Research Statement

Yiping Lu (Stanford University), Email:yplu@stanford.edu

The recent breakthroughs in machine learning (ML) and artificial intelligence have enabled data-driven extraction of information for useful predictions across a wide range of science and engineering disciplines.

My research is motivated by the use of mathematical and domain-specific fundamental principles to enhance both the accuracy and transparency of insights and decisions in science and engineering. I aim to understand the statistical properties and computational efficiencies of domain-specific principles informed data-driven algorithms. I believe that combining domain-specific principles with data-driven learning provides a wide range of benefits which inspire my research philosophy. Here I list a few of them:

- A direct use of machine learning methods is highly data intensity. But this data may not always be available due to privacy considerations, changing environments, etc. However, using domain-specific insights, one may be able to model a meaningful relationship between available observed data/signal and target variables/signals. In these cases, one can use learning techniques to obtain structural model-informed solutions that recover the target signal.
- Blindly applying learning algorithms in real-world decision making problems may lead to trustworthy and safety issues with significant social and economic impact. Often, however, dealing with arbitrary issues of this sort may not be tractable both statistically or computationally. Thus, introducing an economic-based model can help alleviate these potential problems.
- Structural models are transparent for counterfactual prediction. However, structural models typically have approximations that may limit their predictive power. Thus, one can use a hybrid approach. For instance, by applying data-driven non-parametric learning for some features of the structural model, which lead themselves to direct estimation from observations. Hybrid approaches may enjoy the flexibility of machine learning while keeping the structural model transparent and explainable simultaneously.
- Structural models generally involve a fundamental law, often encoded, for instance, by means of a differential equation in high dimensions, which is generally hard to solve. Machine learning provides a way to solve structural models with the potential to break the curse of dimensionality via Monte-Carlo.

Although there have been several recent breakthroughs which suggest that machine learning and AI in general hold significant promise in scientific research [19,20], we still lack fundamental understanding that enables a full synergy between machine learning and structural models in science (including physical and social sciences).

- Machine learning for science and scientific discovery needs to be data efficient, especially when scientific data is expansive to collect. However, a sample complexity theory for scientific machine learning problems is still in its infancy.
- How can we inject the desired physical priors into machine learning model to build more structured and physically grounded systems?
- Scientific machine learning models need to be robust to the environment in which they will

be used for prediction, but they are often vulnerable to perturbations, biased towards spurious relations, and even sensitive to simple domain shift [21].

So, while structural modeling has great potential, its proper use in combination with machine learning also brings new challenges in rigorous analysis and the development of performance guarantees. This provides a whole host of fundamental theoretical problems of great interest, in my opinion:

- First of all, structural models should be identifiable. Although *inverse problems* is an area that has been well studied, especially in the physical sciences, identifiable results are still missing for most social science models.
- *Computational methods* in inverse problems and discovery of solutions to differential questions often neglect practical statistical out-of-sample guarantees. For example, direct sample average approximations to the variational form of certain PDEs will (as my research shows) result in an enlarged variance in the frequency domain (which is relevant in several applications). I'm interested in building a non-parametric statistical framework to study optimal performance guarantees over various relevant criteria informed by applications.
- *Machine Learning Theory/Statistics* typically deals with finite dimensional approximations. However, scientific machine learning models is more naturally posed in an infinite dimensional operator setting in which features such as smoothness of the solution and a continuous domain are much more natural primitive objects to consider. I am interested in studying how these statistical properties can be learned and studied in infinite-dimensional problems.

My research aims to understand these questions and reduce the trial-and-error cost for scientific machine learning. To fully understand these problems, my goal is to apply an interdisciplinary research approach across computational physics, probability and statistics, control theory, signal processing/inverse problem and operations research. Over these years, I have accumulated a rich research experience in all of these areas, which has also equipped me with a broad skill set and vision for my future research. I will discuss my research experience.

Research Experience A significant proportion of my research considers phenomena described by differential equations, including applications in many science and engineering disciplines such as physics, material science, operation research, macroeconomic etc. My previous research tends to investigate data-driven differential equation models from the following three perspectives:

Encoding Physics Information into a Model. My initial research focus on interpreting many popular neural networks as different numerical discretizations of (stochastic) differential equations [1,2]. Based on this perspective, we were able to combine physical information with the deep neural network architecture to boost the performance and transparency at the same time.

- **Learning PDEs from Data** In [3,4], we present an initial attempt to learn evolution PDEs from data. By fully exploiting the relation between the orders of differential operators and the orders of sum rules of filters, we proposed PDE-Net which learns differential operators by learning convolution kernels (filters) and apply neural networks or other machine learning methods to approximate the unknown nonlinear responses. This approach combines the representation power of the deep neural networks and the transparency of the PDE models which lead to better generalization property towards diverse initial conditions.
- **Learning Inverse Problem** In [5], we apply this idea to inverse problems. With the understanding

of the physic behind the task, we additionally learn an terminal time for different noise level which leads to better generalization across different noise level and different noise statistics.

- **Optimize Neural Network via Control Theory** In [6,7], we apply this idea towards understanding the training process of the deep neural networks. The control theory view-point enables us to design 4-5 times faster algorithm for adversarial training in practice [6] and provide the convergence proof for stochastic gradient descent training multi-layer networks in mean-field via exploring the (local) Polyak-Lojasiewicz property [7].

Sample Complexity of Scientific Machine Learning. How large the sample size and how much computational power are needed to reach a prescribed performance level is always the core problem for theoretical computational science research. We answers this question in a series of work [10,11,12].

- **Statistical and Computational Analysis for ML Based PDE Solver** In [10,11], we focus on a prototype elliptic PDE $\mathcal{L}u = f$. We aim to build an estimaotr for u from random observations $\{(x_i, f(x_i) + \eta)\}_{i=1}^n$ of right hand side function f. We establish the information theoretical lower bounds for learning the equation's solution from sampled data and the first matching upper bound for both (modified version of) Deep Ritz Method (DRM) and Physics Informed Neural Network (PINN). We observed that DRM enlarge the variance of sampling a high-frequency single and a modification is needed to achieve optimal rate. In [11], we explain an implicit acceleration of using a Sobolev norm as the objective function for training, inferring that the optimal number of epochs of DRM becomes larger than the number of PINN when both the data size and the hardness of tasks increase in low dimension, although both DRM and PINN can achieve statistical optimality.
- **Statistical Analysis for Operator Learning** In [12], we consider the optimal learning rate for learning a linear operator between two infinite dimensional Hilbert spaces. We provided a novel lower bound to the literature and showed that multi-level machine learning is essential to achieve an optimal learning rate. This example showed a fundamental difference between infinite dimension machine learning and finite dimension one both in sample complexity and algorithmic design.

Building Robust learning Algorithm via Inductive Bias. Overparametraization, *i.e*, having more model parameters than necessary, is the core factor behind the success of modern machine learning. However, overparametraization also enables the model to fit any noisy signal which makes the model extremely vulnerable. My research aims to build robust overparametrized model via understanding the inductive bias.

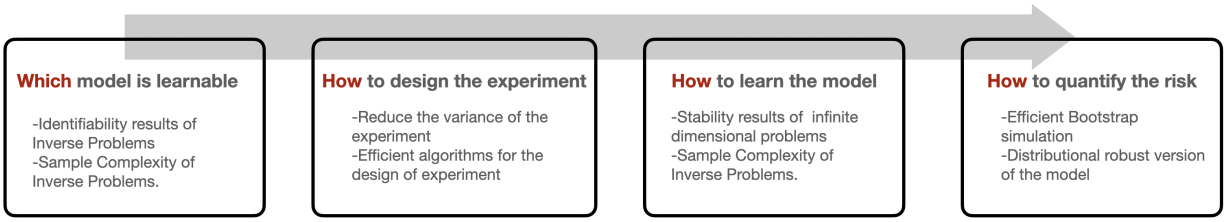
- **Neural Collapse and Imbalanced Classification** In [14], we analyze Neural Collapse, a highly symmetric geometry of neural networks that emerges during the terminal phase of training which leads to better generalization performance, better robustness, and better interpretability. We analyze a surrogate model called the unconstrained layer-peeled model where all the critical points are strict saddle points except the global minimizers that exhibit the neural collapse phenomenon. Based on this theory, we proposed importance tempering, which improve the decision boundary and achieve consistently better results for overparameterized models [15] while traditional importance weighting have less or even no effect on overparameterized models both empirically and theoretically. We also provided theoretical guidance for temperature choosing

under both label shift and spurious correlation setting.

- **Robust Traininig with Noisy Labels** In [8], we discussed the how overparameterized neural network can perform towards corrupted label, which is unavoidable when we collect physics signals in experiments. We who that although overparameterization can leads to memorization of all signals, however the discrepancy of convergence speed of different types of information can help us to distinguish signal from the noise. This leads to a theoretical understanding on how distillation algorithm can help within the situation that the supervision signal is noisy.

Research Plan and Goal

Moving forward, I plan to harness my theoretical skills to continue developing effective solutions with fundamental understandings and theoretical guarantees. I strive to explore the entire pipeline of the scientific machine-learning system. First, we need to understand which model is identifiable and learnable. Then I aim to design the data collection process and model learning algorithms to reduce the variance of the learned model and the computational complexity of the whole process. Finally, I aim to design fast simulation algorithms to quantify the learned model's risk. It's an exciting time to work on the interaction between mathematical modeling, computational methods, and machine learning in this entire pipeline. I list my research plan and goal as follows.



The entire pipeline for scientific machine learning system.

Which Model is Learnable? Defining learning tasks is one of the critical problems for scientific machine learning, while recent research has not paid that much attention to this. For example, what signal should be collected to learn a certain target quantity? Which contains more information with a fixed budget, cleaner data, or as much noisy data as possible contains more information for non-parametric Bayesian filtering? How should the data collecting process and machine learning algorithm be adapted if we aim to learn multi-scale physic? Based on our *first optimal* sample complexity bounds for scientific machine learning, I will continually investigate sample complexity from an information theoretical viewpoint.

How to design the experiment? Correlation is not causal. While machine learning for scientific problems has been widely studied in recent years, causal inference for physics problems as the potential to explore. Causally learned models enjoy stable generalization capabilities of data-driven inference. I plan to study causal inference from the following two perspectives. Firstly, causal inference always implements a randomized experiment that lets one know the effect of specific environmental changes. This provides more information to reconstruct the underlying model. How identifiable results and statistical sample complexity can be improved is interesting. Secondly, stochastic simulation of randomized experiments always has a large variance for high-dimensional

physics. We aim to find computational methods for optimal experiment design. Although finding optimal design is known to be NP-hard, our recent work [16] finds out that experiment design with synthetic control is equivalent to phase synchronization. We numerically solved the problem via the generalized power method and provided *first global optimal guarantee* for experiment design under certain statistical generative models. I plan to extend our approach to specific physics problems.

How to learn our model? Data-driven algorithms always needs a sample average approximation. While we consider the sample average approximation in terms of an optimization framework, stability results of the approximated objective function are needed. However, stability is surprisingly not always satisfied for constrained and non-smooth infinite-dimensional optimization problems. I plan to work on multi-stage stochastic control (fundamental in operation research) [17] and (super-)martingale optimal transport (a key problem for auction theory and robust hedging). Both are examples of important problems where sample average approximation fails to converge to the optimal solution. In this project, we aim to construct smoothed relaxations of these objective functions. Our approach has a deep connection with the *distributional robust optimization* and *outlier detection*. I aim to build a computational feasible and sample efficient algorithm to learn the infinite dimensional non-smooth problem based on our stability resutls.

How to quantify the risk? Quantifying the impact of input uncertainty, environmental uncertainty, and model misspecification needs substantial computation demand, especially when a computational expansive model is included. In [18], I understand this problem under a semi-parametric framework. Combining my expertise in semi-parametric statistics and stochastic simulation, I aim to design efficient simulation algorithms for quantifying the risk of a certain model. I am also working on extending my framework for infinite-dimensional models, where the variance may explode up to infinity.

Significance of Proposed Work

With the rapid development of sensors, computational power, and data storage in the past decade, a huge amount of data can be easily collected and efficiently stored. Such a vast quantity of data offers new opportunities for the data-driven discovery of physical laws. From the experimental design to data analysis, from physics law discovery to make useful predictions, AI can help to make the full process of scientific discovery fully automated. However, the lack of guidance of model and algorithm designs, and robustness of the trained models is a clear threat to this area.
Overcoming the obstacles mentioned before needs interdisciplinary research, varying from numerical analysis, statistics to applied probability and even physics and economics. At the same time, it also raises new challenges to theoretical analysis. I have armed myself broad vision of mathematical background and the skills to model real-world motivations as computational feasible model to investigate through previous research. It lays a solid foundation for my future career path to build theoretical foundation for scientific machine learning and make them more transparent and robust. I'll continue conducting concrete contribution to the analysis and design of consistent, statistically and computationally efficient algorithms for scientific machine learning, where the variance of simulation may explode up to infinity.

Reference

My Contributions

[1] Yiping Lu, AoxiaoZhong, Quanzheng Li, Bin Dong. "Beyond Finite Layer Neural Network:Bridging Deep Architects and Numerical Differential Equations" Thirty-fifth International Conference on Machine Learning (ICML), 2018.
[2] Yiping Lu*, Zhuohan Li*, Di He, Zhiqing Sun, Bin Dong, Tao Qin, Liwei Wang, Tie-yan Liu. Understanding and Improving Transformer Architecture From the View of Multi-particle Dynamic System.(*equal contribution.) arXiv:1906.02762, 2019.
[3] Zichao long*,Yiping Lu*, Xianzhong Ma*, Bin Dong. "PDE-Net:Learning PDEs From Data", Thirty-fifth International Conference on Machine Learning (ICML) 2018 (*equal contribution)
[4] Zichao long, Yiping Lu, Bin Dong. PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. Journal Of Computational Physics 2019.
[5] Xiaoshuai Zhang*,Yiping Lu*, Jiaying Liu, Bin Dong. "Dynamically Unfolding Recurrent Restorer: A Moving Endpoint Control Method for Image Restoration" (*equal contribution, joint first author) International Conference on Learning Representations(ICLR) 2019
[6] Dinghuai Zhang*, Tianyuan Zhang*, Yiping Lu*, Zhanxing Zhu, Bin Dong. "You only propagate once: Accelerating adversarial training using maximal principle." Conference on Neural Information Processing Systems (Neurips) 2019 (*equal contribution, joint first author)
[7] Yiping Lu, Chao Ma, Yulong Lu, Jianfeng Lu, Lexing Ying. A Mean-field Analysis of Deep ResNet and Beyond: Towards Provable Optimization Via Overparameterization From Depth ICML2020.
[8] Bin Dong, Yiping Lu, Jikai Hou, Zhihua Zhang. "Distillation=Early stopping? Extracting knowledge from anisotropic information retrieval" NeurIPS2019 Workshop on ML with Guarantees.
[9] Bin Dong, Haochen Ju, Yiping Lu, Zuoqiang Shi. " CURE: Curvature Regularization For Missing Data Recovery." SIAM Journal on Imaging Sciences, 2020, 13(4): 2169-2188.
[10] Yiping Lu, Haoxuan Chen, Jianfeng Lu, Lexing Ying, Jose Blanchet. Machine Learning For Elliptic PDEs: Fast Rate Generalization Bound, Neural Scaling Law and Minimax Optimality. International Conference on Learning Representations (ICLR) 2022
[11] Yiping Lu, Jose Blanchet,Lexing Ying. Sobolev Acceleration and Statistical Optimality for Learning Elliptic Equations via Gradient Descent, Neurips 2022.
[12] Jikai Jin, Yiping Lu, Jose Blanchet, Lexing Ying Minimax Optimal Kernel Operator Learning via Multilevel Training arXiv : 2209.14430
[13] Huishuai Zhang, Da yu, Yiping Lu, Di He. Adversarial Noises Are Linearly Separable for (Nearly) Random Neural Networks arXiv:2206.04316
[14] Ji W, Lu Y, Zhang Y, et al. An unconstrained layer-peeled perspective on neural collapse. International Conference on Learning Representations(ICLR) 2022.
[15]Yiping Lu, Wenlong Ji, Zach Izzo, Lexing Ying, Importance Tempering: Group Robustness for Overparameterized Models. arxiv: 2209.08745
[16] Yiping Lu, Jiajin Li, Lexing Ying, Jose Blanchet, Synthetic Principle Component Design: Fast Experiment Design with Synthetic Control (Submitted)
[17] Yiping Lu, Jiajin Li, Jose Blanchet, Epi-Convergent Data-Driven Multistage Stochastic Linear Programming via Wasserstein Distributionally Robust Relaxation (In Preparation)

[18] Yiping Lu, Lexing Ying, Jose Blanchet, Orthogonal Bootstrapping: Fast simulation of input uncertainty (In Preparation)

Other References

[19] Jia W, Wang H, Chen M, et al. Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning SC20: International conference for high performance computing, networking, storage and analysis. IEEE, 2020: 1-14.
[20] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596(7873): 583-589.
[21] Amodei D, Olah C, Steinhardt J, et al. Concrete problems in AI safety. arXiv preprint arXiv:1606.06565, 2016.