AI for Mathematics: Design and Practice of AI Agents

Bin Dong, Peking University

Artificial intelligence is transforming the paradigm of mathematical research, yet significant bottlenecks remain, including the difficulty of discovering cross-domain connections across the vast mathematical literature and the substantial gap between natural language proofs and formal verification. This talk presents three mathematical research agents developed by our group at Peking University: Archon, a formal verification agent powered by the formal retrieval engine LeanSearch; Rethlas, a reasoning agent powered by the informal retrieval engine Matlas; and Iteris, a research agent for computational mathematics. We highlight the design principles behind these agents and showcase representative cases in which they have been applied to solve open mathematical problems. We conclude with a brief outlook on future directions for AI-assisted mathematical research.

Harness In-Context Operator Learning with Chain of Operators

Ling Guo, Shanghai Normal University

While scientific foundation models show immense promise in solving numerical problems, they often struggle with out-of-distribution (OOD) generalization. When confronted with complex OOD problems, these models typically require expensive parameter fine-tuning or retraining. To achieve zero-shot adaptation without weight updates, numerical models require the equivalent of a "harness", analogous to the reasoning chains and tool use that successfully adapt frozen Large Language Models to complex tasks. In this talk, we explore this frontier based on In-Context Operator Networks (ICON), a numerical foundation model paradigm that infers operators directly from numerical prompts. To this end, we introduce Chain of Operators (CHOP), a framework that harnesses a frozen ICON to OOD tasks. CHOP systematically decomposes complex problems by routing numerical prompts through explicit, closed-form elementary transformations and multiple frozen-model calls, effectively mapping them back to the in-distribution regime. Across conservation laws, mean-field control, and cross-region air-quality forecasting, CHOP consistently reduces prediction error relative to direct frozen-ICON inference. Moreover, elementary-operator sequences discovered for one task remain effective across operator mappings and distinct PDE families. Taken together, these results demonstrate how ICON's intrinsic multi-operator in-context learning capability, coupled with an explicit inference-time harness, can serve as a cornerstone of modular, agentic scientific computing.

Discovering State Variables and Continuous-Time Dynamics from Video

Kuang Huang, The University of Hong Kong

Dynamical systems play a central role in scientific modeling, traditionally described by hand-chosen state variables, such as angle and angular velocity for a swinging pendulum, together with a set of differential equations governing their evolution. We present a data-driven method that automates this modeling process directly from raw video, discovering: (i) a set of state variables that preserves the continuous-time dynamics of the observed system, and (ii) a vector field on the resulting state space that captures its equation of motion. We evaluate the approach on a range of physical systems through both quantitative and qualitative analyses, including the recovery of characteristic frequencies and the detection of chaotic and limit-cycle behavior. These results suggest a path toward a data-driven modeling process that assists human scientists in accelerating scientific discovery.

Advances in Self-Supervised Image Denoising: From Gaussian Noise to Real-World Noise

Hui Ji, National University of Singapore

Image denoising is a fundamental task in image restoration and a key component in solving inverse imaging problems. Deep learning has achieved remarkable success in denoising, particularly through supervised learning; however, its reliance on ground-truth images for network training limits its broader applicability in real-world scenarios. Recent research has increasingly shifted towards truth-free learning paradigms, where models are trained directly on noisy data without requiring clean references. In this talk, I will present a sequence of works on self-supervised image denoising, progressively addressing more complex noise models, from i.i.d. Gaussian noise to pixel-wise independent heteroscedastic noise, and ultimately to pixel-wise correlated noise observed in real-world data. Built upon data augmentation and mask-and-predict training strategies, our approaches enable networks to learn denoising solely from noisy observations. Experimental results demonstrate that the proposed self-supervised denoisers achieve performance comparable to supervised methods, offering a practical solution for data-limited imaging applications.

Structure-Preserving Construction of Collision Operators for Kinetic Equations from Molecular Dynamics

Huan Lei, Michigan State University

We introduce a data-driven approach to learn generalized collision operators from molecular dynamics. Unlike conventional models (e.g., Landau), the present operator takes a symmetry-breaking non-stationary form that depends not only on the relative velocity but also on the average velocity of the collision pair, capturing heterogeneous energy transfer arising from collective interactions with the environment. The constructed model strictly preserves frame-indifference, conservation laws, and physical constraints such as the H-theorem. To enable efficient numerical evaluation, we develop a fast spectral separation method that represents the kernel as a low-rank tensor product of univariate basis functions. This formulation admits an O(N log N) algorithm and structure-preserving discretization. Numerical results demonstrate that the proposed model accurately captures challenging 1D3V plasma dynamics in the moderately coupled regime beyond the standard Landau model while maintaining high computational efficiency and structure-preserving properties.

Integrated Mathematical Modeling and Machine Learning Approach to Complex Biological Systems

Bo Li, University of California San Diego

Experiments and computer simulations provide abundant biological data f or creating artificial intelligence for biology. We develop an integrated m athematical modeling and data-driven machine learning approach to study ing complex biological systems, aiming at understanding biological intelli gence and specific rules governing the interactions and collective behavio rs of biological molecules and cells. This talk presents several case studie s to demonstrate how such an approach can lead to a quantitative understa nding of an underlying biological system. These studies include solving th e Poisson-Boltzmann equation with complicated molecular geometry for biomolecular solvation, identifying rate parameters in biochemical reactio n networks modeling cell signaling pathways, and predicting the kinetics of an expanding bacterial colony composed of millions of cells. Challenge s in the further development of rigorous mathematical theories and efficie nt learning techniques will be discussed.

Learning macroscopic and mesoscopic dynamics from data
Qianxiao Li, National University of Singapore

Classical modelling and scientific computing workflows for dynamical systems begin from a physically motivated model (e.g., ODE, SDE, PDE), investigate its well-posedness properties, and then design convergent numerical algorithms to solve them to obtain physical insight. However, for complex dynamical processes, obtaining an accurate and well-motivated model is the most challenging step. In this talk, we propose an alternative, learning-based modelling framework, which begins with the design of a model class instead of individual models, with physical constraints and mathematical and numerical well-posedness built-in. Using machine learning, we then identify particular models on computational or experimental data, with physical interpretation of the learned component as an a posteriori analysis step. We demonstrate this new modelling philosophy on several problems arising from science and engineering.

Random batch and Langevin algorithms: convergence, ergodicity and diffusion approximation
Jianguo Liu, Duke University

Random minibatching is a common computational principle in optimization, interacting particle simulation, and sampling: a large deterministic average is replaced by a much smaller random average at each step. This substantially reduces computational cost, but also creates new stochastic dynamics whose approximation errors and long-time behavior must be understood. This talk covers several results in three related directions. First, for stochastic gradient descent, I will discuss the uniform-in-time weak diffusion approximations in the finite-variance regime. Second, for interacting particle systems, I will introduce the Random Batch Methods (RBM) with a focus on the convergence analysis of RBM with replacement, which can also be viewed as the kinetic Monte Carlo applied to the particle system. Third, for Langevin-type sampling algorithms, I will talk about some recent progress of theoretical analysis for the stochastic gradient Langevin dynamics and the implicit Langevin Monte Carlo. This talk is mainly based on joint works with Lei Li (SJTU), Yuliang Wang (Duke), and others.

Geometric Priors in Deep Learning for Imaging via Quasiconformal Methods

Ronald Lui, The Chinese University of Hong Kong

Deep learning has achieved impressive success in image processing, yet its performance often degrades when images are noisy, blurred, or incomplete. In many applications, prior geometric information about underlying structures, such as shape or anatomical constraints, is available and can provide valuable regularization. In this talk, we present a quasiconformal framework for incorporating geometric priors into imaging models. Quasiconformal theories offer a mathematically principled way to control geometric distortions, enabling shape constraints to be naturally embedded into tasks such as image registration and segmentation. Using quasiconformal theories, we develop deep learning based quasiconformal models with geometric priors for various imaging tasks. Experimental results demonstrate improved robustness and accuracy in challenging imaging scenarios.

A deep learning method for computing the spectrum of the Schrodinger operator: numerics and analysis

Pingbing Ming, Chinese Academy of Sciences

We present a novel deep learning method for computing eigenstates of the Schrödinger operator. The proposed approach combines a novel loss function with an innovative neural network architecture that incorporates a-prior knowledge of the problem. These improvements enable the method to handle both high-dimensional problems and problems posed on irregular bounded domains. We successfully compute up to the first 30 eigenvalues for various Schrödinger operators. We analyze the generalization error in the framework of the Barron space, in which a regularity result has been established with the aid of operator theory. We also establish a Barron regularity result for the many-particle Schrödinger operator. As an application, we apply the method to fractional Schrödinger eigenvalue problems and the fractional isospectral problem. This is a joint work with Yixiao Guo and Hao Yu.

Deep Approximation via Deep Learning

Zuowei Shen, National University of Singapore

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, which can be viewed as a series of nested feature extractors. The key idea of a deep learning network is to convert layers of compositions to layers of tunable 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 the mathematical theory behind this new approach and the approximation rate of deep networks; 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.

Mathematical Limits of Explaining AI

Atsushi Suzuki, The University of Hong Kong

Attempts to explain AI, or to construct explainable AI, have been widely pursued. However, existing methods either explain only a small part of th e properties of practical AI systems used for complex tasks, such as gener al-purpose chat AI, in numerical or textual form, or construct AI systems t hat may be explainable but whose performance is degraded. Although it is clear that realising an AI system equipped with a complete and comprehe nsible explanation would bring great benefits to human society, existing r esearch has not clarified whether such a thing is possible in principle. The refore, our study takes the following research question: “For complex task s, does there exist a pair consisting of an AI system with performance hig h enough for practical use and an explanation that completely describes it s properties while being sufficiently short to be understandable to human s?" Our study proves a novel inequality that represents a fundamental limi t on AI explanations and gives a negative answer to the question above. B ased on the negative result, we clarify the directions in which research on AI explanations should proceed.

Mathematics as a Partner in Building AI Reasoning - Evaluating and Improving Mathematical Intelligence

Yang Wang, The University of Hong Kong

Mathematics has long provided a rigorous language for precise reasoning. It is now also becoming an important foundation for evaluating and improving the reasoning capabilities of artificial intelligence. This talk presents a connected line of work that studies mathematical reasoning in large language models from two perspectives: how to assess it more reliably, and how to use verifiable mathematical structure to improve it. On the evaluation side, E-GSM investigates whether models that perform well on short grade-school word problems can maintain their reasoning ability when the same mathematical ideas are embedded in longer and more realistic contexts. UGMathBench extends this investigation to undergraduate-level mathematics, examining consistency across diverse subjects and carefully varied versions of the same problem. Together, these studies show that final-answer accuracy alone is insufficient for evaluating mathematical reasoning; robustness across context, formulation, and domain is equally important. The second part of the talk turns from evaluation to improvement. Since many mathematical answers can be verified with high reliability, mathematics provides a natural source of feedback for training reasoning models. TFPI studies how to make this feedback-driven training process more efficient by reducing redundant reasoning while preserving useful learning signals. Composition-RL further explores how to expand verifiable training data by composing existing problems into new training tasks. Taken together, these works illustrate a productive symbiosis between mathematics and AI: mathematics provides rigorous tools for evaluating and training AI reasoning, while AI creates new opportunities for generating, exploring, and scaling mathematical problem solving.

Modeling Effects of Randomness in High Entropy Alloys

Yang Xiang, Hong Kong University of Science and Technology

High entropy alloys (HEAs) are single-phase crystals that consist of random solid solutions of multiple elements in approximately equal proportions. This class of novel materials has exhibited superb mechanical properties, such as high strength, which is associated with the motion of dislocations. We obtain a stochastic continuum model based on the Peierls-Nabarro framework for dislocations in an HEA from an atomistic model that incorporates the atomic-level randomness and short-range order. This approach provides a fundamental explanation for the origin of the high strength of HEAs based on the stochastic effects on the intrinsic strength. We also develop a machine learning framework to unravel the complex relationship between chemical composition and materials properties in HEAs.

Overparameterization as a general mechanism for accelerating gradient descent

Jinchao Xu, KAUST

Overparameterization is often viewed as a defining feature of modern machine learning. This talk argues that many advanced fast solvers in scientific computing can be interpreted through a similar lens. By enlarging the representation, hidden structures and effective correction directions become more accessible, leading to significantly faster gradient-based algorithms. Examples from numerical PDEs, together with MgNet and large language models, illustrate this perspective and its potential relevance to modern machine learning.

On the Trainability of Quantum Approximate Optimization Algorithms

Shengyu Zhang, Tencent Quantum Lab

Variational quantum algorithms (VQAs) are a cornerstone of quantum artificial intelligence. Among them, the quantum approximate optimization algorithm (QAOA) represents a canonical class of VQAs that could efficiently solve certain combinatorial optimization problems. Nevertheless, these algorithms often face significant trainability challenges, most notably the barren plateau (BP) phenomenon, where the gradient vanishes exponentially as the problem size increases. While prior research in this area has largely relied on heuristic approaches, in this talk we will rigorously establish conditions for the presence and absence of barren plateaus in QAOA. We first prove that QAOA encounters barren plateaus for almost all graphs, including both weighted and unweighted instances. We then demonstrate that QAOA on graphs with sufficient symmetry—such as cycles and complete graphs—does not exhibit barren plateaus. Our proofs leverage the dynamical Lie algebras associated with these variational circuits. These results provide new insights for the analysis and design of scalable variational quantum algorithms.

DeepParticle: learning PDE dynamics by a deep neural network minimizing Wasserstein distance on data generated from an interacting particle method

Zhiwen Zhang, The University of Hong Kong

High-dimensional PDEs are hard to compute with traditional mesh-based methods, especially with large gradients or unknown concentrations. Mesh-free methods are more appealing, but slow and expensive for long computations. We present DeepParticle, an approach integrating Deep Learning (DL), mini-batch Optimal Transport (OT), and interacting particle (IP) , first through a case study of Fisher-Kolmogorov-Petrovsky-Piskunov (FKPP) front speeds in incompressible flows. PDE analysis reduces the problem to computing the principal eigenvalue of an advection-diffusion operator. Feynman-Kac representation enables a genetic IP algorithm to evolve particle distribution to a time-invariant measure for front speed extraction. This measure is parameterized by the Peclet number. We learn this family of measures by training a physically parameterized DNN on affordable IP data at moderate Peclet numbers, then predict at larger, more expensive Peclet numbers. Our method extends to learning and generating other PDE dynamics, for which we show a second case study on aggregation patterns in Keller-Segel chemotaxis systems, and compare with diffusion models.

What can One Expect when Training Neural Networks to Solve PDEs?

Hongkai Zhao, Duke University

When neural networks (NNs) are used as a type of nonlinear parametric representation to solve partial differential equations (PDEs), they often display frequency-dependent learning dynamics that can differ from those seen in direct function approximation tasks, resulting from a balance between the frequency bias of the NN representation and that of the underlying differential operator. Using elliptic PDEs as examples, we show how these two factors compete and balance in different situations. Once the balance is determined, it is important to design computational strategies to counter the resulting bias to improve training efficiency. We propose a simple operator aware preconditioning strategy that rebalances the optimization landscape and the learning dynamics. Extensive experiments, including multiscale problems, show that our approach restores more balanced learning dynamics across modes and substantially improves both convergency and accuracy.

A parameterized Wasserstein Hamiltonian flow approach for solving the Schrödinger equation

Haomin Zhou, Georgia Institute of Technology

I will present a reformulation of the Schrödinger equation using a generative model and an associated computational method to solve it in higher dimensions. The reformulation stems from a parameterization approach that can be used to design algorithms simulating geometric flows on the Wasserstein manifold, the probability density space equipped with the optimal transport metric. The approach leverages optimal transport theory and Machine Learning techniques, such as push-forward operators and neural networks, resulting in a system of ODEs for the neural network parameters. The resulting methods are meshless, basisless, sample-based schemes that scale well to higher-dimensional problems. The strategy is applicable to the Wasserstein Hamiltonian flows, including the Schrödinger equation. This presentation is based on joint work with Shu Liu (Florida State), Hao Wu (Hong Kong Poly U), and Xiaojing Ye (Georgia State).