Oppenheim Workshop 2024
Machine Learning and AI: Theory and Practice

(held in conjunction with the Oppenheim Lecture)

Date: 02 December 2024

Venue: Department of Mathematics
Seminar Room 1 (S17-04-06) National
University of Singapore

Speakers:

- Bin Yu, Statistics, EECS, CompBio and Simons Inst, UC Berkeley
  Veridical Data Science and Alignment in Medical AI
- Fei Wu, Zhejiang University
  Domain-Specific LLMs as well as their Application
- Shinji Nishimoto, Osaka University
  Insights into Perceptual and Cognitive Representations: Bridging the Human Brain and AI
- Kenji Fukumizu, The Institute of Statistical Mathematics
  Neural Fourier Transform: Learning Group Representation from Data
- Susan Wei, School of Mathematics and Statistics, The University of Melbourne
  What's Degeneracy Got to Do with it? Understanding Deep Neural Networks through the Local Learning Coefficient
- Zuowei Shen, National University of Singapore, Mathematics
  Deep Approximation via Deep Learning
- Qianxiao Li, NUS Math
  Learning, Approximation and Control
- Tan Nguyen, National University of Singapore, Mathematics
  Monomial Matrix Group Equivariant Neural Functional Networks
- Jialiang Li, National University of Singapore, Statistics
  Robust model averaging prediction of longitudinal response with ultrahigh-dimensional covariates
- Yi Li, Nanyang Technological University
  Subspace Embedding and Linear Regression Problems

Jointly organized by

NUS National University of Singapore | Department of Mathematics Faculty of Science

NUS National University of Singapore | IMS Institute for Mathematical Sciences

Programme

Monday 02 December 2024	
9.30 am to 9.35 am	**Weiqing Ren** Opening Remarks
9.35 am to 10.20 am	**Bin Yu (Chair: Weiqing Ren)** Title: Veridical Data Science and Alignment in Medical AI Abstract: Alignment and trust are crucial for the successful integration of AI in healthcare including digital twin projects, a field involving diverse stakeholders such as me dical personnel, patients, administrators, public health officials, and taxpayers, all of whom influence how these concepts are defined. This talk presents a series of collaborative medical case studies where AI algorithms progressively become, from transparency to more opaque thus with increasing difficulty of alignment assessment. These range from tree-based methods for trauma diagnosis, to LLM-based emergency department co-pilot, and mechanistic circuits for structured data extraction from pathology reports. They are guided by Veridical Data Science (VDS) principles—Predictability, Computability, and Stability (PCS)—for the goal of building trust and interpretability, enabling doctors to assess alignment. The talk concludes with a discussion on applying VDS to medical foundation models and next steps for evaluating AI algorithm alignment in healthcare.
10.20 am to 10.40 am	Coffee/Tea Break (Math Staff Lounge)
Session: Learning Theory (Chair: Vincent Tan)	
10.40 am to 11.00 am	**Qianxiao Li** Title: Learning, Approximation and Control Abstract: We discuss some problems and recent results on the interface of deep learning, approximation theory and control theory. Through a dynamical system viewpoint of deep residual architectures, the study of model complexity in deep learning can be formulated as approximation or interpolation problems that can be studied using control theory, but with a mean-field twist. In a similar vein, training deep architectures can be formulated as optimal control problems in the mean-field sense. We provide some basic mathematical results on these new control problems that so arise, and discuss some applications in improving efficiency, robustness and adaptability of deep learning models.n theory and control theory. Through a dynamical system viewpoint of deep residual architectures, the study of model complexity in deep learning can be formulated as approximation or interpolation problems that can be studied using control theory, but with a

	mean-field twist. In a similar vein, training deep architectures can be formulated as optimal control problems in the mean-field sense. We provide some basic mathematical results on these new control problems that so arise, and discuss some applications in improving efficiency, robustness and adaptability of deep learning models.
11.00 am to 11.20 am	**Kenji Fukumizu** Title: Neural Fourier Transform: Learning Group Representation from Data Abstract: In this study, we introduce a novel deep learning framework designed to infer group representations from data under the assumption that the data space is subject to an unknown group action. The data consists of examples of the group action, comprising a point and its transformation under a group element, or sequences generated through the successive application of a group element. Utilizing an autoencoder architecture, our approach maps the data to a latent space in a manner that is equivariant to the group action, achieving linear group action on the latent variables and thus approximating a group representation. Further, by applying block-diagonalization, we approximately decompose the representation into irreducible representations. We call this method the Neural Fourier Transform. This presents a generalized, data-driven approach to Fourier transform. We validate our framework across various scenarios, including one notable case where the group and its actions are not explicitly known, yet our method successfully learns the equivariant mapping and a group representation using only sequential data. Employing image sequences altered by transformations affecting color or shape, we demonstrate that our derived irreducible representations effectively disentangle the underlying generative processes of the data. Theoretical results supporting our methodology are also presented.
11.20 am to 11.40 am	**Susan Wei** Title: What's Degeneracy Got to Do with it? Understanding Deep Neural Networks through the Local Learning Coefficient Abstract: Deep neural networks (DNN) are singular statistical models that exhibit complex degeneracies. In this work, we introduce a quantity known as the Local Learning Coefficient (LLC) which precisely quantifies the degree of degeneracy in DNNs. Although the LLC is designed to address the limitations of traditional complexity measures, it coincides with familiar notions of complexity when the model is regular or minimally singular. We introduce a scalable estimator for the LLC and apply it across diverse DNN architectures including deep linear networks up to 100M parameters, ResNet image models, and transformer language models. Empirical evidence suggests that the LLC provides valuable insights into how common deep learning training heuristics might influence the effective complexity of DNNs. Ultimately, the LLC

	emerges as a valuable tool for reconciling the apparent contradiction between deep learning's complexity and the principle of parsimony.
11.40 am to 12.00 pm	**Yi Li** Title: Subspace Embedding and Linear Regression Problems Abstract: In this talk, I shall give a brief overview of the subspace embedding problem and related linear regression problems from a theoretical computer science perspective. Specifically, the ℓ_p subspace embedding problem asks to find a matrix S for a given n × d matrix A (where n>>d) such that ||SAx||_p ≈ ||Ax||_p for all d-dimensional vectors x, assuming p >= 1 is constant. This matrix S is often constructed using Lewis weights sampling. Using subspace embeddings, one can reduce the scale of the standard linear regression problem min_x ||Ax − b||_p. I shall also cover several regression problem variants and discuss the main results from the literature.
12.00 pm to 2.00 pm	Lunch (Math Staff Lounge)
2.00 pm to 2.45 pm	**Zuowei Shen (Chair: Hui Ji)** Title: Deep Approximation via Deep Learning Abstract: The primary task of many applications is approximating/estimating a function through samples drawn from a probability distribution on the input space. The deep approximation is to approximate a function by compositions of many layers of simple functions, that can be viewed as a series of nested feature extractors. The key idea of deep learning network is to convert layers of compositions to layers of tuneable parameters that can be adjusted through a learning process, so that it achieves a good approximation with respect to the input data. In this talk, we shall discuss mathematical theory behind this new approach and approximation rate of deep network; we will also show how this new approach differs from the classic approximation theory, and how this new theory can be used to understand and design deep learning networks.
2.45 pm to 3.00 pm	Coffee/Tea (Math Staff Lounge)
Session: Applications of AI and Learning (Chair: Li Qianxiao)	

3.00 pm to 3.20 pm	**Shinji Nishimoto** <u>Title</u>: Insights into Perceptual and Cognitive Representations: Bridging the Human Brain and AI <u>Abstract</u>: Our brains process vast amounts of sensory information to infer what is happening in the world, and with additional cognitive processing, they enable us to compute, judge, make decisions, and generate purposeful actions. By analyzing human brain activity during naturalistic tasks, such as watching videos or performing various cognitive tasks, we aim to understand how information is represented in the brain and to explore how these representations might resemble or differ from those of emerging intelligent agents, such as AI. In this talk, I will present our recent findings and insights into the perceptual and cognitive representations that underlie human thought and action.
3.20 pm to 3.40 pm	**Jialiang Li** <u>Title</u>: Robust Model Averaging Prediction of Longitudinal Response with Ultrahigh-Dimensional Covariates <u>Abstract:</u> Model averaging is an attractive ensemble technique to construct fast and accurate prediction. Despite of having been widely practiced in cross-sectional data analysis, its application to longitudinal data is rather limited so far. We consider model averaging for longitudinal response when the number of covariates is ultrahigh. To this end, we propose a novel two-stage procedure in which variable screening is first conducted and then followed by model averaging. In both stages, a robust rank-based estimation function is introduced to cope with potential outliers and heavy-tailed error distributions, while the longitudinal correlation is modelled by a modified Cholesky decomposition method and properly incorporated to achieve efficiency. Asymptotic properties of our proposed methods are rigorously established, including screening consistency and convergence of the model averaging predictor, with uncertainties in the screening step and selected model set both taken into account. Extensive simulation studies demonstrate that our method outperforms existing competitors, resulting in significant improvements in screening and prediction performance. Finally, we apply our proposed framework to analyse a human microbiome dataset, showing the capability of our procedure in resolving robust prediction using massive metabolites.

3.40 pm to 4.00 pm	**Fei Wu** <u>Title</u>: Domain-Specific LLMs as well as their Application <u>Abstract</u>: In recent years, some large language models (e.g., OpenAI's ChatGPT, and Google's PaLM) have been shown to exhibit more general intelligence than previous AI models across a variety of domains and tasks. These LLMs can generate novel and unexpected responses—a significant departure from earlier routine models that were limited to generating predictable and formulaic responses. In this talk, I will introduce how to train domain-specific LLMs for certain tasks such education and legal domains. The main topics consist of domain-specific SFT and the integration of data-driven and knowledge-guided techniques.
4.00 pm to 4.20 pm	**Tan Nguyen** <u>Title</u>: Monomial Matrix Group Equivariant Neural Functional Networks <u>Abstract</u>: Neural functional networks (NFNs) have recently gained significant attention due to their diverse applications, ranging from predicting network generalization and network editing to classifying implicit neural representation. Previous NFN designs often depend on permutation symmetries in neural networks' weights, which traditionally arise from the unordered arrangement of neurons in hidden layers. However, these designs do not take into account the weight scaling symmetries of ReLU networks, and the weight sign flipping symmetries of sin or Tanh networks. In this paper, we extend the study of the group action on the network weights from the group of permutation matrices to the group of monomial matrices by incorporating scaling/sign-flipping symmetries. Particularly, we encode these scaling/sign-flipping symmetries by designing our corresponding equivariant and invariant layers. We name our new family of NFNs the Monomial Matrix Group Equivariant Neural Functional Networks (Monomial-NFN). Because of the expansion of the symmetries, Monomial-NFN has much fewer independent trainable parameters compared to the baseline NFNs in the literature, thus enhancing the model's efficiency. Moreover, for fully connected and convolutional neural networks, we theoretically prove that all groups that leave these networks invariant while acting on their weight spaces are some subgroups of the monomial matrix group. We provide empirical evidences to demonstrate the advantages of our model over existing baselines, achieving competitive performance and efficiency.
4.20 pm to 5.30 pm	Free discussion at the Math Lounge