Final Report for AEOLUS: Advances in Experimental Design, Optimal Control, and Learning for Uncertain Complex Systems

DOE Award: DE-SC0021659

Principal Investigator: Dr. Clayton Webster

Period of Performance: 06/15/2021 – 06/14/2024
UT Award: 202100051001AWD
UT Account: 26083980

Contents

1. Summary of scientific accomplishments

The AEOLUS Center is dedicated to developing a unified optimization-under-uncertainty framework for: (1) learning predictive models from data; and (2) optimizing experiments, processes, and designs governed by these models, all driven by complex, uncertain energy systems. AEOLUS addresses the critical need for principled, rigorous, scalable, and structure-exploiting capabilities for exploring parameter and decision spaces of complex forward simulation models. This report summarizes the key highlights of our research during the period of performance. We have made significant progress on several thrusts, including:

- a non-intrusive inference reduced order model for fluids using deep multistep neural networks [7];
- closure learning for nonlinear model reduction using deep residual neural networks [6];
- a data-driven learning framework for the analytic continuation of imaginary time using Adams Bashforth residual neural networks [5];
- an asymptotically compatible meshfree method for solving nonlocal equations with random coefficients [1];
- Gaussian smoothing gradient descent methods for minimizing high-dimensional functions [2];
- improved performance of stochastic gradients with Gaussian smoothing for neural network training [3]; and
- anisotropic Gaussian smoothing for gradient-based optimization [4].

The highlights from these works are described in §2–§5.

2. Data-driven reduced order models for fluid dynamics

Reduced order models (ROMs) are essential for enabling rapid simulation of complex fluid systems, but traditional projection-based approaches often suffer from instability when applied to nonlinear problems. In this project, we developed data-driven approaches that leverage deep neural networks to construct stable and accurate reduced order models without requiring intrusive modifications to existing simulation codes.

2.1. Non-intrusive inference reduced order model using deep multistep neural networks

In [7], we introduced a non-intrusive framework for learning reduced order models directly from simulation data. Unlike traditional Galerkin projection methods that require access to the governing equations and can produce unstable reduced systems, our approach learns the optimal reduced dynamics from data using deep neural networks based on linear multistep methods.

The key innovation is the use of linear multistep neural networks (LMNet) inspired by implicit Adams-Moulton schemes from numerical analysis. Given snapshots of the full order solution projected onto a reduced basis, LMNet learns to predict the evolution of the reduced coefficients over time. The implicit formulation provides enhanced stability compared to explicit methods, which is critical for long-time integration of chaotic fluid systems.

We demonstrated the approach on two-dimensional flow past a circular cylinder at Reynolds number 100. The results show that LMNet-ROM significantly outperforms standard projection-based ROMs in terms of both accuracy and stability. The non-intrusive nature of the method makes it applicable to complex legacy codes where modifying the solver is impractical or impossible.

2.2. Closure learning for nonlinear model reduction using deep residual neural networks

A fundamental challenge in reduced order modeling is the closure problem: when projecting nonlinear dynamics onto a low-dimensional subspace, unresolved scales interact with resolved scales in ways that

standard projection cannot capture. In [6], we addressed this challenge by using deep residual neural networks to learn closure terms that account for the effect of truncated modes.

Our approach augments the standard Galerkin ROM with a learned correction term. The residual neural network takes the current reduced state as input and outputs a correction that compensates for truncation errors. Training uses data from the full order model, with the network learning to minimize the discrepancy between the corrected ROM prediction and the true dynamics.

The residual network architecture is particularly well-suited to this task because the correction term represents a perturbation to the Galerkin dynamics rather than the full dynamics itself. This structure allows the network to focus on learning the closure contribution while the underlying physics is captured by the projection. Numerical experiments on fluid flow problems demonstrate that the closure-corrected ROM achieves substantially improved accuracy compared to both standard Galerkin ROMs and purely data-driven approaches, while maintaining computational efficiency suitable for many-query applications such as optimization and uncertainty quantification.

3. Analytic continuation of noisy data using Adams Bashforth residual neural network

The analytic continuation problem in quantum many-body physics requires recovering spectral functions from imaginary-time Green's function data. This inverse problem is fundamentally ill-posed, and standard Maximum Entropy (MaxEnt) methods struggle when the input data is corrupted by noise. In [5], we developed a data-driven learning framework using Adams Bashforth residual neural networks (AB-ResNet) that achieves superior accuracy under noisy conditions.

Our approach is inspired by the connection between residual networks and numerical methods for ordinary differential equations. We designed network architectures based on first, second, and third-order Adams-Bashforth schemes (AB1, AB2, AB3), where higher-order methods provide improved stability properties. The AB3-ResNet architecture is strongly stable, while AB1 and AB2 are only conditionally stable. This stability hierarchy directly impacts performance on noisy data.

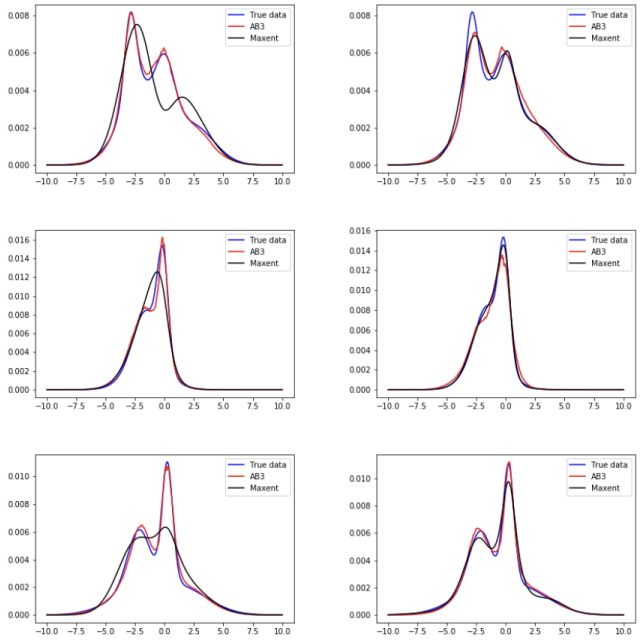


Figure 1: Spectral density functions recovered by AB3-ResNet and MaxEnt. Left column: noise level 10^{-2}. Right column: noise level 10^{-3}. AB-ResNet correctly identifies spectral peaks even under high noise conditions where MaxEnt fails.

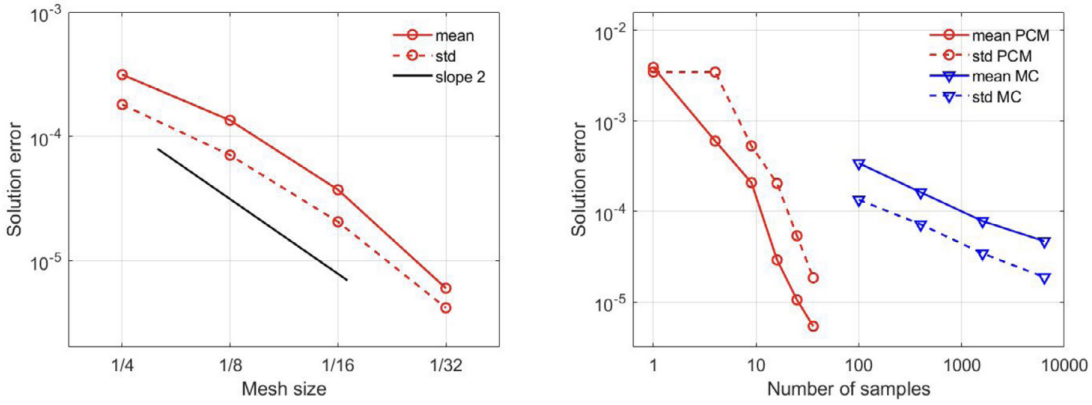
Figure 1 compares AB3-ResNet against MaxEnt across different noise levels. At low noise (10^{-3}), both methods accurately recover the spectral density. However, at higher noise levels (10^{-2}), MaxEnt fails to resolve the spectral peaks while AB-ResNet maintains accurate predictions. The mean absolute errors on the test dataset are 6.8×10^{-4}, 3.8×10^{-4}, and 2.6×10^{-4} for AB1, AB2, and AB3, respectively, confirming that higher-order methods yield higher accuracy.

Beyond accuracy improvements, AB-ResNet provides a direct mapping from Green's function to spectral density, avoiding the iterative optimization required by MaxEnt. The computational cost is reduced by an order of magnitude: AB-ResNet requires $\mathcal{O}(10)$ seconds compared to $\mathcal{O}(100)$ seconds for MaxEnt. This combination of improved noise robustness and computational efficiency makes AB-ResNet a practical tool for analytic continuation in quantum many-body calculations.

4. Meshfree method for solving nonlocal diffusion in heterogeneous media

Nonlocal diffusion models capture long-range interactions that classical local PDEs cannot represent, making them valuable for modeling anomalous transport in heterogeneous media. However, when the diffusion coefficient is uncertain, quantifying the resulting uncertainty in the solution requires solving the nonlocal equation many times for different parameter realizations, which quickly becomes computationally prohibitive. In [1], we developed an asymptotically compatible meshfree method combined with sparse grid stochastic collocation to efficiently solve nonlocal diffusion problems with random coefficients.

Our approach represents the random diffusivity using the Karhunen-Loève decomposition, reducing the infinite-dimensional stochastic input to a finite-dimensional parameter space. We then employ a probabilistic collocation method (PCM) with sparse grids to sample this parameter space efficiently. For each sample, we solve the deterministic nonlocal diffusion problem using an optimization-based meshfree quadrature rule that does not require a mesh and naturally handles the nonlocal integral operators.



(a) Convergence with $\delta, h \to 0$ in the physical space. (b) Convergence with increasing sample numbers.

Figure 2: Asymptotic compatibility study on a 2D spatial domain with 2D parametric space. (a) Convergence with 100 Monte Carlo samples. (b) Sparse grid convergence for Smolyak levels $\ell = 1, \ldots, 6$.

A key theoretical contribution is proving that our numerical scheme is asymptotically compatible: as the nonlocal horizon parameter approaches zero, the discrete solution converges to the solution of the corresponding local PDE. We also establish rigorous convergence rates in the random parameter space, showing that the sparse grid collocation achieves algebraic or sub-exponential convergence as the number of collocation points increases. Figure 2 confirms these theoretical predictions on benchmark problems.

The sparse grid approach provides substantial computational savings compared to standard Monte Carlo simulation. For problems with moderate parametric dimension, the PCM method achieves com-

parable accuracy with orders of magnitude fewer samples. This efficiency gain enables uncertainty quantification for nonlocal models in applications where Monte Carlo would be impractical, including materials science, subsurface flow, and biological transport phenomena.

5. Gaussian smoothing methods for gradient-based optimization

Gradient-based optimization is fundamental to machine learning and scientific computing, yet standard methods often struggle with non-convex landscapes, noise, and high dimensionality. In this project, we developed a family of Gaussian smoothing techniques that improve gradient-based optimization by incorporating nonlocal information about the loss landscape, leading to more robust convergence and better solutions.

5.1. Gaussian smoothing gradient descent for function minimization

In [2], we introduced Gaussian Smoothing Gradient Descent (GSmoothGD), a method that replaces standard gradients with nonlocal gradients derived from Gaussian-smoothed versions of the objective function. The smoothing operation averages function values over a neighborhood, which reduces high-frequency noise and small local variations while preserving the overall landscape structure. This enables the optimizer to escape shallow local minima and navigate toward better solutions.

Computing the smoothed gradient exactly requires high-dimensional integration, which becomes intractable as dimension increases. We address this through Monte Carlo approximation (MC-GSmoothGD), sampling random directions to estimate the smoothed gradient efficiently. A key theoretical contribution is proving that MC-GSmoothGD converges regardless of the function's smoothness or the problem dimension, making it applicable to a broad class of optimization problems.

We also developed adaptive strategies for adjusting the smoothing radius during optimization. Large radii provide global exploration early in optimization, while smaller radii enable fine-grained local refinement as the algorithm approaches a solution. Numerical experiments on standard non-convex benchmarks demonstrate that GSmoothGD outperforms momentum-based methods, classical gradient descent, and other smoothing approaches.

5.2. Improved stochastic gradients with Gaussian smoothing for neural networks

Training deep neural networks involves optimizing highly non-convex loss functions with stochastic gradients computed from mini-batches of data. In [3], we extended Gaussian smoothing to stochastic optimization, developing GSmoothSGD and GSmoothAdam variants of standard optimizers.

A key contribution is deriving analytically efficient formulations for computing Gaussian-smoothed loss functions in feedforward and convolutional neural networks. Rather than relying on expensive zero-order approximations that require many function evaluations, our approach integrates directly with automatic differentiation frameworks. This makes the smoothed optimizers practical for large-scale deep learning applications.

Theoretical analysis shows that Gaussian smoothing simplifies the loss landscape by attenuating small fluctuations that can trap standard optimizers in suboptimal local minima. The smoothed variants also exhibit enhanced robustness to noise in gradient estimates and improved generalization performance. Experiments on image classification tasks demonstrate consistent improvements over unsmoothed counterparts, with the computational overhead remaining modest due to our efficient analytical formulations.

5.3. Anisotropic Gaussian smoothing for gradient-based optimization

Standard isotropic Gaussian smoothing applies uniform smoothing in all directions, but this may not be optimal when the loss landscape has different characteristics along different directions. In [4], we introduced anisotropic Gaussian smoothing that adapts the smoothing directionality to match the local geometry of the objective function.

We developed a family of algorithms including AGS-GD, AGS-SGD, and AGS-Adam that replace standard gradients with nonlocal gradients derived from anisotropic Gaussian smoothing. The anisotropy is controlled by the covariance matrix of the Gaussian distribution, which can be adjusted to provide stronger smoothing along directions with high curvature and weaker smoothing along directions that are already well-behaved. This directional adaptation helps algorithms escape local minima more effectively while maintaining fast convergence along favorable directions.

We provide convergence analyses for both convex and non-convex L-smooth functions, extending theoretical results from the isotropic case. In stochastic settings, we show that the algorithms converge to a neighborhood of the optimum, with the neighborhood size determined by the smoothing parameters. The paper includes practical implementation guidance using Monte Carlo estimation techniques aligned with zero-order optimization methods, making the approach accessible for practitioners working on challenging optimization problems in machine learning and scientific computing.

References

[1] Y. Fan, X. Tian, X. Yang, X. Li, C. G. Webster, and Y. Yu. An asymptotically compatible probabilistic collocation method for randomly heterogeneous nonlocal problems. *Journal of Computational Physics*, 465:111376, 2022.

[2] A. Starnes, A. Dereventsov, and C. G. Webster. Gaussian smoothing gradient descent for minimizing functions (GSmoothGD). *arXiv*, 2023.

[3] A. Starnes and C. G. Webster. Improved performance of stochastic gradients with Gaussian smoothing. *arXiv*, 2023.

[4] A. Starnes, G. Zhang, V. Reshniak, and C. G. Webster. Anisotropic Gaussian Smoothing for Gradient-based Optimization. *arXiv*, 2024.

[5] X. Xie, F. Bao, T. Maier, and C. G. Webster. Analytic continuation of noisy data using Adams Bashforth residual neural network. *Discrete & Continuous Dynamical Systems - S*, 15(4):877–892, 2022.

[6] X. Xie, C. G. Webster, and T. Iliescu. Closure Learning for Nonlinear Model Reduction Using Deep Residual Neural Network. *Fluids*, 5(1,39):1–15, 2020.

[7] X. Xie, G. Zhang, and C. G. Webster. Non-intrusive inference reduced order model for fluids using deep multistep neural network. *Mathematics*, 7(8):1–15, 2019.

A. Professional activities

A.1. Invited presentations

1. 2024: Keynote lecture on *Advancing personalized recommendations with reinforcement learning techniques*, IEEE International Conference on Data Mining, 4th Workshop on AI for Nudging and Personalization (WAIN), Abu Dhabi, United Arab Emirates.
2. 2024: 2024 Conference on the Mathematical Theory of Deep Neural Networks, Philadelphia, PA.
3. 2024: 2024 SIAM Conference on Mathematics of Data Science, Atlanta, GA.
4. 2024: 2024 SIAM Conference on Uncertainty Quantification, Trieste, Italy.
5. 2024: 18th Copper Mountain Conference on Iterative Methods, Copper Mountain, CO.
6. 2024: 4th International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI), Jeju Island, South Korea.
7. 2024: Department of Mathematics, Middle Tennessee State University, Murfreesboro, TN.
8. 2024: Department of Mathematics, North Carolina State University, Raleigh, NC.
9. 2024: Department of Mathematics, University of Nevada, Las Vegas, NV.
10. 2024: Fields Institute, University of Toronto, Toronto, ON.
11. 2023: Keynote lecture on *Hyperpersonalization in healthcare*, IEEE International Conference on Data Mining, 3rd Workshop on AI for Nudging and Personalization (WAIN), Shanghai, China.
12. 2023: 10th International Congress on Industrial and Applied Mathematics (ICIAM), Tokyo, Japan.
13. 2023: Department of Scientific Computing, Florida State University, Tallahassee, FL.
14. 2023: Department of Mathematics, Missouri S&T University, Rolla, MO.
15. 2022: Keynote lecture on *Improving entropy of reinforcement learning approaches for personalized recommendation tasks*, IEEE International Conference on Data Mining, 2nd Workshop on AI for Nudging and Personalization (WAIN), Orlando, FL.
16. 2022: Adaptivity, High Dimensionality and Randomness, The Erwin Schrödinger International Institute for Mathematics and Physics (ESI), University of Vienna, Vienna, Austria.
17. 2022: Approximation of high-dimensional parametric PDEs in forward UQ, The Erwin Schrödinger International Institute for Mathematics and Physics (ESI), University of Vienna, Vienna, Austria.
18. 2022: Applied Mathematics Seminar, Argonne National Laboratory, Lemont, IL.
19. 2021: Babuska Forum Lecture on *Sparsity-enforced regularizations for optimal learning of high-dimensional systems from random sampling*, Oden Institute for Engineering and Computational Sciences, The University of Texas at Austin, Austin, TX.
20. 2021: Department of Mathematics, International School for Advanced Studies, Trieste, Italy.
21. 2021: Department of Mathematics, University of Pittsburgh, Pittsburgh, PA.

22. 2021: Oden Institute for Computational Science & Engineering, University of Texas at Austin, Austin, TX.

23. 2021: Department of Mathematics, Texas A&M University, College Station, TX.

24. 2021: Department of Mathematics, Virginia Polytechnic and State University, Blacksburg, VA.

25. 2021: Department of Mathematics, University of Nevada, Las Vegas, NV.

26. 2021: RAMSES: Reduced order models; Approximation theory; Machine learning; Surrogates, Emulators and Simulators, SISSA, International School for Advanced Studies, Trieste, Italy.

27. 2021: Safety and Security of Deep Learning, Institute for Computational and Experimental Research in Mathematics (ICERM), Brown University, Providence, RI.

28. 2021: 2021 SIAM Conference on Computational Science and Engineering, Virtual.

29. 2021: Workshop on Scientific Computing and Applications, University of Nevada, Las Vegas, NV.

30. 2021: SIAM/CAIMS Annual Meeting, Toronto, Canada.

A.2. Workshops and conferences co-organized

1. 2023: 10th International Congress on Industrial and Applied Mathematics (ICIAM), Minisymposium on *Black box methods for efficient learning in high-dimensional scientific computing* (3 parts, 12 talks), Tokyo, Japan. Co-organized with Nick Dexter (FSU) and Guannan Zhang (ORNL).

2. 2021: Safety and Security of Deep Learning, Institute for Computational and Experimental Research in Mathematics (ICERM), Brown University, Providence, RI. Workshop website: https://icerm.brown.edu/events/htw-21-ssdl/

3. 2021: 2021 SIAM Conference on Computational Science and Engineering, Minisymposium on *In quest of predictable and robust machine learning: theoretical and applied perspectives* (2 parts, 8 talks), Virtual. Co-organized with Viktor Reshniak (ORNL).

A.3. Other notable professional activities

- EDITOR-IN-CHIEF
  - 2020–present: Numerical Methods for Partial Differential Equations
- EDITORIAL BOARDS
  - 2018–present: Numerische Mathematik
  - 2018–present: Results in Applied Mathematics
  - 2018–2021: SIAM Mathematics in Industry (Book Series)
  - 2013–present: International Journal of Computer Mathematics (IJCM)
  - 2011–present: International Journal for Uncertainty Quantification (IJ4UQ)