DATA-DRIVEN DISCOVERY OF GOVERNING PHYSICAL LAWS AND THEIR PARAMETRIC DEPENDENCIES IN ENGINEERING, PHYSICS AND BIOLOGY

J. Nathan Kutz[†]*, Samuel H. Rudy*[†]*, Alessandro Alla*[*] *and Steven L. Brunton*[**]

[†] Department of Applied Mathematics, University of Washington, Seattle, WA 98195-3925
[*] Department of Scientific Computing, Florida State University, Tallahasee, FL 32306
[**]Department of Mechanical Engineering, University of Washington, Seattle, WA 98195

ABSTRACT

We propose a regression method based upon group sparsity that is capable of discovering parametrized governing dynamical equations of motion of a given system by time series measurements. The method balances model complexity and regression accuracy by selecting a parsimonious model via Pareto analysis. This gives a promising new technique for disambiguating governing equations from simple parametric dependencies in physical, biological and engineering systems.

***Index Terms*—** data-driven discovery, dynamical systems, sparse regression, parametric systems

1. INTRODUCTION

The derivation of governing equations and physical laws has dominated scientific discovery for centuries. In a precursor to the scientific revolution, Galileo first wrote down equations of motion from empirical observations which would later be found to be solutions to Newton's celebrated law of motion $\mathbf{F} = m\mathbf{a}$. Modern sensor technologies and computational power are revolutionizing data-driven discovery by leveraging the core underpinnings of machine learning: optimization and regression. Remarkably, we show that the level of mathematical sophistication need not rise above solving large overdetermined linear systems $\mathbf{Ax} = \mathbf{b}$. For such systems, critical concepts from model selection, pareto analysis (parsimony) and sparse regression allow us to discover governing dynamics from time series measurements alone, thus providing an enabling framework for the derivation of governing equations and laws in, for instance, biological systems where first principals may not be possible.

The rise of data science has relied on three principle components: abundant (big) data and storage, high-performance computing, and statistical methods. In combination, critical low-rank features of the data can be extracted, clustered, classified and often interpreted. This forms the basis of machine learning architectures and predictive analytics. For the discovery of governing equations, we measure a physical or biological system in time. The time series measurements are a manifestation of the underlying nonlinear interactions and processes that produce the empirical observations. By building a large library of potential functional forms of the dynamics, we can use sparsity promoting optimization methods for solving a resulting $\mathbf{Ax} = \mathbf{b}$ and selecting the best terms for representing the dynamics of the time series. As we demonstrate on examples from the physical and biological sciences, this provides a robust mathematical framework for discovering the underlying governing equations of a measured system, thus providing a computational framework for adaptively processing signals from sensor arrays.

2. RELATION TO PRIOR WORK

The data-driven discovery of dynamical systems has emerged as an important tool for characterizing physical and biological systems where first principles derivations are not feasible. As early as the late 1980s attempts were made to provide a principled approach [1]. More recent contributions leveraging symbolic regression and an evolutionary algorithm [2, 3], along with automated inference techniques [4, 5] and spare regression [6, 7, 8] were capable of directly determining nonlinear dynamical system from data. Both the evolutionary [3] and sparse [6] symbolic regression methods avoid overfitting by selecting parsimonious models that balance model accuracy with complexity via Pareto analysis [7]. More recently, a number of other sparse regression schemes have been introduced for model recovery with innovations particularly well suited for handling corrupted data [9], network inference [10] and regressions which do not use a pre-selected set of basis functions [11]. One can also frame the model discovery problem in a probabilistic framework where Gaussian processes are used for probabilistic inference over functions [12, 13]. Overall, these methods are advancing our ability to use computations and data for discovery governing physical laws.

S. Rudy, S. L. Brunton and J. N. Kutz acknowledge support from the Defense Advanced Research Projects Agency (DARPA HR0011-16-C-0016). J. N. Kutz acknowledges support from the Air Force Office of Scientific Research (AFOSR) grant FA9550-15-1-0385. A. Alla is supported by US Department of Energy grant number DE-SC0009324.

978-1-5386-1251-4/17/$31.00 ©2017 IEEE

Many dynamical models, however, exhibit critical parametric dependencies. Thus not only must the model be identified, but also its time-dependent parameter(s) evolution. Discovering the time-evolution of parameters separately from the underlying nonlinear dynamics itself is extremely challenging, and beyond the scope of the current methods developed thus far. In this work, we develop an optimization procedure for parametric model discovery. Specifically, we show that using a *group* regression architecture allows us to discover the underlying dynamical system along with the time-dependent evolution of any parameters. Our algorithm, termed *Sequential Grouped Threshold Ridge regression (SGTR)* is demonstrated to discover parametrized dynamics and be a robust alternative to such techniques as *group LASSO* (GLASSO) [14].

3. DATA-DRIVEN DISCOVERY OF DYNAMICS

The formulation of the data-driven discovery architecture will result in a large, overdetermined linear system of equations. In what follows, we demonstrated how this linear system is constructed and outline its solution via sparse regression.

3.1. Problem formulation and model selection

Consider the system of nonlinear differential equations

$$\dot{\mathbf{u}} = f(\mathbf{u}, \mu(t)) \tag{1}$$

where the dot denotes time-differentiation, $\mathbf{u} \in \mathbb{R}^n$ and $\mu(t)$ represents the explicit parametric dependence of the governing system. We assume that we have time-series measurements of the state space $\mathbf{x}(t)$, but do not know the dynamics $f(\cdot)$ or its parametric dependency μ. Thus both these quantities need to be discovered from the time-series measurements. One fundamental assumption is made in our data-driven architecture: there are only a few terms that comprise the function $f(\mathbf{u}, \mu(t))$. Across the physical, engineering and biological sciences, this assumption holds true in modeling such diverse phenomenon as quantum mechanics and fluid flow. Identifying the few terms, however, is a challenging task.

The method used here constructs libraries of potential right hand side functions so that

$$\dot{\mathbf{X}} = [\mathbf{g}_1(\mathbf{u})\ \mathbf{g}_2(\mathbf{u})\ \mathbf{g}_3(\mathbf{u})\ \ldots\ \mathbf{g}_M(\mathbf{u})] = \boldsymbol{\Theta}\boldsymbol{\Xi} \tag{2}$$

where $\mathbf{X} \in \mathbb{R}^{m\times n}$ so that each column of $\mathbf{X}$ is a component of the state variable $\mathbf{u}$ while each row denotes one of the m time measurements. The column vectors $\mathbf{g}_j(\mathbf{u})$ represent different potential right hand side functions which are evaluated for each time measurement. In total, the M functions $\mathbf{g}_j(\mathbf{u})$ comprise the matrix of library functions $\boldsymbol{\Theta}$. Finally, the matrix $\boldsymbol{\Xi} \in \mathbb{R}^{M\times n}$ determines the weighting of each column of $\boldsymbol{\Theta}$ on the matrix $\dot{\mathbf{X}}$. Each column (2) results in a linear system of equations with the matrix $\mathbf{A} \in \mathbb{R}^{m\times M}$. Importantly, the assumption that only a few terms comprise $f(\mathbf{u}, \mu(t))$ results in a sparse solution matrix $\boldsymbol{\Xi}$.

3.2. Sparse regression

Solving $\mathbf{A}\mathbf{x} = \mathbf{b}$ can easily and efficiently be done using least-squares regression using, for instance, the Morse-Penrose pseudo-inverse. However, the solution vector for the overdetermined systems we consider will be dense. In practice, this means that the solution $\boldsymbol{\Xi}$ in (2) will weight all library elements given by the $\mathbf{g}_j(\mathbf{u})$. So the discovered dynamics will have no parsimony. Indeed, in such a case, the nonlinear dynamics would be a linear combination of thousands of terms that comprise the library.

Sparse regression solves the large overdetermined system by promoting parsimony. In the most ideal case, one would solve the optimization problem

$$\hat{\boldsymbol{\Xi}} = \arg\min_{\boldsymbol{\Xi}} \|\boldsymbol{\Theta}\boldsymbol{\Xi} - \dot{\mathbf{X}}\|_2^2 + \lambda\|\Xi\|_0 \tag{3}$$

where the solution $\hat{\boldsymbol{\Xi}}$ is constructed with an ℓ_0 penalty $\lambda\|\Xi\|_0$. This would lead to a desired solution which has a minimal (parsimonious) number of terms that are nonzero. Unfortunately, the ℓ_0 penalty leads to a combinatorially intractable np-hard optimization procedure. In the last decade, a number of theoretical papers have shown that the ℓ_1 norm can serve as a proxy for sparsity. Moreover, many optimization problems with an ℓ_1 penalization can be formulated as a convex optimization problem so that one is led to consider () with the ℓ_0 penalty replaced by $\lambda\|\Xi\|_1$. There are a variety of algorithms that solve this ℓ_1 penalized equation, or a closely related variant, including a number of greedy algorithms, LASSO, elastic net, hard thresholding etc.

A successful alternative to ℓ_0 or ℓ_1 optimization is the use of a sequential least-squares thresholding procedure [6, 8]. Such a strategy has more limited rigorous guarantees on convergence [15] when compared to many leading ℓ_1 optimization strategies. However, it works exceptionally well in practice across a wide variety of models and data [6, 8]. In this case, one again considers (3.2) but with the ℓ_2 penalty

$$\hat{\boldsymbol{\Xi}} = \arg\min_{\boldsymbol{\Xi}} \|\boldsymbol{\Theta}\boldsymbol{\Xi} - \dot{\mathbf{X}}\|_2^2 + \lambda\|\Xi\|_2^2. \tag{4}$$

Once a solution is found, only the largest terms are retained, or alternatively a threshold is applied to remove small terms, and the procedure is done again. After a number of rounds of thresholding, the solution converges to a sparse solution matrix $\boldsymbol{\Xi}$ which we show to robustly select the terms associated with the nonlinear dynamics. As with any regression procedure, cross-validating the solution is imperative in order to assess the robustness and efficacy of the results. By design, the sparse solution sought represents a parsimonious solution so that a Pareto analysis can be performed which balances model complexity with prediction error [7]. Such methods are closely related to model selection and the computation of Kullback-Leibler divergence metrics for Akaike or Bayesian information criteria (AIC and BIC respectively). In our algorithm, we use AIC as a cross-validation metric in order to

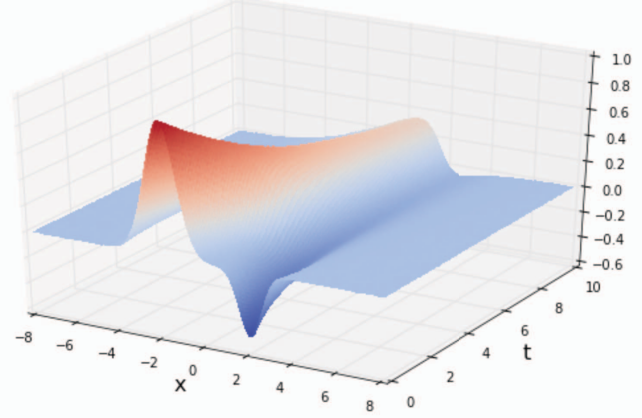

Fig. 1. Dynamics of the parametrized Burgers' equation (7).

assess the selected model as well as those models that are also near the Pareto front. Such a procedure provides a principled method for evaluating the regression procedure.

3.3. Generalizations

The basic framework of (2) with the sequential optimization procedure (4) can be modified to accommodate two important innovations. First, the library functions $\mathbf{g}_j(\mathbf{u})$ can be modified to handle functions of the form

$$\mathbf{g}_j(\mathbf{u}) = \frac{H_j(\mathbf{u})}{G_j(\mathbf{u})} \tag{5}$$

where $H_j(\mathbf{u})$ and $G_j(\mathbf{u})$ are typically simple functional forms. This modification is necessary to model many biological networks that have simple relationships in the numerator and denominator on the right hand side. Without this modification, (5) would have an infinite series expansion and a parsimonious solution could not be achieved. Thus the optimization is modified so that the sparsest set of vectors spanning the null space give the correct solution forms [7].

A second important innovation allows the functions $\mathbf{g}_j(\mathbf{u})$ to also depend on spatial derivatives [8]. This modification to the library gives a framework for the derivation of partial differential equations that govern spatio-temporal phenomenon. Thus the library elements can take the form

$$\mathbf{g}_j(\mathbf{u}) = \mathbf{g}_j(\mathbf{u}, \mathbf{u}_x, \mathbf{u}_{xx}, \cdots) \tag{6}$$

where the subscripts denotes partial differentiation with respect to the spatial variable x. In this generalization, the library functions can now depend upon spatial derivatives. Many important physical and biological systems have underlying diffusion and/or dispersive processes that can thus be captured with library elements that explicitly account for the spatial derivatives, both linear or nonlinear.

Algorithm 1 SGTR($\{\mathbf{\Theta}_i, \dot{\mathbf{X}}_i\}_{i=1}^m, \lambda, \epsilon, \text{maxit}, f(\mathbf{u}) = \|\mathbf{u}\|_2$)

$n, d = \mathbf{\Theta}_i$.shape
$\mathbf{\Xi} = 0 \in \mathbb{R}^{d \times m}$

Train baseline ridge predictor for each timestep
for $iter = 1, \ldots, m$:
$\mathbf{\Xi}[:, i] = arg\,min_{\mathbf{\Xi}} \|\mathbf{\Theta}_i \mathbf{\Xi} - \dot{\mathbf{X}}_i\|_2^2 + \lambda \|\Xi\|_2^2$

Throw out bad variables and repeat
for $iter = 1, \ldots,$ maxit:
    # Find which variables may be thrown out
$bad\_var = \{j : f(\mathbf{\Xi}[j, :]) > \epsilon\}$
$good\_var = \{j : j \notin bad\_var\}$
$\mathbf{\Xi}[bad\_var, :] = 0$
    # Refit to good variables
for $iter = 1, \ldots, m$:
$\mathbf{\Xi}[good\_var, i] = arg\,min_{\mathbf{\Xi}} \|\mathbf{\Theta}_i[:, good\_var]\mathbf{\Xi} - \dot{\mathbf{X}}_i\|_2^2 + \lambda \|\mathbf{\Xi}\|_2^2$

Get unbiased estimates of good variables
for $iter = 1, \ldots, m$:
$\mathbf{\Xi}[good\_var, i] = arg\,min_w \|\mathbf{\Theta}_i[:, good\_var]\mathbf{\Xi} - \dot{\mathbf{X}}\|_2^2$
return $\mathbf{\Xi}$

4. PARAMETRIC MODEL DISCOVERY

The methodology of the last section does not account for any parametric dependency in data. However, as is clear from (1), the underlying governing equations will change as a function of the parameter $\mu(t)$. The parametric dependency can be smooth or abrupt. The model can simply change its parameters, or alternatively, the underlying model itself can abruptly change to a new set of governing equations. Such modifications to (1) are beyond the scope of current methods to handle, i.e. the sparse regression techniques cannot simultaneously discover the governing model and the parametric dependency $\mu(t)$. In this work, we develop a critical innovation for model discovery that discovers both $f(\cdot)$ and $\mu(t)$ simultaneously using a group sparsity procedure.

4.1. Group Sparsity Constraints

Our methodology is based upon innovations using group sparsity constraints. Thus if data is sampled from the same system, then the sparsity pattern of the solution should be the same across the samples. This is the basis of the GLASSO [14] method. Instead of the LASSO regression however, we use a sequential least-square thresholding method [6] for performing the sparse regression. Generically, our SGTR algorithm seeks solutions to the linear systems $\mathbf{\Theta}_j \mathbf{\Xi}_j = \dot{\mathbf{X}}_j$ for each time slice $j = 1, \ldots, m$ where each $\mathbf{\Xi}_j$ is sparse and required to have the same sparsity pattern,

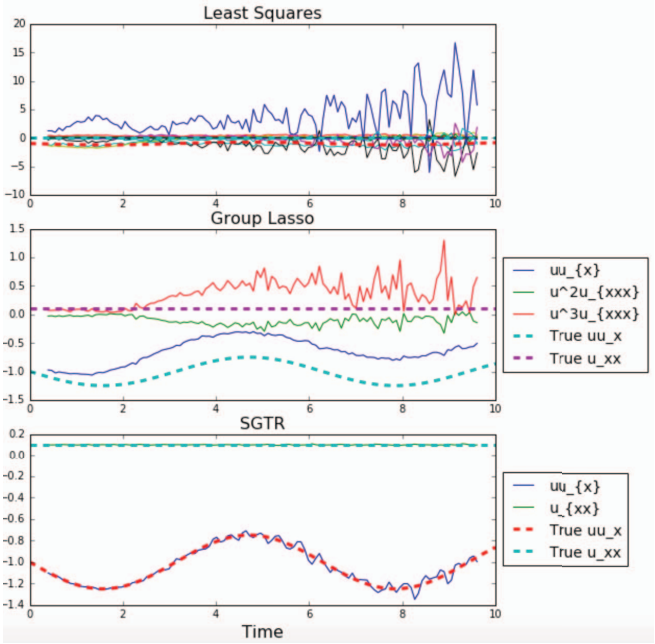

Fig. 2. Comparison of regression methods for solving (2).

with potentially different values, as the other $\boldsymbol{\Xi}_k$ $(k \neq j)$. This may be achieved using GLASSO, but in many cases it requires a good initial guess for the algorithm to converge in a reasonable amount of time. We instead propose SGTR which has significant advantages over GLASSO. It is computationally efficient, easy to write, and can be adjusted to include arbitrarily complex penalties on the coefficient vectors by adjusting the function $f(\mathbf{u})$. Details of the algorithm are shown in Algorithm 1.

4.2. Example dynamical systems

We demonstrate the application of the SGTR, and compare its performance to GLASSO, on a parametrized version of the Burgers' evolution equation

$$u_t + \left(1 + \frac{1}{4}\sin(t)\right) uu_x - Du_{xx} = 0 \qquad (7)$$

where the diffusion parameter $D = 0.1$. For this example, the model selection uses the library structure of (6). The nonlinearity is modulated by a sinusoidal function $\mu(t) = \sin(t)/4$, which makes the data-driven methods for model discovery perform very poorly. However, the SGTR method is robust and efficient and can easily identify both the model $f(\mathbf{u})$ as well as the time-dependent parameter $\mu(t)$. Recall that the temporal evolution form $\mu(t)$, which is sinusoidal for this example, is not part of the library of functions $\boldsymbol{\Theta}$. Only the spatial library elements (atoms) are included in the library. Figure 1 demonstrates the evolution of initial data subject to Burgers' equation. Given the time snapshots of this system, various regression formats can be used to discover the underlying physics. Figure 2 shows a comparison of solutions of (2) using least-squares regression, GLASSO and SGTR when a small amount of measurement noise ($\approx 1\%$) is added to the measurements. Both least-squares and GLASSO fail to capture the correct dynamics, whereas SGTR quickly produces the correct model and parametric dependency.

5. CONCLUSION AND OUTLOOK

We have demonstrated a new sparse regression algorithm that capitalizes on group sparsity in order to discover parametrized dynamical systems from time series measurements. The method disambiguates between the model dynamics and time dependent parameter changes in order to provide a robust and computationally efficient framework for discovering both $f(\cdot)$ and $\mu(t)$ in the governing equations (1). The method provides a important innovation for emerging data driven discovery methods arising across the physical, engineering, and biological sciences. It also overcomes a fundamental limitation of existing methods by allowing for a direct recovery of parametric dependencies observed in most complex systems. Of course, the parametric dependencies in many important physical systems are simple to consider, such as the Reynolds number in the Navier-Stokes equation. However, many more complex and high-dimensional parametrizations may be encountered in practice which would limit the ability to separate the model from its parametric dependency. The work also highlights the utility of iterative thresholding schemes for the recovery of sparse representations, thus suggesting directions for further rigorous mathematical explorations. Indeed, it remains an open question how to guarantee the sparse selection procedure. This will be explored in future work.

6. REFERENCES

[1] James Crutchfield and Bruce McNamara, "Equations of motion from a data series," *Comp. sys.*, vol. 1, pp. 417–452, 1987.

[2] Josh Bongard and Hod Lipson, "Automated reverse engineering of nonlinear dynamical systems," *PNAS*, vol. 104, no. 24, pp. 9943–9948, 2007.

[3] Michael Schmidt and Hod Lipson, "Distilling free-form natural laws from experimental data," *Science*, vol. 324, no. 5923, pp. 81–85, 2009.

[4] Michael Schmidt, Ravishankar Vallabhajosyula, J. Jenkins, Jonathan Hood, Abhishek Soni, John Wikswo, and Hod Lipson, "Automated refinement and inference of analytical models for metabolic networks," *Phys. bio.*, vol. 8, no. 5, pp. 055011, 2011.

[5] Bryan Daniels and Ilya Nemenman, "Automated adaptive inference of phenomenological dynamical models," *Nat. comm.*, vol. 6, 2015.

[6] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *PNAS*, vol. 113, no. 15, pp. 3932–3927, 2016.

[7] Niall M Mangan, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz, "Inferring biological networks by sparse identification of nonlinear dynamics," *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, vol. 2, no. 1, pp. 52–63, 2016.

[8] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz, "Data-driven discovery of partial differential equations," *Science Advances*, vol. 3, no. 4, pp. e1602614, 2017.

[9] Giang Tran and Rachel Ward, "Exact recovery of chaotic systems from highly corrupted data," *SIAM Multiscale Model. Simul.*, vol. 15, pp. 1108–1129, 2017.

[10] S. Chen, A. Shojaie, and D. Witten, "Network reconstruction from high-dimensional ordinary differential equations," *J. Amer. Stat. Assoc. (to appear)*, 2017.

[11] Manuel Mai, Mark D. Shattuck, and Corey S. O'Hern, "Reconstruction of ordinary differential equations from time series data," *arXiv preprint arXiv:1605.05420*, 2016.

[12] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis, "Machine learning of linear differential equations using gaussian processes," *Journal of Computational Physics*, vol. 348, pp. 683–693, 2017.

[13] Maziar Raissi and George Em Karniadakis, "Hidden physics models: Machine learning of nonlinear partial differential equations," *arXiv preprint arXiv:1708.00588*, 2017.

[14] Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani, "A sparse-group lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[15] Thomas Blumensath and Mike E Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629–654, 2008.