Mathematical Foundations & Methodologies for Artificial Intelligence & Data-Driven Scientific Computing
Workshop on Mathematical Foundations of AI Models
Abstracts

17 Aug 2026–21 Aug 2026

July 23, 2026

Abstracts

Tutorial

Principles and Practice of Deep Representation Learning Part 1

Yi Ma

The University of Hong Kong, Hong Kong SAR

Pending

Principles and Practice of Deep Representation Learning Part 2

Peng Wang

University of Macau, Macau SAR

Pending

Invited Talks

Training Neural Networks at Any Scale

Volkan Cevher

Swiss Federal Technology Institute of Lausanne, Switzerland

At the heart of deep learning's transformative impact lies the concept of scale–encompassing both data and computational resources, as well as their interaction with neural network architectures. Scale, however, presents critical challenges, such as increased instability during training and prohibitively expensive model-specific tuning. Given the substantial resources required to train such models, formulating high-confidence scaling hypotheses backed by rigorous theoretical research has become paramount.

To bridge theory and practice, the talk explores a key mathematical ingredient of scaling in tandem with scaling theory: the numerical solution algorithms commonly employed in deep learning, spanning domains from vision to language models. We unify these algorithms under a common master template, making their foundational principles transparent. In doing so, we reveal the interplay between adaptation to smoothness structures via online learning and the exploitation of optimization geometry through non-Euclidean norms. Our exposition moves beyond simply building larger models–it emphasizes strategic scaling, offering insights that promise to advance the field while economizing on resources.

Anatomy of a Hallucination: How Diffusion Solvers Shape Generation Errors

Grigorios Chrysos

University of Wisconsin, USA

Diffusion models consistently produce stunningly realistic images, but they also frequently synthesize images that hallucinate. Building a reliable pipeline free of structural hallucinations forces us to look closely at our methods: do we rely on stochastic or deterministic sampling? Since stochastic methods inject noise that can cause the generation path to drift, deterministic methods often feel like the obvious fix. But does the empirical data actually support this intuition? This talk explores whether choosing a deterministic method genuinely prevents errors, or if it simply alters the specific hallucination profile. In the second part, we augment the question: can we use any of the gradient-based methods for reducing hallucinations?

Approximating PDEs with Neural Networks via Residual Minimization of Monotone Discretizations

Carlos Esteve-Yagüe

University of Alicante, Spain

In recent years, advancements in deep learning and new optimization algorithms have motivated the use of artificial neural networks to solve non-linear problems in high-dimensional setups. One of the crucial steps in implementing any deep learning method is the choice of the loss functional used to train the network parameters, typically through gradient-based optimization. This talk focuses on the approximation of Partial Differential Equations (PDEs) using artificial neural networks. I will show that by utilizing a monotone discretization of the underlying differential operator, one can ensure the associated loss functional has a unique critical point. Moreover, for sufficiently coarse discretizations, we prove a dimension-robust Polyak-Łojasiewicz inequality. This implies linear convergence of the associated gradient flow at a rate that avoids the curse of dimensionality. Building on this theoretical foundation, we propose a multi-level training algorithm that exploits the faster convergence available on the coarser grids of the discretization.

TBA

Kenji Fukumizu

The Institute of Statistical Mathematics, Japan

Pending

Finding Latent Structure at Very Low SNR: Heterogeneity Analysis in Cryo-EM

Marc Aurèle Gilles

Princeton University, USA

Biomolecules are not static objects: their motions and conformational changes are often central to their biological function. Cryogenic electron microscopy offers a way to study these motions by imaging large ensembles of molecules, each potentially captured in a different conformation. The central challenge is that each molecule is observed only through an extremely noisy two-dimensional projection, making the recovery of its underlying three-dimensional conformational landscape a difficult high-dimensional inverse problem.

I will introduce the cryo-EM heterogeneity problem and describe classical and deep-learning approaches for reconstructing conformational variability. These include RECOVAR, which combines linear dimensionality reduction with kernel regression to represent a continuum of three-dimensional structures. I will then discuss how one can use these representations to estimate conformational populations, identify stable states, and infer plausible transition pathways between them. A recurring theme will be the effect of noise and uncertainty, which can obscure the underlying landscape, create spurious heterogeneity, and complicate the validation of reconstructed structures.

How Mathematics Shapes AI for Science

Youngjoon Hong
Seoul National University, S. Korea

Artificial intelligence is becoming a powerful tool for scientific modeling, simulation, and discovery. At the same time, many of its central challenges are deeply mathematical, involving approximation, stability, generalization, dynamics, and uncertainty. In this talk, I will discuss how mathematical analysis can shape the development of AI for Science. I will first present examples from scientific machine learning and operator learning for partial differential equations, including finite element-based neural operator methods. I will also discuss related mathematical perspectives on neural network approximation, Log Barron spaces, and Transformer architectures. I will then turn to generative models, including diffusion models, normalizing flows, and flow matching, and explain how they can be understood through stochastic dynamics, transport, and continuity equations. Finally, I will highlight applications and open mathematical problems in AI for Science, with an emphasis on reliable and interpretable scientific AI.

Deep Learning based Surface Reconstruction from PointClouds

Myungjoo Kang
Seoul National University, S. Korea

In this presentation, we propose a novel deep learning framework for surface reconstruction from unorganized point clouds by leveraging implicit surface representations via level set functions. Our method guarantees watertightness and exhibits robust adaptability to varying topologies. The core of our approach involves solving a p-Poisson equation to learn the signed distance function (SDF) with high precision, facilitated by a variable splitting strategy that introduces the SDF gradient as an auxiliary field. To further enhance the physical plausibility and stability of the

reconstruction, we impose a curl-free constraint on the auxiliary variable, reflecting the irrotational property of conservative vector fields. Extensive numerical experiments demonstrate that the proposed integration of partial differential equation modeling and vector field constraints enables accurate and topologically consistent surface reconstruction, without requiring any predefined geometric priors.

Beyond Learning: Toward a Mathematical Theory of Computation for Artificial Intelligence

Gitta Kutyniok

Ludwig Maximilian University of Munich, Germany

Artificial intelligence is commonly studied through the lenses of approximation, optimization, and learning theory. This perspective has led to major advances in our understanding of expressivity, trainability, generalization, and robustness of modern AI systems.

However, these analyses implicitly assume that the underlying computations can actually be carried out. This raises a more fundamental question: what are the computational limits of AI systems themselves? At the same time, the rapidly increasing energy demand of large-scale AI systems highlights that computation is not only a theoretical issue, but also a practical and societal bottleneck.

In this talk, I will argue that a comprehensive mathematical foundation of artificial intelligence requires extending the current learning-centered viewpoint toward a theory of computation for AI. I will discuss recent results showing that several problems arising in machine learning, inverse problems, and scientific computing are not computable within the classical Turing framework, thereby revealing fundamental limitations of current digital computing paradigms.

These observations naturally motivate the exploration of next-generation computing approaches, including neuromorphic and analog computing. I will present recent mathematical advances on spiking neural networks, including results on expressivity, universal approximation, and stability, which provide a rigorous foundation for emerging computing architectures. The talk concludes with a perspective on how computability theory, approximation theory, optimization, and learning theory may jointly contribute to a future theory of computation for AI.

TBA

Yi Ma

The University of Hong Kong, Hong Kong SAR

Pending

Human-like Common Sense as Core Inductive Bias for General Intelligence

Roland Memisevic

University of Toronto, Canada

Learning theory postulates that no model can learn without inductive biases - built-in assumptions about the model architecture or the environment it operates in. In this talk I will argue that an appreciation of inductive biases has taken a back-seat in recent years, partly due to the widely adopted race to scale large language models towards "general intelligence". I will discuss a variety of tasks that are very easy for humans to solve but that are very hard, or even impossible to learn, for any current frontier AI models. They include simple algorithmic reasoning tasks, compositional image understanding tasks, and simple real-world robotics tasks with partial observability. I will also discuss how these can be solved easily by models with appropriate inductive biases, such as recurrence and local (foveated) vision. Based on these results, I will argue that human-inspired models and human-like common sense are vastly undervalued guiding principles in building AI models today.

The Geometry of Training Dynamics

Govind Menon

Brown University, USA

I will provide an update of several recent developments in training dynamics for deep learning. This includes a collection of rigorous results on the deep linear network along with general geometric principles for all feed forward neural networks. In particular, we will explain the principle that "neural architecture determines a canonical Riemannian geometry".

Training Task Diversity Shapes In-Context Learning via Low-Dimensional Subspaces

Qing Qu

University of Michigan, USA

The transformer's emergent ability to perform in-context learning (ICL) has sparked a wide range of studies designed to understand its underlying mechanism. Existing works often study how training task diversity, defined either as the number of ICL training task vectors or as the number of function classes from which the task vectors are drawn, shapes both the generalization capabilities and the learning dynamics of ICL. While both definitions have uncovered many interesting phenomena, many observations under the latter definition remain theoretically unexplained. This talk presents a minimal analytical model under which these phenomena provably emerge from the properties of the pre-training data. By modeling the pre-training task vectors as a mixture of low-rank Gaussians, we show that pre-training task diversity, defined by the number of non-overlapping columns between the subspaces that parameterize the covariance matrices, improves both the generalization and the optimization trajectory of ICL with linear attention. In particular, we show that our model can explain (i) why ICL can achieve out-of-distribution generalization, and (ii) why pre-training with multiple tasks can shorten the ICL training plateau. We conclude by showing how our results empirically extend to nonlinear transformers and nonlinear function classes. Overall, our work presents a mathematically tractable framework to unify existing observations.

TBA

Quang Khoat Than

Hanoi University of Science and Technology, Vietnam

Pending

TBA

Rene Vidal

University of Pennsylvania, USA

Pending

Test-Time Guidance for Flow-Based Generative Models via Enhanced Sampling

Bao Wang

University of Utah, USA

Generative models that transport a simple source distribution to a complex data distribution—such as diffusion and flow-based models—are central to high-fidelity data generation. Test-time guidance can further steer pretrained models toward user-specified high-reward regions without costly retraining. However, existing guidance methods face critical limitations: they struggle with non-differentiable rewards, fail to navigate complex landscapes, and often lack theoretical guarantees on generation performance. We will talk about our recently proposed *Source Parallel Tempering (SPT)*, a gradient-free test-time guidance framework that operates entirely in source space, leveraging its simpler geometry to avoid the complexities of the data manifold. SPT couples a local exploration kernel with parallel tempering, enabling efficient barrier crossing and robust discovery of high-reward modes.

Functional Scaling Laws for LLM Pretraining

Lei Wu

Peking University, China

Scaling laws are among the most influential empirical regularities in LLM pretraining: final performance improves predictably as data, model size, and compute scale. However, classical scaling laws primarily describe the endpoint of training, leaving open the dynamical mechanism that produces these laws and how scaling behavior is shaped by hyperparameters. In this talk, I will present Functional Scaling Laws (FSL), a framework that extends scaling laws from final-loss prediction to full loss-trajectory prediction. Starting from power-law kernel regression as a toy model, we derive stochastic training dynamics and reveal an intrinsic-time structure that unifies training curves across scales and protocols. This functional perspective provides a dynamical theory of scaling laws. It explains how signal learning and noise forgetting jointly determine the loss dynamics, and shows how learning-rate and batch-size schedules shape scaling behavior through this balance. Experiments on LLM pretraining support FSL as a framework for interpreting and designing large-scale pretraining.

TBA

Jinchao Xu

King Abdullah University of Science and Technology, Saudi Arabia

Pending

Accelerating LLM Pre-Training through Flat-Direction Dynamics Enhancement

Kun Yuan

Peking University, China

Pre-training Large Language Models requires immense computational resources, making optimizer efficiency essential. The optimization landscape is highly anisotropic, with loss reduction driven predominantly by progress along flat directions. While matrix-based optimizers such as Muon and SOAP leverage fine-grained curvature information to outperform AdamW, their updates tend toward isotropy–relatively conservative along flat directions yet potentially aggressive along sharp ones. To address this limitation, we first establish a unified Riemannian Ordinary Differential Equation (ODE) framework that elucidates how common adaptive algorithms operate synergistically: the preconditioner induces a Riemannian geometry that mitigates ill-conditioning, while momentum serves as a Riemannian damping term that promotes convergence. Guided by these insights, we propose LITE, a generalized acceleration strategy that enhances training dynamics by applying larger Hessian damping coefficients and learning rates along flat trajectories. Extensive experiments demonstrate that LITE significantly accelerates both Muon and SOAP across diverse architectures (Dense, MoE), parameter scales (130M–1.3B), datasets (C4, Pile), and learning-rate schedules (cosine, warmup-stable-decay). Theoretical analysis confirms that LITE facilitates faster convergence along flat directions in anisotropic landscapes, providing a principled approach to efficient LLM pre-training.

How Does Learning Progress in Neural Network Training? River-Valley Landscapes and Optimizer Dynamics

Chulhee Yun

Korea Advanced Institute of Science & Technology, S. Korea

Despite the success of modern neural network training, it remains unclear which aspects of parameter-space motion are actually responsible for learning progress. A

common view is that the essential training dynamics are largely captured by a low-dimensional subspace associated with the dominant, high-curvature directions of the loss landscape. In this talk, I will present recent evidence challenging this view: although these sharp directions dominate local curvature and gradient dynamics, movement outside the dominant subspace appears to be essential for progress. These observations motivate a river-valley view of neural network loss landscapes. In this picture, sharp directions form the walls of a narrow valley and generate large oscillatory gradients, while effective learning proceeds along much flatter "river" directions. This perspective provides a useful framework for understanding several recurring phenomena in neural network optimization, including the effectiveness of exponential moving averages of model weights, the characteristic loss trajectory induced by warmup–stable–decay learning rate schedules, and the benefits of momentum. It also sheds light on the behavior of more recent optimizers. Schedule-free methods can be interpreted as producing low-loss trajectories that track progress along the river, whereas orthogonalization-based methods such as Muon can accelerate motion in river directions but may also amplify oscillations against the valley walls. Finally, I will describe how this tradeoff motivates AMUSE (Anytime MUon with Stable Gradient Evaluation), a new method designed to combine Muon's rapid river-direction progress with the stabilizing effect of schedule-free averaging.

Multigrade Neural Network Approximation: Residual Refinement and Layer-wise Rates

Shijun Zhang

Hong Kong Polytechnic University, Hong Kong SAR

We present recent theoretical results on multigrade neural network approximation. Multigrade deep learning constructs deep networks grade by grade: previously learned grades are frozen, and each newly added grade-wise subnetwork is composed on top of the previous grades and trained to fit the residual left by the current approximation, yielding a structured hierarchical refinement process. We first show that, for any continuous target function on a hypercube, there exists a fixed-width ReLU multigrade construction whose residuals converge uniformly to zero with strict decay at every nontrivial grade. We then discuss a related layer-wise approximation theorem, showing that intermediate readouts of a fixed-width architecture are certified approximants with errors controlled at progressively finer geometric scales. Together, these results interpret depth as residual refinement through composition and multiscale approximation, with potential relevance to scientific machine learning for differential equations.

Inference-Time Learning Through Context

Zhihui Zhu

Ohio State University, USA

Large language models exhibit a remarkable ability to learn and adapt from context without updating their parameters. In this talk, I will present our recent work on both understanding how foundation models extract task information from context and designing contexts that enable continual improvement during inference. I will first discuss a geometric analysis on in-context learning, providing new insights into how task representations emerge and evolve across layers. I will then discuss how these insights motivate a broader paradigm of inference-time learning, in which context is actively constructed rather than passively consumed. Building on iterative refinement frameworks such as AlphaEvolve, we view context as an evolving memory that stores hypotheses, intermediate solutions, and feedback. Drawing inspiration from optimization and sequential Monte Carlo, we develop principled approaches for designing and updating context over time. Overall, understanding how models read context and how we can systematically write and evolve context may provide a foundation for the next generation of adaptive AI systems.
