AFRL-AFOSR-VA-TR-2020-0086

Sparse Modeling and Machine Learning for Nonlinear Partial Differential Equations

Hayden Schaeffer
CARNEGIE MELLON UNIVERSITY

05/28/2020
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
06-07-2020

2. REPORT TYPE
Final Performance

3. DATES COVERED *(From - To)*
15 Apr 2017 to 14 Apr 2020

4. TITLE AND SUBTITLE
Sparse Modeling and Machine Learning for Nonlinear Partial Differential Equations

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-17-1-0125

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Hayden Schaeffer

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CARNEGIE MELLON UNIVERSITY
5000 FORBES AVENUE
PITTSBURGH, PA 15213-3815 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2020-0086

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The main goal of the grant was to construct new approaches and algorithms for learning dynamical systems from data. The objective of this research was to develop and analyze new approaches for discovering the underlying governing equations that model a given dataset. We assumed that the data was generated by some unknown dynamic process, typically satisfying a time-dependent differential equation. The technical strategies were based on sparse optimization (with limited sampling) and structured networks. This work involved sparsity-promoting optimization based on the l1 penalty, which was used to regularize the recovery process. The results include several new algorithmic and theoretical developments.

15. SUBJECT TERMS
inference, compressed sensing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			FAHROO, FARIBA
Unclassified	Unclassified	Unclassified	UU		19b. TELEPHONE NUMBER *(Include area code)* 703-696-8429

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Final Report, April 2020
Project Title: Sparse Modeling and Machine Learning for Nonlinear Partial Differential Equations
Grant Number: FA9550-17-1-0125
Program Officer: Dr. Fariba Fahroo, (Formerly: Dr. Jean-Luc Cambier), AFOSR Computational Mathematics
Principal Investigator: Hayden Schaeffer, Carnegie Mellon University

1 Project Descriptions and Goals

This is the final report for the Young Investigator Program Grant, Number: FA9550-17-1-0125 managed by Dr. Fariba Fahroo (formerly managed by Dr. Jean-Luc Cambier), AFOSR Computational Mathematics.

The main goal of the grant was to construct new approaches and algorithms for learning dynamical systems from data. The objective of this research was to develop and analyze new approaches for discovering the underlying governing equations that model a given dataset. We assumed that the data was generated by some unknown dynamic process, typically satisfying a time-dependent differential equation. The technical strategies were based on sparse optimization (with limited sampling) and structured networks. This work involved sparsity-promoting optimization based on the ℓ^1 penalty, which was used to regularize the recovery process. The results include several new algorithmic and theoretical developments.

2 Significant Findings and Conclusions

We summarize the significant findings, which can also be found in the papers supported by this grant [2–12].

Some highlights of the results produced by this grant include:

- introducing a random sampling strategy for recovering high-dimensional dynamical systems from data based on our random burst framework,
- providing sparse recovery results for dependent (time-series) data,
- developing a group-sparse recovery model for learning dynamical systems sampled from different sources and unknown bifurcation regimes,
- providing convergence guarantees for the SINDy algorithm,
- constructing methods for learning sparse dynamical systems for noisy state samples,
- developing new neural network architectures and theory for computational physics problems.

Preprints, publications, and codes are available online, and links to access them are provided in this final report.

2.1 Overview

The main computational and algorithmic goals in this grant were to develop methods for learning dynamical systems from trajectories (time-series samples). We have constructed ℓ^1 optimization approaches that successful solve various learning problems [2, 4–8]. Additionally, we have provided convergence theory for alternative approaches [12]. We developed new deep neural network (DNN) and machine learning algorithms for uncovering unknown dynamical systems from data [9, 11].

Applications related to the goals of the AFOSR include: making data-based predictions, learning nonlinear reduced order dynamics, and computing surrogate models for complex systems.

Technical Summary: Before summarizing our findings, we provide an overview of the mathematical problem. Given time-series data, $x(t) \in \mathbb{R}^n$ (likely in a high dimension $n \gg 1$), we fit the data to a dynamical system, $\dot{x} = f(x)$ with $x(t_0) = x_0$. The goal is to learn the unknown function $f = (f_1, \ldots, f_n)$, from samples of the trajectory $x(t)$. This problem is difficult since, in high dimensions, the data is often under-determined and thus one does not have enough samples to determine the function f using standard approaches. We used sparse approximations and neural networks to construct tractable methods with various approximation guarantees.

2.2 Extracting Dynamical Systems from High-Dimensional Data using Randomness

In [6–8], we used the sparse structure of the governing equations (the fact that the dynamics only depend on a limited number of nonlinear interactions) along with recent results from random sampling theory to develop methods for extracting dynamical systems from under-sampled data.

In [7], we proposed three random sampling strategies that led to the **exact recovery** of first-order dynamical systems even when we are given fewer samples than unknowns. The strategies balance between information on the system and control over the sampling process. The first strategy made no assumptions on the behavior of the data, and required a certain number of random initial samples. The second strategy utilized the structure of the governing equation to limit the number of random initializations needed. The third strategy leveraged any intrinsic randomness or chaos in the data to construct a nearly deterministic sampling strategy. Each are viable in different applications.

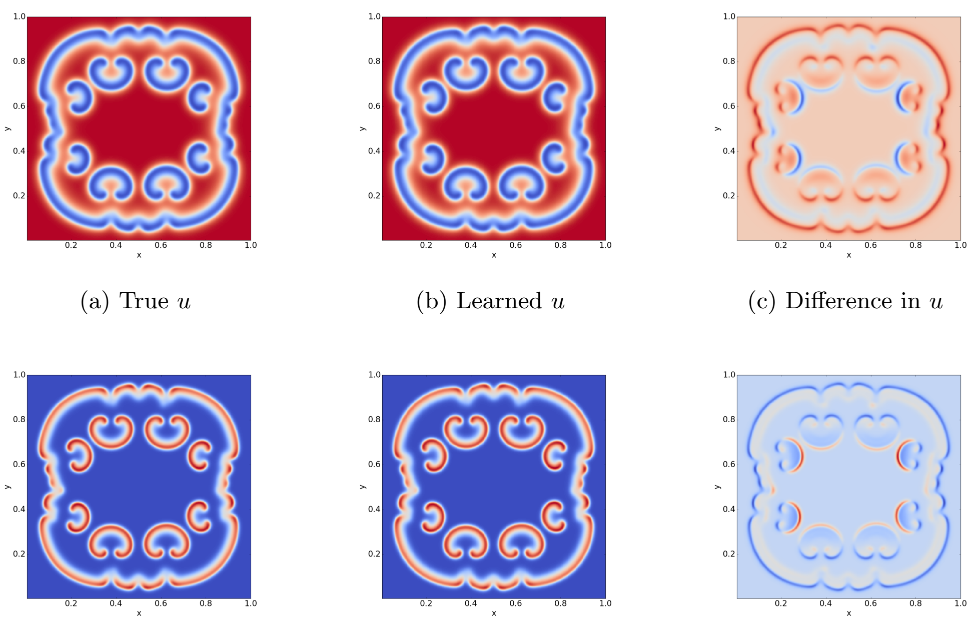
As the dimension increase, the number of candidate functions grows rapidly. Using prior information on locality or connectivity between coordinates, we developed tractable algorithms for very high dimensional structured systems [8]. Computational results show that we can reconstruct the underlying PDE from measurements of the state variable. Additionally, we have theoretical guarantees of success beyond the standard compressive sensing results.

2.3 Sparse Approximation from Data with Outliers

In [2], we constructed a sparse learning approach for approximating non-linear systems from noisy, limited, and/or dependent data. When the data is acquired from a time-series, the measurements will often have non-trivial dependencies. In addition, the measurements may have errors which can bias the learning process (e.g. outliers or noise). This is an important issue to deal with, since it is related to modern applications. The main contribution is the development of a sparse-outlier optimization problem that can recover the parameters of the model and remove outliers. We also provided guarantees of successful recovery via the optimization problem, where the sampling matrix is formed from dependent data.

2.4 Dictionary Matrix and Noise

Another approach to deal with moderate levels of noise is to reformulate the dynamical system as an integral equation [4]. In particular, instead of learning an approximation for $\dot{x} = f(x)$ with $x(t_0) = x_0$, we learned an approximation to $x(t) - x_0 = \int_{t_0}^{t} f(x(s))\mathrm{d}s$ with initial data $x(t_0) = x_0$. For moderate noise levels, the algorithm produced an approximation to the underlying model as well as a smooth approximation to the noisy dynamics. Computational experiments on simulated data showed that this approach is stable with respect to data-size, robust to noise, and accurate when dynamic behavior is included in the dataset. Figure 2 shows an example of the method applied to data generated by the Rössler System. This system contains fast and slow dynamics, which we corrupt with additive noise. We restrict the variables to the slow ones, in this case the first three components. In this example, we are able to recover the slow (dominate) dynamics (the curve in the right figure).



(a) True u (b) Learned u (c) Difference in u

(d) True v (e) Learned v (f) Difference in v

Figure 1: From [8]: Learning a two variable reaction-diffusion system. (a)(d) The true evolution of both variables at $T = 1000$. (b)(e) The learned system at $T = 1000$ using our model. (c)(f) The difference between the true evolution and the learned evolution.

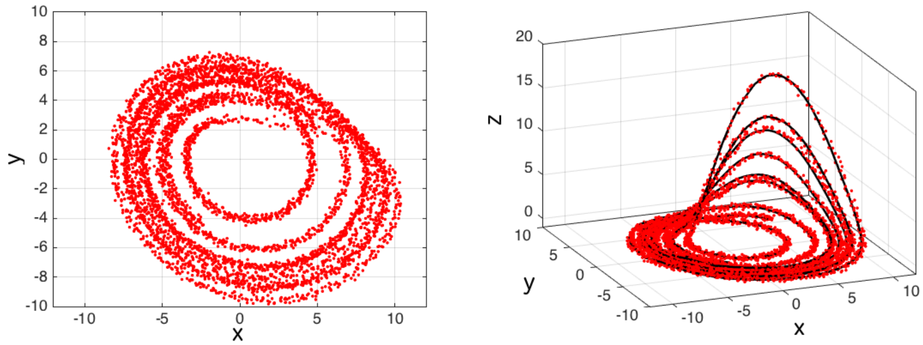
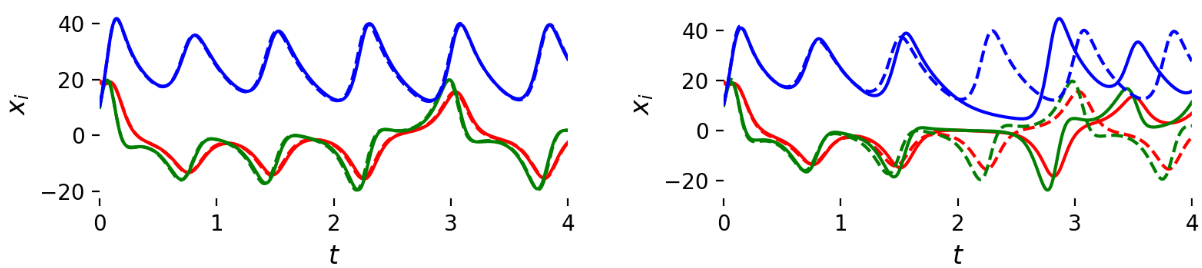


Figure 2: Fast-Slow Approximation The left figure plots the noisy data projected onto the xy-plane with noise level 3.5%. The right figure plots the noisy data along with the data-driven approximation from [4].

2.5 New Deep Neural Network Architectures

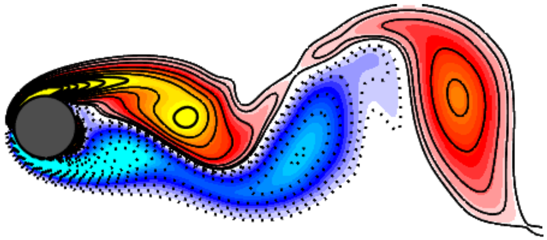
As an alternative to stationary dictionaries, we developed a new DNN architecture for learning dynamical systems [9]. The network is learned via an optimal control problem, with constraints $\dot{x} = f(x)$, where f is parameterized by shallow multilayer perceptrons with nonlinear differential terms. This new architecture incorporated relevant correlations between spatio-temporal samples and thus, experimentally, needed fewer parameters to get an accurate approximation. We demonstrated our approach on dynamical systems, reduced order models for fluids, and conservation laws. In addition, we showed that this approach can be used to lower the parameter cost for deep networks used in image classification.



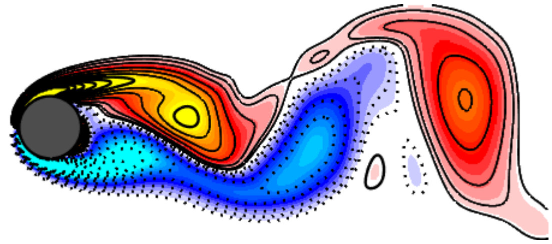
(a) Our Approach with 263 parameters.

(b) Previous DNN approach with 269 parameters.

Figure 3: Time-series data generated by Lorenz system and the corresponding learned processes using our DNN approach. The original data (dashed) and the learned series (solid) are plotted, where RGB corresponds to x_1, x_2, and x_3 respectively.



(a) Our reduced order model using a new DNN architecture with 8 modes



(b) DMD method with 8 modes

Figure 4: Comparison between our DNN approach vs. standard reduced order models..

2.6 Related Theoretical Advances

We also investigated convergence of various algorithms related to the goals of this grant.

In [12], we proved that the sparse identification of nonlinear dynamics framework proposed in [1] approximates local minimizers of an unconstrained ℓ_0-penalized least-squares problem. We also provided sufficient conditions for general convergence, rate of convergence, and conditions for one-step recovery. This showed that the algorithm enforces sparsity via an ℓ_0 penalty and provided theoretical verification to several observed phenomena.

The most popular algorithm in optimization is the gradient descent method. For problems in machine learning, the objective functions are often non-convex and thus the energy landscape can include saddles and local maxima. In [5], we provided larger step-size restrictions for which gradient descent based algorithms (almost surely) avoid strict saddle points. We proved that given one uniformly random initialization, the probability that gradient descent with a step-size up to $2/L$ will converge to a strict saddle point is zero (where L is the Lipschitz constant of the gradient). We showed that the assumptions are robust in the sense that functions which do not satisfy the assumptions are meager with respect to analytic functions.

3 Codes

Our codes for sparse learning of dynamical systems, high-dimensional functions, and bifurcation regimes can be found on:

• Extracting structured dynamical systems using sparse optimization with very few samples:
`https://github.com/linanzhang/SparseCyclicRecovery`

• On the Convergence of the SINDy Algorithm:
`https://github.com/linanzhang/SINDyConvergenceExamples`

• Extracting Sparse High-Dimensional Dynamics from Limited Data:
`https://github.com/GiangTTran/ExtractingSparseHighDimensionalDynamicsFromLimitedData`

• Learning Dynamical Systems and Bifurcation via Group Sparsity:

https://github.com/GiangTTran/GroupHardIterativeThresholdingAlgorithm

References

[1] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

[2] Lam Si Tung Ho, Hayden Schaeffer, Giang Tran, and Rachel Ward. Recovery guarantees for polynomial approximation from dependent data with outliers. *arXiv preprint arXiv:1811.10115*, 2018.

[3] Hayden Schaeffer. A penalty method for some nonlinear variational obstacle problems. *Communications in Mathematical Sciences*, 16(7):1757–1777, 2018.

[4] Hayden Schaeffer and Scott G McCalla. Sparse model selection via integral terms. *Physical Review E*, 96(2):023302, 2017.

[5] Hayden Schaeffer and Scott G McCalla. Extending the step-size restriction for gradient descent to avoid strict saddle points. *arXiv preprint arXiv:1908.01753*, 2019.

[6] Hayden Schaeffer, Giang Tran, and Rachel Ward. Learning dynamical systems and bifurcation via group sparsity. *arXiv preprint arXiv:1709.01558*, 2017.

[7] Hayden Schaeffer, Giang Tran, and Rachel Ward. Extracting sparse high-dimensional dynamics from limited data. *SIAM Journal on Applied Mathematics*, 78(6):3279–3295, 2018.

[8] Hayden Schaeffer, Giang Tran, Rachel Ward, and Linan Zhang. Extracting structured dynamical systems using sparse optimization with very few samples. *arXiv preprint arXiv:1805.04158*, 2018.

[9] Yifan Sun, Linan Zhang, and Hayden Schaeffer. Neupde: Neural network based ordinary and partial differential equations for modeling time-dependent data. *arXiv preprint arXiv:1908.03190*, 2019.

[10] Linan Zhang and Hayden Schaeffer. Stability and error estimates of bv solutions to the abel inverse problem. *Inverse Problems*, 34(10):105003, 2018.

[11] Linan Zhang and Hayden Schaeffer. Forward stability of resnet and its variants. *Journal of Mathematical Imaging and Vision*, pages 1–24, 2019.

[12] Linan Zhang and Hayden Schaeffer. On the convergence of the sindy algorithm. *Multiscale Modeling & Simulation*, 17(3):948–972, 2019.