LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-815070

Nonlinear manifold-based reduced order model

Y. Choi

September 29, 2020

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

FULL TECHNICAL FINAL REPORT

Nonlinear Manifold-based reduced order model

Youngsoo Choi

20-FS-007

Abstract

Traditional linear subspace reduced order models (LS-ROMs) are able to accelerate physical simulations, in which the intrinsic solution space falls into a subspace with a small dimension, i.e., the solution space has a small Kolmogorov n-width. However, for physical phenomena not of this type, e.g., any advection-dominated flow phenomena such as in traffic flow, atmospheric flows, and air flow over vehicles, a low-dimensional linear subspace poorly approximates the solution. To address cases such as these, we have developed a fast and accurate physics-informed neural network ROM, namely nonlinear manifold ROM (NM-ROM), which can better approximate high-fidelity model solutions with a smaller latent space dimension than the LS-ROMs. Our method takes advantage of the existing numerical methods that are used to solve the corresponding full order models. The efficiency is achieved by developing a hyper-reduction technique in the context of the NM-ROM. Numerical results show that neural networks can learn a more efficient latent space representation on advection-dominated data from 1D and 2D Burgers' equations. A speedup of up to 2.6 for 1D Burgers' and a speedup of 11.7 for 2D Burgers' equations are achieved with an appropriate treatment of the nonlinear terms through a hyper-reduction technique. Finally, a posteriori error bounds for the NM-ROMs are derived that take account of the hyper-reduced operators.

Background and Research Objectives

Physical simulations are influencing developments in science, engineering, and technology more rapidly than ever before. However, high-fidelity, forward physical simulations are computationally expensive and, thus, make intractable any decision-making applications, such as design optimization, inverse problems, optimal controls, and uncertainty quantification, for which many forward simulations are required to explore the parameter space in the outer loop.

To compensate for the computational expense issue, the projection-based reduced order models (ROMs) take advantage of both the known governing equation and the data. ROMs generate the solution data from the corresponding physical simulations and then compress the data to find an intrinsic solution subspace, which is represented by a linear combination of basis vectors, i.e., LS-ROMs. This condensed solution representation is plugged back into the (semi-)discretized governing equation to reduce the number of unknowns, resulting in an over-determined system, i.e., more equations than unknowns. Note that the full governing equations are used to constrain the LS-ROM through this substitution. Therefore, this can be considered as a physics-informed surrogate model. Additionally, the existing numerical methods for the corresponding full order model (FOM) is utilized in the LS-ROM solution process. Therefore, the LS-ROM fully respects the original discretization of the governing equations that describe/approximate the underlying physical laws, unlike black-box approaches.

In spite of its successes, the linear subspace solution representation suffers from not being able to represent certain physical simulation solutions with a small basis dimension, such as advection-dominated or sharp gradient solutions. This is because LS-ROMs work only for physical problems in which the intrinsic solution space falls into a subspace with a small dimension, i.e., the solution space has a small Kolmogorov n-width. Unfortunately, even though problems that are advection-dominated or have sharp gradient solutions are important, they do not have small Kolmogorov n-width. Such physical simulations include, but are not limited to, the hyperbolic equations with high Reynolds number, the Boltzmann transport equations, and the traffic flow simulations.

In order to overcome the issues that arise from LS-ROM, we have developed a fast and accurate physics-informed neural network ROM with a nonlinear manifold solution representation, i.e., the nonlinear manifold ROM (NM-ROM). The NM-ROM is able to accelerate advection-dominated simulations with a small latent space. It has a great solution representability due to the transition from linear subspace to nonlinear manifold.

Scientific Approach and Accomplishments

The detailed description of the technical approach can be found in [1]. Briefly explaining the technical approach here, a parameterized nonlinear dynamical system is considered, characterized by a system of nonlinear ordinary differential equations (ODEs), which can be considered as a resultant system from semi-discretization of Partial Differential Equations (PDEs) in space domains. The state variable is represented by the nonlinear manifold, i.e.,

(1) $$\boldsymbol{w} \approx \tilde{\boldsymbol{w}} = \boldsymbol{w}_{\text{ref}} + \boldsymbol{g}(\hat{\boldsymbol{w}}),$$

where w is a state variable and g is a nonlinear function that maps from the low dimensional latent space to full space dimension. The nonlinear map is constructed by training an autoencoder, using solution data from high-fidelity model. **Figure 1** depicts an autoencoder, which compress the original high-fidelity model data to a latent space via an encoder and decompress it back to an approximate output via a decoder. Notice that the decoder maps the latent space dimension to a full order model dimension, which we use as a nonlinear manifold solution representation.

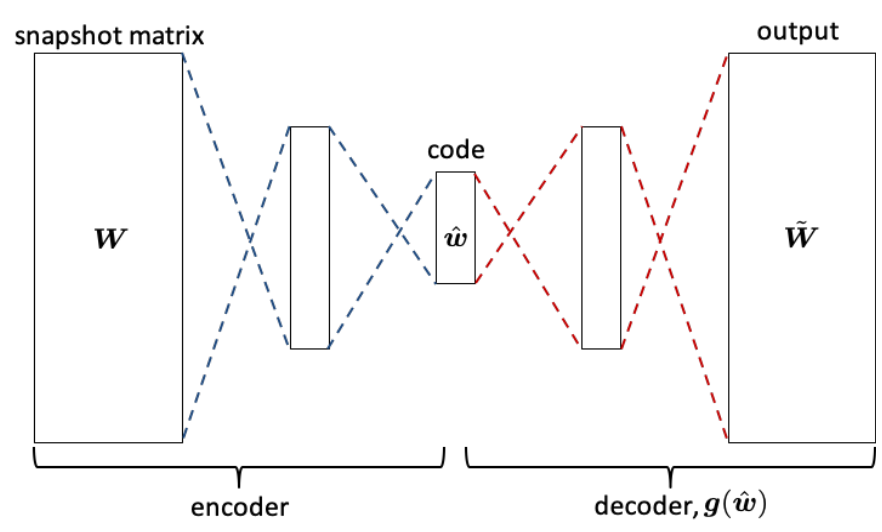

Figure 1: A schematic description of an autoencoder.

A crucial development of our NM-ROM is the structure of the decoder. We use a shallow

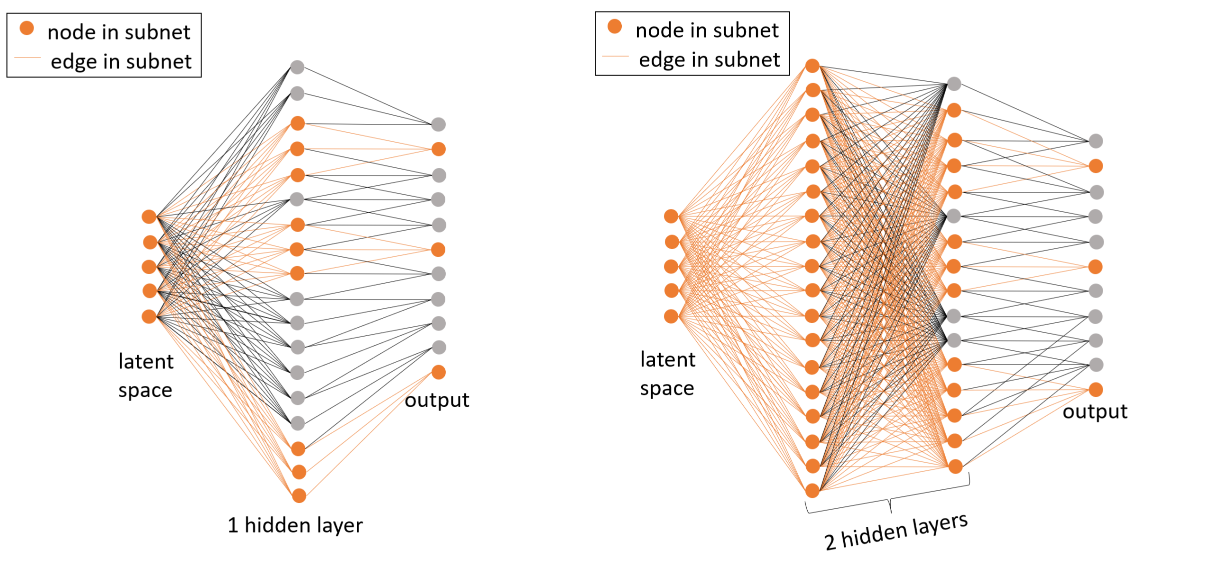

Figure 2: Illustration of the effect on the sparsity of the active path for shallow network vs deep network. The shallow network provides a sparser network than the deep network in the subnet. Therefore, the shallow network is expected to achieve a higher speed-up than the deep network.

masked neural network in order to achieve a sparsity structure that is required for the successful hyper-reduction technique. **Figure 2** compares a shallow and deep neural network in the context of sampled hyper-reduced outputs. The orange disks and edges represent the active paths that are actually used to compute the selected outputs. The gray nodes and edges are completely ignored in the computation. Note that the shallow network is able to provide much sparser structure than the deep one, which explains why we use a shallow neural network as a decoder.

The nonlinear manifold solution representation in Eq. (1) is plugged into the governing equation, resulting in an over-determined system, i.e., more equations than unknowns. There are two ways of closing the system: i) Galerkin (denoted as NM-Galerkin-HR) and (ii) least-squares Petrov–Galerkin (denoted as NM-LSPG-HR) approaches. The Galerkin approach first projects the over-determined ODE to the reduced continuous ODE, i.e.,

(2) $$\dot{\hat{\boldsymbol{w}}} = ((\boldsymbol{Z}^T\boldsymbol{\Phi}_r)^\dagger \boldsymbol{Z}^T \boldsymbol{J}_g(\hat{\boldsymbol{w}}))^\dagger (\boldsymbol{Z}^T\boldsymbol{\Phi}_r)^\dagger \boldsymbol{Z}^T \boldsymbol{f}(\boldsymbol{w}_{ref} + \boldsymbol{g}(\hat{\boldsymbol{w}}), t; \boldsymbol{\mu}).$$

Then, the time domain is discretized by a time integrator. For example, if the backward Euler time integrator is used, then the fully discretized reduced system for Galerkin approach becomes

(3) $$\hat{\boldsymbol{w}}_n = \hat{\boldsymbol{w}}_{n-1} + \Delta t((\boldsymbol{Z}^T\boldsymbol{\Phi}_r)^\dagger \boldsymbol{Z}^T \boldsymbol{J}_g(\hat{\boldsymbol{w}}))^\dagger (\boldsymbol{Z}^T\boldsymbol{\Phi}_r)^\dagger \boldsymbol{Z}^T \boldsymbol{f}(\boldsymbol{w}_{ref} + \boldsymbol{g}(\hat{\boldsymbol{w}}_n), t_n; \boldsymbol{\mu}).$$

On the other hand, the NM-LSPG-HR discretize the time domain first, resulting in the following nonlinear residual function if the backward Euler time integrator is used,

(4) $$\tilde{\boldsymbol{r}}^n_{BE}(\hat{\boldsymbol{w}}_n; \hat{\boldsymbol{w}}_{n-1}, \boldsymbol{\mu}) \equiv \boldsymbol{g}(\hat{\boldsymbol{w}}_n) - \boldsymbol{g}(\hat{\boldsymbol{w}}_{n-1}) - \Delta t \boldsymbol{f}(\boldsymbol{w}_{ref} + \boldsymbol{g}(\hat{\boldsymbol{w}}_n), t_n; \boldsymbol{\mu}).$$

Then, the nonlinear residual in (4) is closed by minimize the norm of the residual, i.e.,

(5) $$\hat{\boldsymbol{w}}_n = \underset{\hat{\boldsymbol{v}}\in\mathbb{R}^{n_s}}{\operatorname{argmin}} \frac{1}{2}\|(\boldsymbol{Z}^T\boldsymbol{\Phi}_r)^\dagger \boldsymbol{Z}^T \tilde{\boldsymbol{r}}_{BE}^n(\hat{\boldsymbol{v}};\hat{\boldsymbol{w}}_{n-1},\boldsymbol{\mu})\|_2^2.$$

Here Z is a sampling matrix and f is a nonlinear function from the underlying governing equation. It turns out that the NM-LSPG-HR is more robust than the NM-Galerkin-HR in term of accuracy. For example, we apply our NM-ROMs to 2D viscous Burgers equation, which can be described by

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

where we take Reynolds number, $Re = 1/\nu = 10,000$, in order to make the problem advection-dominated. The corresponding full order model is solved using the finite difference with spatial degrees of freedom of 3,600 and the backward Euler time integrator. **Figure 3** shows the comparison among four different reduced order models, i.e., LS-Galerkin is the LS-ROM with Galerkin projection, LS-LSPG is the LS-ROM with the least-squares Petrov–Galerkin projection. It also shows two projection error curves, i.e., LS proj. error is the error level, where LS-ROM can reach at best. On the other hand, NM proj. error is the error level, where NM-ROM can reach at best. The reduced dimension varies from 3 to 20, and the corresponding maximum relative errors are reported in Figure 3. Note that NM-LSPG is the best performed model. Note that the maximum relative error for NM-LSPG is even lower than the LS proj. error, exemplifying the great solution representability of the nonlinear manifold for the advection-dominated problems.

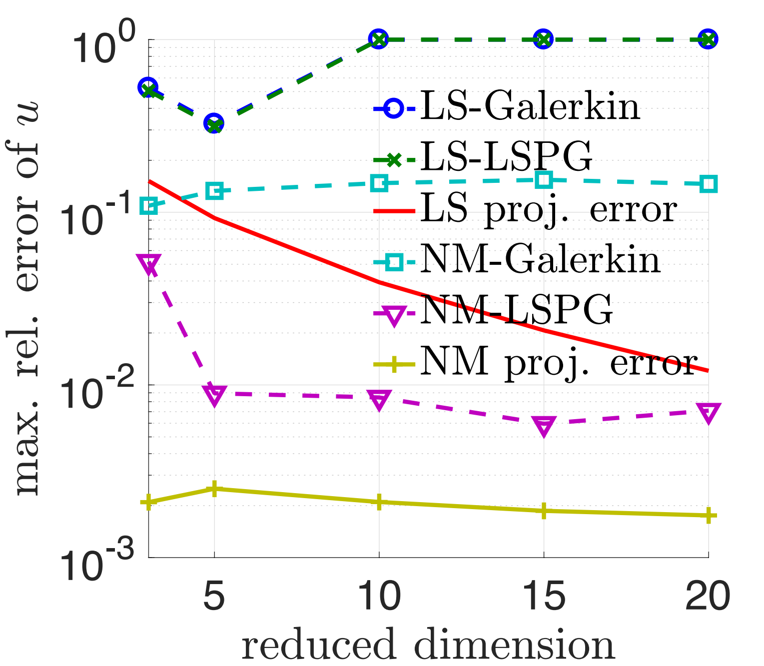

Figure 3: reduced dimension vs. maximum relative error

The speed-up of the NM-ROM is also achieved by applying an efficient hyper-reduction. **Table 1** compares both accuracy and speed-up between NM-LSPG-HR and LS-LSPG-HR for the 2D viscous Burgers equation. Although LS-LSPG-HR achieves a great speed-up (e.g., around 27), but its solution is very inaccurate (e.g., a relative error greater than 30%). On the other hand, our NM-LSPG-HR is able to achieve both a great speed-up (e.g., greater than 10) and accuracy (e.g., a relative error less than 1 %).

	NM-LSPG-HR			LS-LSPG-HR		
Residual basis dimension	55	56	51	59	53	53
The number of residual samples	58	59	54	59	58	59
Maximum relative error (%)	0.93	0.94	0.95	34.38	37.73	37.84
Wall-clock time (sec)	12.15	12.35	12.09	5.26	5.02	4.86
Speed-up	11.58	11.39	11.63	26.76	28.02	28.95

Table 1. The performance comparison between NM-LSPG-HR and LS-LSPG-HR.

Our NM-ROM can also predict the solution of a parameter point that was not used to build the ROM. Figure 4 shows how the maximum relative error of NM-ROMs varies as the parameter values vary. The NM-ROM is generated by using the full order model data from the parameter values, 0.9, 0.95, 1.05, and 1.1. The NM-ROM is able to generate good approximate solutions at the parameter points that were not used in the training phase.

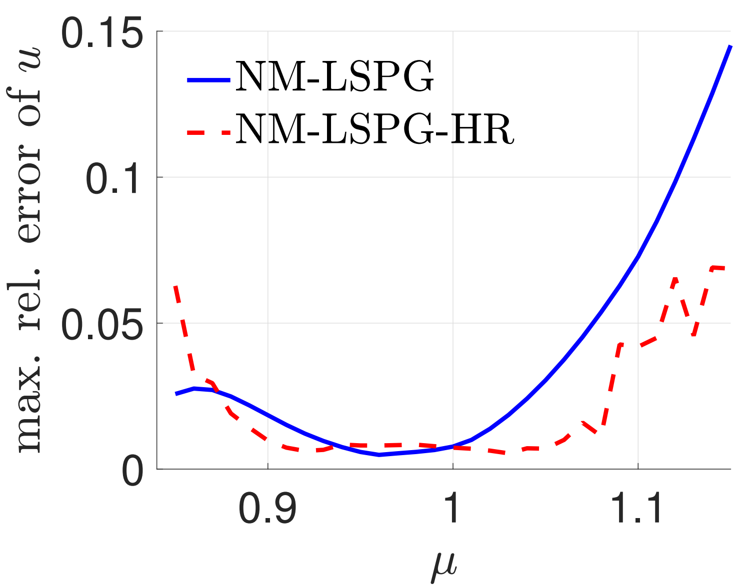

Figure 4: The comparison of the NM-LSPG-HR and NM-LSPG on the maximum relative error over the parameter variation.

Mission Impact

Reduced-order modeling is specifically named as an R&D priority in the following core competencies: the high-energy-density science, high-performance computing, simulation, and data science, and advanced materials and manufacturing in Investment Strategy for science and technology. N program will support this technology to be applied to hydrodynamics problems.

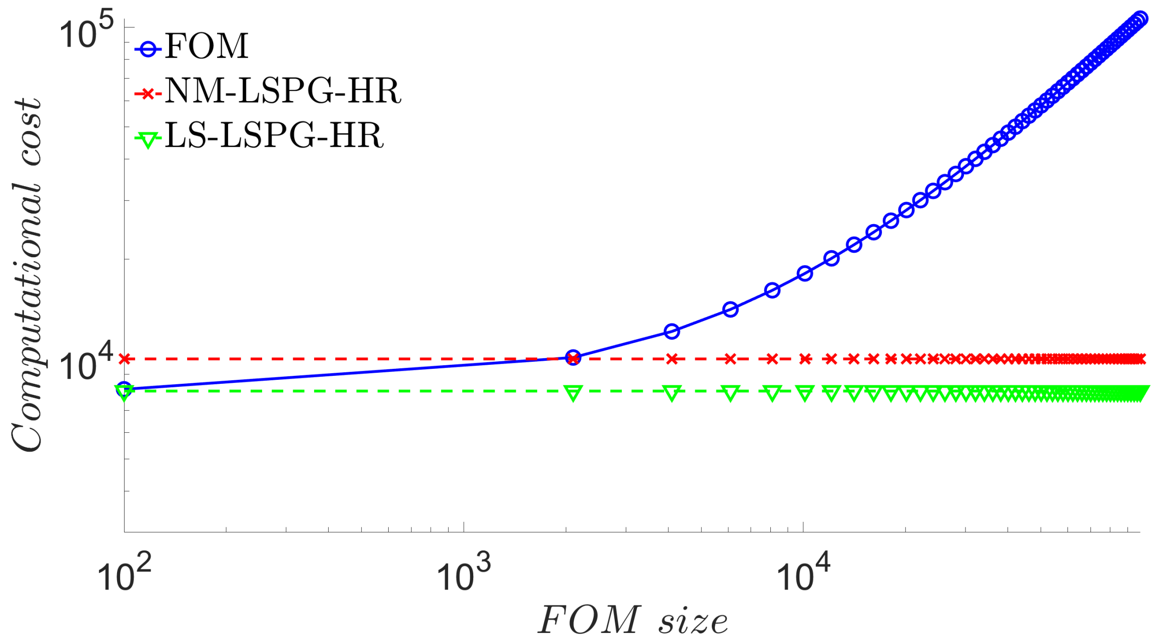

Figure 5: Computational cost vs full order model size. The figure shows that the higher the speed-up will be achieved; the larger the underlying full order model problems is.

Conclusion

The feasibility study has successfully demonstrated that the nonlinear manifold representation enables a reduced order model to achieve both accuracy and speed-up for small-scale problems. This verified methodology needs to be tested for large-scale problems with the funding in LDRD ER scale. Our flop count calculation shows that the NM-ROM will produce even higher speed-up if it is applied to a larger scale problem. See **Figure 5**.

References

[1] Kim, Youngkyu, Youngsoo Choi, David Widemann, Zohdi Tarek. "A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder." *arXiv preprint* (2020): arXiv:2009.11990

Publications and Presentations

Choi, Youngsoo, Gabriele Boncoraglio, Spenser Anderson, David Amsallem, and Charbel Farhat. "Gradient-based constrained optimization using a database of linear reduced-order models." *Journal of Computational Physics* (2020): 109787 [IM number: LLNL-JRNL-796678].

Hoang, Chi, Youngsoo Choi, Kevin Carlberg. "Domain-decomposition least-squares Petrov–Galerkin (DD-LSPG) nonlinear model reduction." *arXiv preprint* (2020): arXiv:2007.11835 [IM number: LLNL-JRNL-812648].

Kim, Youngkyu, Youngsoo Choi, David Widemann, Zohdi Tarek. "A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder." *arXiv preprint* (2020): arXiv:2009.11990 [IM number: LLNL-JRNL-814844].