimized using the Adam optimizer over 2000 epochs with a batch size of 32. We also apply early stopping [41] with a patience of 300 and reduce the learning rate on plateau with an initial learning rate of  $10^{-3}$ . The implementation was done in PyTorch and used the PyTorch Sparse and SparseLinear packages.

## 4 Numerical experiment: 2D Burgers' equation

We compare the DD LS-ROM of [25] and the proposed DD NM-ROM with HR for the 2D steady-state Burgers equation. The DD LS-ROM can be regarded as a specific instance within the realm of DD NM-ROMs, where the encoders and decoders defined in Equation (4) are exchanged for linear

operators derived through singular value decomposition. We compute the relative error as

$$e = \left( \frac{1}{n_\Omega} \sum_{i=1}^{n_\Omega} \left( \left\| \mathbf{x}_i^\Omega - \mathbf{g}_i^\Omega(\widehat{\mathbf{x}}_i^\Omega) \right\|_2^2 + \left\| \mathbf{x}_i^\Gamma - \mathbf{g}_i^\Gamma(\widehat{\mathbf{x}}_i^\Gamma) \right\|_2^2 \right) / \left( \left\| \mathbf{x}_i^\Omega \right\|_2^2 + \left\| \mathbf{x}_i^\Gamma \right\|_2^2 \right) \right)^{1/2}. \quad (5)$$

All training and computations were performed on the Lassen machine at Lawrence Livermore National Laboratory, which consists of an IBM Power9 processor with NVIDIA V100 (Volta) GPUs, clock speed between 2.3-3.8 GHz, and 256 GB DDR4 memory. The code can be found at [https://anonymous.4open.science/r/DDNMROM\\_NeurIPS-4160/](https://anonymous.4open.science/r/DDNMROM_NeurIPS-4160/).

The implementation was done sequentially, but to highlight potential advantages of a parallel implementation, the reported wall clock time for computing subdomain-specific quantities for the SQP solver is taken to be the largest wall clock time incurred among all subdomains. The wall clock time for the remaining steps of the SQP solver is set to the overall wall clock time.

We consider the 2D steady-state Burgers' equation

$$u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = \nu \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), \quad u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} = \nu \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) \quad (6)$$

for  $(x, y) \in [-1, 1] \times [0, 0.05]$  with viscosity  $\nu = 0.1$ . As in [25], we use the exact solution  $u_{ex} = -2\nu \frac{\partial}{\partial x} \psi / \psi$ ,  $v_{ex} = -2\nu \frac{\partial}{\partial y} \psi / \psi$ , where  $\psi(x, y; a, \lambda) = a(1+x) + (e^{\lambda(x-1)} + e^{-\lambda(x-1)}) \cos(\lambda y)$  and  $(a, \lambda)$  are parameters, and its restriction to the boundary as Dirichlet boundary conditions. The PDE is discretized using centered finite differences with 482 uniformly spaced grid points in the  $x$ -direction and 26 uniformly spaced grid points in the  $y$ -direction. For ROM training, we collected 6400 FOM snapshots corresponding to varying  $(a, \lambda) \in [1, 10^4] \times [5, 25]$  (see Fig. 1) in a uniform  $80 \times 80$  grid. We use ROMs to predict the out-of-sample case  $(a, \lambda) = (7692.5384, 21.9230)$ .

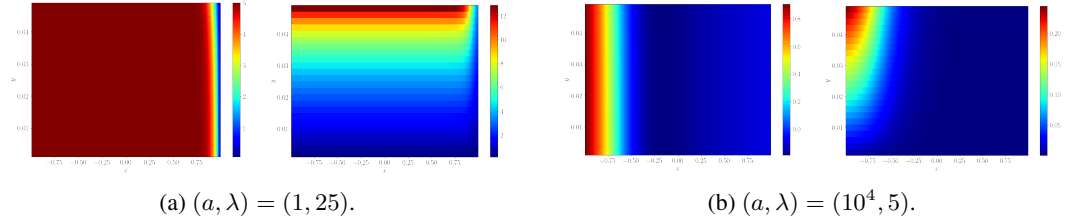


Figure 1: FOM  $u$  and  $v$  components for different  $(a, \lambda)$ . The distance of the shock from the left boundary and its steepness are determined by  $a$  and  $\lambda$ , respectively.

First we use DD problem with 4 uniformly sized subdomains in a  $2 \times 2$  configuration and vary the ROM sizes  $n_i^\Omega$  and  $n_i^\Gamma$ . Table 1 shows that NM-ROM has an order of magnitude lower error than LS-ROM with and without HR when comparing ROMs of the same size. In the non-HR case, LS-ROM only achieves order  $10^{-3}$  error for a ROM with 96 total DoF (error =  $2.66 \times 10^{-3}$ ), while NM-ROM can achieve a similar error with only 36 DoF (error =  $2.42 \times 10^{-3}$ ) and a higher speedup (speedup = 26.2) compared to LS-ROM with similar accuracy (speedup = 18.3). LS-ROM achieves a much higher speedup in the HR cases while retaining similar errors from the non-HR cases. NM-ROM also retains high accuracy after HR, and gains an extra 15-20 times speedup after applying HR.

Next we examine the per-subdomain reduction in the required number of autoencoder parameters for different subdomain configurations compared to the monolithic single-domain NM-ROM. We use the notation  $2 \times 1$  subdomains to indicate 2 subdomains in the  $x$ -direction and 1 subdomain in the  $y$ -direction. As expected, from Table 2, we see that the maximum number of NN parameters per subdomain decreases significantly as more subdomains are used. Furthermore, the total number of NN parameters in the DD cases also decreases relative to the single-domain case. We also note that the error increases as more subdomains are used. We kept the ROM size  $(n_i^\Omega, n_i^\Gamma) = (6, 3)$  constant for each subdomain configuration to isolate the effect of DD on the number of NN parameters, but this may cause overfitting in the 16 subdomain case. More careful hyper-parameter tuning is necessary to mitigate increases in error as the number of subdomains is increased.

	$n_i^\Omega$	$n_i^\Gamma$	DoF	Error	Speedup	Error (HR)	Speedup (HR)
LS-ROM	6	3	36	$2.06 \times 10^{-2}$	48.7	$1.78 \times 10^{-2}$	340.0
	8	4	48	$1.98 \times 10^{-2}$	30.0	$1.44 \times 10^{-2}$	347.6
	10	5	60	$1.50 \times 10^{-2}$	16.3	$1.16 \times 10^{-2}$	329.6
	16	8	96	$2.66 \times 10^{-3}$	18.3	$3.23 \times 10^{-3}$	280.4
NM-ROM	6	3	36	$2.42 \times 10^{-3}$	26.2	$2.60 \times 10^{-3}$	44.7
	8	4	48	$1.28 \times 10^{-3}$	21.7	$1.64 \times 10^{-3}$	43.9
	10	5	60	$1.09 \times 10^{-3}$	15.0	$1.19 \times 10^{-3}$	43.6
	16	8	96	$7.87 \times 10^{-4}$	13.9	$9.80 \times 10^{-4}$	37.5

Table 1: Relative error and speedup for LS-ROM and NM-ROM with and without HR for varying ROM size. We use  $N_i^B = 100$  HR nodes per subdomain in the HR case.

Subdomains	Max # subdomain params.	Reduction	Total # params.	Error
$1 \times 1$	$2.995 \times 10^6$	0.0 %	$2.995 \times 10^6$	$1.08 \times 10^{-3}$
$2 \times 1$	$1.147 \times 10^6$	61.7 %	$2.307 \times 10^6$	$1.27 \times 10^{-3}$
$2 \times 2$	$5.257 \times 10^5$	82.4 %	$2.384 \times 10^6$	$2.42 \times 10^{-3}$
$4 \times 2$	$2.617 \times 10^5$	91.3 %	$2.391 \times 10^6$	$4.26 \times 10^{-3}$
$8 \times 2$	$1.297 \times 10^5$	95.7 %	$2.406 \times 10^6$	$4.58 \times 10^{-2}$

Table 2: Max number of NN parameters per subdomain, the per-subdomain reduction in number of NN parameters, the total number of parameters, and the corresponding error for different subdomain configurations. For the single-domain case, an NM-ROM of dimension  $n = 9$  is used. For the DD cases,  $(n_i^\Omega, n_i^\Gamma) = (6, 3)$ , resulting in 9 DoF per subdomain. HR was not used to evaluate the NM-ROMs in these examples.

## 5 Conclusion

We extended the DD framework of [25] and compute ROMs using NM-ROM with HR as presented in [32]. Our experiments on the 2D Burgers’ equation show that NM-ROM achieves an order of magnitude lower relative error than LS-ROM in nearly all cases tested. While LS-ROM with HR achieves much higher speedup than NM-ROM with HR, NM-ROM is still the clear winner in terms of ROM accuracy for a given ROM size. Moreover, HR allows NM-ROM to gain an extra 15-20 time speedup compared to the non-HR cases. While the speedup is not as drastic as for LS-ROM, these speedup gains for NM-ROM are the highest that have been achieved for NM-ROM to our knowledge. We also showed that using the DD approach significantly decreases the number of required NN parameters per subdomain compared to the monolithic single-domain NM-ROM. In future work, we plan to apply DD NM-ROM to more challenging problems, including those with slowly decaying Kolmogorov  $n$ -width and to time-dependent problems. Other directions for future research include a greedy sampling strategy when choosing which FOM snapshots to compute for NM-ROM training and applying the DD NM-ROM framework to decomposable or component-based systems.

## Acknowledgments and Disclosure of Funding

This work was performed at Lawrence Livermore National Laboratory. A. N. Diaz was supported for this work by a Defense Science and Technology Internship (DSTI) at Lawrence Livermore National Laboratory and a 2021 National Defense Science and Engineering Graduate Fellowship. Y. Choi was supported for this work by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, as part of the CHaRMNET Mathematical Multifaceted Integrated Capability Center (MMICC) program, under Award Number DE-SC0023164 and partially by LDRD (21-SI-006). M. Heinkenschloss was supported by AFOSR Grant FA9550-22-1-0004 at Rice University. Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344. IM review: LLNL-CONF-854737.

## References

- [1] P. F. Antonietti, P. Pacciarini, and A. Quarteroni. A discontinuous Galerkin reduced basis element method for elliptic problems. *ESAIM Math. Model. Numer. Anal.*, 50(2):337–360, 2016. doi: 10.1051/m2an/2015045. URL <https://doi.org/10.1051/m2an/2015045>.
- [2] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*, volume 6 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005. doi: 10.1137/1.9780898718713. URL <https://doi.org/10.1137/1.9780898718713>.
- [3] A. C. Antoulas, I. V. Gosea, and A. C. Ionita. Model reduction of bilinear systems in the Loewner framework. *SIAM J. Sci. Comput.*, 38(5):B889–B916, 2016. URL <https://doi.org/10.1137/15M1041432>.
- [4] A. C. Antoulas, C. A. Beattie, and S. Gugercin. *Interpolatory Model Reduction*, volume 21 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2020. doi: 10.1137/1.9781611976083. URL <https://doi.org/10.1137/1.9781611976083>.
- [5] J. Barnett, I. Tezaur, and A. Mota. The Schwarz alternating method for the seamless coupling of nonlinear reduced order models and full order models. *arXiv:2210.12551*, 2022. doi: 10.48550/ARXIV.2210.12551. URL <https://doi.org/10.48550/ARXIV.2210.12551>.
- [6] J. Barnett, C. Farhat, and Y. Maday. Neural-network-augmented projection-based model order reduction for mitigating the Kolmogorov barrier to reducibility. *J. Comput. Phys.*, 492:Paper No. 112420, 20, 2023. doi: 10.1016/j.jcp.2023.112420. URL <https://doi.org/10.1016/j.jcp.2023.112420>.
- [7] P. Benner and T. Breiten. Two-sided projection methods for nonlinear model order reduction. *SIAM J. Sci. Comput.*, 37(2):B239–B260, 2015. doi: 10.1137/14097255X. URL <http://dx.doi.org/10.1137/14097255X>.
- [8] P. Benner and T. Breiten. Chapter 6: Model order reduction based on system balancing. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation: Theory and Algorithms*, Computational Science and Engineering, pages 261–295, Philadelphia, 2017. SIAM. doi: 10.1137/1.9781611974829.ch6. URL <https://doi.org/10.1137/1.9781611974829.ch6>.
- [9] M. Buffoni, H. Telib, and A. Iollo. Iterative methods for model reduction by domain decomposition. *Comput. & Fluids*, 38(6):1160–1167, 2009. doi: 10.1016/j.compfluid.2008.11.008. URL <https://doi.org/10.1016/j.compfluid.2008.11.008>.
- [10] K. Carlberg, Y. Choi, and S. Sargsyan. Conservative model reduction for finite-volume models. *Journal of Computational Physics*, 371:280–314, 2018. doi: 10.1016/j.jcp.2018.05.019. URL <https://doi.org/10.1016/j.jcp.2018.05.019>.
- [11] K. T. Carlberg, C. Farhat, J. Cortial, and D. Amsallem. The GNAT method for nonlinear model reduction: Effective implementation and application to computational fluid dynamics and turbulent flows. *Journal of Computational Physics*, 242:623 – 647, 2013. doi: 10.1016/j.jcp.2013.02.028. URL <http://dx.doi.org/10.1016/j.jcp.2013.02.028>.
- [12] S. W. Cheung, Y. Choi, D. M. Copeland, and K. Huynh. Local lagrangian reduced-order modeling for the rayleigh-taylor instability by solution manifold decomposition. *Journal of Computational Physics*, 472:111655, 2023. doi: 10.1016/j.jcp.2022.111655. URL <https://doi.org/10.1016/j.jcp.2022.111655>.
- [13] Y. Choi and K. Carlberg. Space–time least-squares petrov–galerkin projection for nonlinear model reduction. *SIAM Journal on Scientific Computing*, 41(1):A26–A58, 2019. doi: 10.1137/17M1120531. URL <https://doi.org/10.1137/17M1120531>.
- [14] Y. Choi, P. Brown, W. Arrighi, R. Anderson, and K. Huynh. Space–time reduced order model for large-scale linear dynamical systems with application to boltzmann transport problems. *Journal of Computational Physics*, 424:109845, 2021. doi: 10.1016/j.jcp.2020.109845. URL <https://doi.org/10.1016/j.jcp.2020.109845>.

- [15] D. M. Copeland, S. W. Cheung, K. Huynh, and Y. Choi. Reduced order models for lagrangian hydrodynamics. *Computer Methods in Applied Mechanics and Engineering*, 388:114259, 2022. doi: 10.1016/j.cma.2021.114259. URL <https://doi.org/10.1016/j.cma.2021.114259>.
- [16] A. N. Diaz, Y. Choi, and M. Heinkenschloss. A fast and accurate domain-decomposition nonlinear manifold reduced order model. *arXiv:2305.15163v1*, 2023. doi: 10.48550/arXiv.2305.15163. URL <https://doi.org/10.48550/arXiv.2305.15163>.
- [17] J. L. Eftang and A. T. Patera. Port reduction in parametrized component static condensation: approximation and *a posteriori* error estimation. *Internat. J. Numer. Methods Engrg.*, 96(5): 269–302, 2013. doi: 10.1002/nme.4543. URL <https://doi.org/10.1002/nme.4543>.
- [18] J. L. Eftang, D. B. P. Huynh, D. J. Knezevic, E. M. Ronquist, and A. T. Patera. Adaptive port reduction in static condensation. *IFAC Proceedings Volumes*, 45(2):695–699, 2012. doi: 10.3182/20120215-3-AT-3016.00123. URL <https://doi.org/10.3182/20120215-3-AT-3016.00123>. 7th Vienna International Conference on Mathematical Modelling.
- [19] I. V. Gosea and A. C. Antoulas. Data-driven model order reduction of quadratic-bilinear systems. *Numer. Linear Algebra Appl.*, 25(6):e2200, 2018. doi: 10.1002/nla.2200. URL <http://dx.doi.org/10.1002/nla.2200>.
- [20] C. Gu. QLMOR: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 30(9):1307–1320, sept. 2011. doi: 10.1109/TCAD.2011.2142184. URL <https://doi.org/10.1109/TCAD.2011.2142184>.
- [21] M. Gubisch and S. Volkwein. Chapter 1: Proper Orthogonal Decomposition for linear-quadratic optimal control. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation: Theory and Algorithms*, Computational Science and Engineering, pages 3–64, Philadelphia, 2017. SIAM. doi: 10.1137/1.9781611974829.ch1. URL <https://doi.org/10.1137/1.9781611974829.ch1>.
- [22] B. Haasdonk. Chapter 2: Reduced basis methods for parametrized PDEs - a tutorial introduction for stationary and instationary problems. In P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, editors, *Model Reduction and Approximation: Theory and Algorithms*, Computational Science and Engineering, pages 65–136. SIAM, Philadelphia, 2017. doi: 10.1137/1.9781611974829.ch2. URL <https://doi.org/10.1137/1.9781611974829.ch2>.
- [23] D. Hartman and L. K. Mestha. A deep learning framework for model reduction of dynamical systems. In *2017 IEEE Conference on Control Technology and Applications (CCTA)*, pages 1917–1922, 2017. doi: 10.1109/CCTA.2017.8062736. URL <https://doi.org/10.1109/CCTA.2017.8062736>.
- [24] M. Hinze and S. Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In P. Benner, V. Mehrmann, and D. C. Sorensen, editors, *Dimension Reduction of Large-Scale Systems*, Lecture Notes in Computational Science and Engineering, Vol. 45, pages 261–306, Heidelberg, 2005. Springer-Verlag. doi: 10.1007/3-540-27909-1\_10. URL [http://doi.org/10.1007/3-540-27909-1\\_10](http://doi.org/10.1007/3-540-27909-1_10).
- [25] C. Hoang, Y. Choi, and K. Carlberg. Domain-decomposition least-squares Petrov-Galerkin (DD-LSPG) nonlinear model reduction. *Comput. Methods Appl. Mech. Engrg.*, 384:Paper No. 113997, 41, 2021. doi: 10.1016/j.cma.2021.113997. URL <https://doi.org/10.1016/j.cma.2021.113997>.
- [26] D. B. P. Huynh, D. J. Knezevic, and A. T. Patera. A static condensation reduced basis element method: approximation and *a posteriori* error estimation. *ESAIM Math. Model. Numer. Anal.*, 47(1):213–251, 2013. doi: 10.1051/m2an/2012022. URL <https://doi.org/10.1051/m2an/2012022>.
- [27] L. Iapichino, A. Quarteroni, and G. Rozza. A reduced basis hybrid method for the coupling of parametrized domains represented by fluidic networks. *Comput. Methods Appl. Mech. Engrg.*, 221/222:63–82, 2012. doi: 10.1016/j.cma.2012.02.005. URL <https://doi.org/10.1016/j.cma.2012.02.005>.

- [28] A. Iollo, G. Sambataro, and T. Taddei. A one-shot overlapping Schwarz method for component-based model reduction: application to nonlinear elasticity. *Comput. Methods Appl. Mech. Engrg.*, 404:Paper No. 115786, 32, 2023. doi: 10.1016/j.cma.2022.115786. URL <https://doi.org/10.1016/j.cma.2022.115786>.
- [29] K. Kashima. Nonlinear model reduction by deep autoencoder of noise response data. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 5750–5755, 2016. doi: 10.1109/CDC.2016.7799153. URL <https://doi.org/10.1109/CDC.2016.7799153>.
- [30] Y. Kim, Y. Choi, D. Widemann, and T. Zohdi. Efficient nonlinear manifold reduced order model. *arXiv preprint arXiv:2011.07727*, 2020. doi: 10.48550/arXiv.2011.07727. URL <https://doi.org/10.48550/arXiv.2011.07727>.
- [31] Y. Kim, K. Wang, and Y. Choi. Efficient space–time reduced order model for linear dynamical systems in python using less than 120 lines of code. *Mathematics*, 9(14):1690, 2021. doi: 10.3390/math9141690. URL <https://doi.org/10.3390/math9141690>.
- [32] Y. Kim, Y. Choi, D. Widemann, and T. Zohdi. A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder. *J. Comput. Phys.*, 451:Paper No. 110841, 29, 2022. doi: 10.1016/j.jcp.2021.110841. URL <https://doi.org/10.1016/j.jcp.2021.110841>.
- [33] K. Lee and K. T. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *J. Comput. Phys.*, 404:108973, 32, 2020. doi: 10.1016/j.jcp.2019.108973. URL <https://doi.org/10.1016/j.jcp.2019.108973>.
- [34] K. Li, K. Tang, T. Wu, and Q. Liao. D3M: A deep domain decomposition method for partial differential equations. *IEEE Access*, 8:5283–5294, 2020. doi: 10.1109/ACCESS.2019.2957200. URL <https://doi.org/10.1109/ACCESS.2019.2957200>.
- [35] S. Li, Y. Xia, Y. Liu, and Q. Liao. A deep domain decomposition method based on fourier features. *Journal of Computational and Applied Mathematics*, 423:114963, 2023. doi: 10.1016/j.cam.2022.114963. URL <https://doi.org/10.1016/j.cam.2022.114963>.
- [36] W. Li, X. Xiang, and Y. Xu. Deep domain decomposition method: Elliptic problems. In J. Lu and R. Ward, editors, *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, volume 107 of *Proceedings of Machine Learning Research*, pages 269–286. PMLR, 20–24 Jul 2020. URL <https://proceedings.mlr.press/v107/li20a.html>.
- [37] Y. Maday and E. M. Rønquist. A reduced-basis element method. *J. Sci. Comput.*, 17(1-4):447–459, 2002. doi: 10.1023/A:1015197908587. URL <https://doi.org/10.1023/A:1015197908587>.
- [38] Y. Maday and E. M. Rønquist. The reduced basis element method: application to a thermal fin problem. *SIAM J. Sci. Comput.*, 26(1):240–258, 2004. doi: 10.1137/S1064827502419932. URL <https://doi.org/10.1137/S1064827502419932>.
- [39] A. J. Mayo and A. C. Antoulas. A framework for the solution of the generalized realization problem. *Linear Algebra Appl.*, 425(2-3):634–662, 2007. doi: 10.1016/j.laa.2007.03.008. URL <https://doi.org/10.1016/j.laa.2007.03.008>.
- [40] M. Ohlberger and S. Rave. Reduced basis methods: Success, limitations and future challenges. *Proceedings of the Conference Algorithmy*, pages 1–12, 2016. URL <http://www.iam.fmph.uniba.sk/amuc/ojs/index.php/algorithmy/article/view/389>.
- [41] L. Prechelt. Automatic early stopping using cross validation: quantifying the criteria. *Neural networks*, 11(4):761–767, 1998. doi: 10.1016/S0893-6080(98)00010-0. URL [https://doi.org/10.1016/S0893-6080\(98\)00010-0](https://doi.org/10.1016/S0893-6080(98)00010-0).
- [42] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations. An Introduction*, volume 92 of *Unitext*. Springer, Cham, 2016. doi: 10.1007/978-3-319-15431-2. URL <https://doi.org/10.1007/978-3-319-15431-2>.



- [43] K. Smetana and T. Taddei. Localized model reduction for nonlinear elliptic partial differential equations: localized training, partition of unity, and adaptive enrichment. *arXiv:2202.09872v1*, 2022. doi: 10.48550/ARXIV.2202.09872. URL <https://doi.org/10.48550/ARXIV.2202.09872>.
- [44] Q. Sun, X. Xu, and H. Yi. Domain decomposition learning methods for solving elliptic problems. *arXiv preprint arXiv:2207.10358*, 2022. doi: 10.48550/arXiv.2207.10358. URL <https://doi.org/10.48550/arXiv.2207.10358>.