IMS Young Mathematical Scientists Forum — Applied Mathematics

Abstracts

Contents

Abstracts are updated as of 29 December 2023.

Chenglong Bao
Tsinghua University, China

Robust AI-aided Imaging Models without Labelled Samples

The observations in practical imaging systems always contain complex noise such that classical approaches are difficult to obtain satisfactory results. In recent years, deep neural networks directly learned a map between the noisy and clean images based on the training on paired data. Despite its promising results in various tasks, collecting the training data is difficult and time-consuming in practice. In this talk, in the unpaired data regime, we will discuss our recent progress for building AI-aided robust models and their applications in image processing. Leveraging the Bayesian inference framework, our model combines classical mathematical modelling and deep neural networks to improve interpretability. Experimental results on various real datasets validate the advantages of the proposed methods. Finally, I will report the recent progresses on solving the preferred orientation problems in cyroEM using the developed tools.

Back to Table of Contents

Zhenning Cai
National University of Singapore, Singapore

Symmetric Gauss-Seidel method for the steady-state Boltzmann equation

The Boltzmann equation is one of the fundamental kinetic models for rarefied gas dynamics. We consider the numerical solver of the steady-state Boltzmann equation using iterative methods based on the symmetric Gauss-Seidel (SGS) method. A naive application of the SGS method will lead to a convergence rate that decreases to zero as the density of the gas increases. To address this problem, a classical method is to use the Euler equations or Navier-Stokes equations, which are models for dense gases that can be considered as the dense limit of the Boltzmann equation, as preconditioners of the iterative methods. However, this requires an additional numerical solver for the Euler or Navier-Stokes equations, which may also be costly. Our solution is to use the dense limit as the preconditioner only when solving the local problem during the SGS iteration. This also eliminates the slowdown of the convergence when the gas gets dense. The proposed approach is also coupled with the multigrid method to further accelerate the convergence.

Back to Table of Contents

Dongdong Chen
University of Edinburgh, UK

Equivariant Imaging: Learning to image without ground truth

Deep networks provide state-of-the-art performance in many inverse imaging problems, ranging from medical imaging to computational photography. In several imaging problems, we typically only have access to compressed measurements of the underlying signals, which complicates most learning-based strategies that typically require pairs of signals and associated measurements for training. Learning only from compressed measurements is generally impossible, as the compressed observations do not contain information outside the range of the forward sensing operator. In this talk, I will present a new end-to-end self supervised framework, called Equivariant Imaging (EI), which overcomes this limitation by exploiting the equivariances present in natural signals. Our proposed learning strategy performs as well as fully supervised methods. Experiments demonstrate the potential of this framework on inverse problems including sparse-view X-ray computed tomography, accelerated MRI, and image inpainting.

Back to Table of Contents

Zhiyan Ding
University of California, Berkeley, USA

Quantum phase estimation and Signal processing

Efficiently estimating eigenvalues of exponentially large Hamiltonians on quantum computers, a task known as quantum phase estimation, is an important problem in quantum chemistry. In this presentation, I will begin with a concise introduction and overview of the quantum phase estimation problem. Next, I will demonstrate the transformation of the quantum phase estimation problem into a signal processing problem, showing the limitations of classical signal processing algorithms for this specific task. Finally, I will introduce a novel and simple signal processing algorithm designed to address this problem. It's worth noting that this presentation assumes no prior knowledge of quantum computing.

Back to Table of Contents

Sebastian Goldt
International School of Advanced Studies (SISSA), Italy

Learning from higher-order correlations, efficiently

Higher-order correlations in data are important, but how do neural networks extract information from them efficiently? We study this question in simple models of single- and two-layer neural networks. We first show that neural networks learn the statistics of their data in a hierarchical way. We then show that while learning from higher-order correlations is expensive in terms of sample complexity, correlations between the latent variables of the data help neural networks accelerate learning. We close by discussing some phase transitions in the higher-order cumulants of inputs with translation symmetry, and discuss their importance for feature learning in neural networks.

Back to Table of Contents

Jiequn Han
Flatiron Institute, USA

Designing High-Dimensional Closed-Loop Optimal Control Using Deep Neural Networks

Designing closed-loop optimal control for high-dimensional nonlinear systems remains a persistent challenge. Traditional methods, such as solving the Hamilton-Jacobi-Bellman equation, suffer from the curse of dimensionality. Recent studies introduced a promising supervised learning approach, akin to imitation learning, that uses deep neural networks to learn from open-loop optimal control solutions.

In this talk, we'll explore this method, highlighting a limitation in its basic form: the distribution mismatch phenomenon, induced by controlled dynamics. To overcome this, we present an improved approach—the initial value problem enhanced sampling method. This method not only provides a theoretical edge over the basic version in the linear-quadratic regulator but also showcases substantial numerical improvement on various high-dimensional nonlinear problems, including the optimal reaching problem of a 7 DoF manipulator. Notably, our method also surpasses the Dataset Aggregation (DAGGER) algorithm, widely adopted in imitation learning, with significant theoretical and practical advantages.

Back to Table of Contents

Marko Hans Weber
National University of Singapore, Singapore

General Equilibrium with Unhedgeable Fundamentals and Heterogeneous Agents

We examine the implications of unhedgeable fundamental risk, combined with agents' heterogeneous preferences and wealth allocations, on dynamic asset pricing and portfolio choice. We solve in closed form a continuous-time general equilibrium model in which unhedgeable fundamental risk affects aggregate consumption dynamics, rendering the market incomplete. Several long-lived agents with heterogeneous risk-aversion and time-preference make consumption and investment decisions, trading risky assets and borrowing from and lending to each other. We find that a representative agent does not exist. Agents trade assets dynamically. Their consumption rates depend on the history of unhedgeable shocks. Consumption volatility is higher for agents with preferences and wealth allocations deviating more from the average. Unhedgeable risk reduces the equilibrium interest rate only through agents' heterogeneity and proportionally to the cross-sectional variance of agents' preferences and allocations.

Back to Table of Contents

Nguyen Hung Minh Tan
National University of Singapore, Singapore

Transformers Meet Image Denoising: Mitigating Over-smoothing in Transformers via Regularized Nonlocal Functionals

Transformers have achieved remarkable success in a wide range of natural language processing and computer vision applications. However, the representation capacity of a deep transformer model is degraded due to the over-smoothing issue in which the token representations become identical when the model's depth grows. In this work, we show that self-attention layers in transformers minimize a functional which promotes smoothness, thereby causing token uniformity. We then propose a novel regularizer that penalizes the norm of the difference between the smooth output tokens from self-attention and the input tokens to preserve the fidelity of the tokens. Minimizing the resulting regularized energy functional, we derive the Neural Transformer with a Regularized Nonlocal Functional (NeuTRENO), a novel class of transformer models that can mitigate the over-smoothing issue. We empirically demonstrate the advantages of NeuTRENO over the baseline transformers and state-of-the-art methods in reducing the over-smoothing of token representations on various practical tasks, including object classification, image segmentation, and language modelling.

Back to Table of Contents

David Itkin
Imperial College London, UK

Are linear strategies nearly optimal when trading with superlinear transaction costs?

Financial managers, when constructing trading strategies, need to take trading costs into account for portfolio construction as these costs can erode financial gains. In this work we consider an investor with mean-variance preferences who trades on a noisy signal in the presence of transaction costs. When the costs are quadratic in the trade size, the celebrated results of Garleanu & Pedersen (2013) establish that the explicit optimal strategy is a linear feedback function of the current signal and position size. However, empirical evidence suggests that for many assets costs are superlinear, but not quadratic. In this case no explicit solution is available and the optimization problem can be computationally intensive to solve. Motivated by the quadratic cost case we consider a tractable one parameter class of linear strategies. We show that under realistic choices of the parameters the best strategy in this class performs nearly as well as the true optimum, which we compute with a brute force numerical method. Our result gives a simple and practical rule of thumb that can be efficiently implemented and yields nearly optimal performance.

This is joint work with Xavier Brokmann, Johannes Muhle-Karbe and Peter Schmidt.

Back to Table of Contents

Lexiao Lai
Columbia University, USA

Global stability of first-order methods for coercive tame functions

We consider first-order methods with constant step size for minimizing locally Lipschitz coercive functions that are tame in an o-minimal structure on the real field. We prove that if the method is approximated by subgradient trajectories, then the iterates eventually remain in a neighbourhood of a connected component of the set of critical points. Under suitable method-dependent regularity assumptions, this result applies to the subgradient method with momentum, the stochastic subgradient method with random reshuffling and momentum, and the random-permutations cyclic coordinate descent method.

Back to Table of Contents

Buyang Li
The Hong Kong Polytechnic University, China

Parametric finite element approximations of surface evolution in geometric flow

---

We briefly review the origin and development of parametric finite element methods for surface evolution in geometric flow, and the development of artificial tangential motions to improve the mesh quality of the numerically computed surfaces, as well as methods that preserve the geometric structures. Then we report some of our recently developed algorithms to improve the mesh quality and our convergence analysis for the algorithms approximating surface evolution in geometric flow.

Back to Table of Contents

Qianxiao Li
National University of Singapore, Singapore

Approximation theory of deep learning for sequence modelling

In this talk, we present some recent results on the approximation theory of deep learning architectures for sequence modelling. In particular, we formulate a basic mathematical framework, under which different popular architectures such as recurrent neural networks, dilated convolutional networks (e.g. WaveNet), encoder-decoder structures, and most recently - transformers - can be rigorously compared. These analyses reveal some interesting connections between approximation, memory, sparsity/low-rank, graphical structures that may guide the practical selection and design of these network architectures.

Back to Table of Contents

Ting Lin
Peking University, China

Universal Approximation and Expressive Power of Deep Neural Networks

In this talk, we will discuss the universal approximation properties of deep neural networks, especially ResNets. We will start from the continuous-time ResNet, and leverage tools from control theory. This approach allows us to explore the expressive power of neural networks by its depth. The connection of this continuous resnets to the deep residual networks will be given. Additionally, we will discuss the generalization on neural networks with symmetry, e.g., the permutation-invariant case and the CNN case.

Back to Table of Contents

Jingwei Liang
Shanghai Jiao Tong University, China

Partial Smoothness: from Functions to Sub-differentials and Beyond

Over the past decade, partial smoothness has played important roles in local convergence analysis, sensitivity analysis and algorithm acceleration. However, one limitation of partial smoothness is its restrictions to functions in finite dimension. In this talk, I will introduce an extension of partial smoothness from functions to setvalued operators. Under such an extension, we can obtain an intuitive geometric interpretation of finite activity identification in proximal operator splitting algorithms, and provide an upper bound for such a property where as existing results only consider the lower bound. Moreover, we can also derive identification results for problems that are degenerate.

Back to Table of Contents

Georg Maierhofer
University of Cambridge, UK

Structure-preserving low-regularity integrators for dispersive nonlinear equations

Dispersive nonlinear partial differential equations describe a range of phenomena, from water waves to spin states in ferromagnetism. Numerical approximation of solutions with limited differentiability (low-regularity) is crucial for simulating emerging structures in random wave fields and domain wall states, but pose a significant challenge to classical algorithms. Recent years have seen the development of tailored low-regularity integrators to address this challenge.

Many dispersive nonlinear equations possess a rich geometric structure, such as a Hamiltonian formulation and conservation laws. To ensure that numerical schemes lead to meaningful results, it is vital to preserve this structure in numerical approximations. This, however, results in an interesting dichotomy: the rich theory of existent structure-preserving algorithms is typically limited to classical integrators that cannot reliably treat low-regularity phenomena, while most prior designs of low-regularity integrators break geometric structure in the equation. In this talk, we will outline recent advances incorporating structure-preserving properties into low-regularity integrators. Starting from simple discussions on the Korteweg–de Vries and the nonlinear Schrödinger equation we will discuss the construction of such schemes for a general class of dispersive equations before demonstrating an application to the simulation of low-regularity vortex filaments.

This is joint work with Yvonne Alama Bronsard, Valeria Banica, Yvain Bruned and Katharina Schratz.

Back to Table of Contents

Qing Qu
University of Michigan, USA

Invariant Low-Dimensional Subspaces in Gradient Descent for Learning Deep Networks

Over the past few years, an extensively studied phenomenon in training deep networks is the implicit bias of gradient descent towards parsimonious solutions. In this work, we first investigate this phenomenon by narrowing our focus to deep linear networks. Through our analysis, we reveal a surprising "law of parsimony" in the learning dynamics when the data possesses low-dimensional structures. Specifically, we show that the evolution of gradient descent starting from orthogonal initialization only affects a minimal portion of singular vector spaces across all weight matrices. In other words, the learning process happens only within a small invariant subspace of each weight matrix, even though all weight parameters are updated throughout training. This simplicity in learning dynamics could have significant implications for both efficient training and a better understanding of deep networks. First, the analysis enables us to considerably improve training efficiency by taking advantage of the low-dimensional structure in learning dynamics. We can construct smaller, equivalent deep linear networks without sacrificing the benefits associated with the wider counterparts. Moreover, we demonstrate the potential implications for efficient training deep nonlinear networks. Second, it allows us to better understand deep representation learning by elucidating the progressive feature compression and discrimination from shallow to deep layers. The study paves the foundation for understanding hierarchical representations in deep nonlinear networks.

Back to Table of Contents

Zhenjie Ren
Université Paris-Dauphine, France

Regularized Mean Field Optimization with Application to Neural Networks

Our recent research focuses on advancing the field of regularized mean field optimization, with the overarching goal of establishing a theoretical foundation for evaluating the effectiveness of the training of neural networks and inspiring novel training algorithms. In this presentation, we will provide a comprehensive overview of the McKean-Vlasov dynamics, which serves as gradient flows, approaching the minimizer of regularized mean field optimization. We will place particular emphasis on examining the long-time behaviour and the particle approximation of such McKean-Vlasov dynamics. Besides the gradient flows, we also introduce and investigate alternative algorithms, such as our recent work on the self-interaction diffusion, to search for the optimal weights of neural networks. Each of these algorithms is ensured to have exponential convergence, and we will showcase their performances though simple numerical tests.

Back to Table of Contents

Leandro Sánchez-Betancourt
University of Oxford, UK

Automated Market Makers Designs beyond Constant Functions

Popular automated market makers (AMMs) use constant function markets (CFMs) to clear the demand and supply in the pool of liquidity. A key drawback in the implementation of CFMs is that liquidity providers (LPs) are currently providing liquidity at a loss, on average. In this paper, we propose two new designs for decentralised trading venues, the arithmetic liquidity pool (ALP) and the geometric liquidity pool (GLP). In both pools, LPs choose impact functions that determine how liquidity taking orders impact the marginal exchange rate of the pool, and set the price of liquidity in the form of quotes around the marginal rate. The impact functions and the quotes determine the dynamics of the marginal rate and the price of liquidity. We show that CFMs are a subset of ALP; specifically, given a trading function of a CFM, there are impact functions and quotes in the ALP that replicate the marginal rate dynamics and the execution costs in the CFM. For the ALP and GLP, we propose an optimal liquidity provision strategy where the price of liquidity maximises the LP's expected profit and the strategy depends on the LP's (i) tolerance to inventory risk and (ii) views on the demand for liquidity. Our strategies admit closed-form solutions and are computationally efficient. We show that the price of liquidity in CFMs is suboptimal in the ALP. Also, we give conditions on the impact functions and the liquidity provision strategy to prevent arbitrages from roundtrip trades. Finally, we use transaction data from Binance and Uniswap v3 to show that liquidity provision is not a loss-leading activity in the ALP.

Back to Table of Contents

Yongsheng Soh
National University of Singapore, Singapore

Group Invariant Regularizers

Symmetries are abound in data. For instance, a collection of natural images still resemble a collection of natural images even after performing rotations or translations to these images. This begs the question: How do we leverage these sort of symmetries in data processing? For instance, can we develop general data processing methods that respect a broad collection of symmetries?

In this talk, we will consider the problem of learning convex regularizers directly from data that respects a very broad family of symmetries. Ideas from the representation theory of compact topological groups as well as semidefinite programming will feature prominently. We will describe an end-to-end recipe for learning group invariant regularizers that can be implemented readily using standard off-the-shelf software.

Back to Table of Contents

Xin Tong
National University of Singapore, Singapore

Ensemble Kalman inversion for high dimensional problems

Ensemble Kalman inversion (EKI) is an ensemble-based method to solve inverse problems.

Its gradient-free formulation makes it an attractive tool for problems with involved formulation.

However, EKI suffers from the ``subspace property'', i.e., the EKI solutions are confined in the subspace spanned by the initial ensemble. It implies that the ensemble size should be larger than the problem dimension to ensure EKI's convergence to the correct solution. Such scaling of ensemble size is impractical and prevents the use of EKI in high dimensional problems. To address this issue, we propose a novel approach using localization and dropout regularization to mitigate the subspace problem. We prove that these methods converge in the small ensemble settings, and the computational cost of the algorithm scales linearly with dimension. We also show that they reach the optimal query complexity, up to a constant factor. Numerical examples demonstrate the effectiveness of our approach.

Back to Table of Contents

Ruiyi Yang
Princeton University, USA

Optimization on Manifolds via Graph Gaussian Processes

Optimization problems on smooth manifolds are ubiquitous in science and engineering. Oftentimes the manifolds are not known analytically and only available as an unstructured point cloud, so that gradient-based methods are not directly applicable. In this talk, we shall discuss a Bayesian optimization approach, which exploits a Gaussian process over the point cloud and an acquisition function to sequentially search for the global optimizer. Regret bounds are established and several numerical examples demonstrate the effectiveness of our method.

Back to Table of Contents

Douglas Dongzhuo Zhou
Shanghai Jiao Tong University, China

Model Reduction of Spatial Neurons through Biophysical Modelling and Data-driven Approaches

Modelling the dynamics of single neurons is the initial stride towards a quantitative understanding of brain computation. However, existing point neuron models fall short in capturing the intricate dendritic effects, which hold paramount importance in neuronal information processing. In this study, we conduct an asymptotic analysis of passive cable theory, leading to the reduction of the PDE-based cable neuron model into an ODE-based point-neuron model. This biophysical model introduces an additional synaptic integration current, originating from the nonlinear interaction among synaptic currents across spatial dendrites. Furthermore, our model adeptly replicates the somatic voltage response of neurons featuring complex dendritic structures, enabling intricate dendritic computations. We also derive a bilinear integration rule through asymptotic analysis and establish an uncomplicated yet effective mapping between a realistic neuron and an artificial neuron. This mapping attains state-of-the-art accuracy in modelling the somatic dynamics of real neurons while demanding less training time and a reduced number of parameters. Our work furnishes a systematic theoretical and computational framework for investigating the dynamics of spatial neurons.

Back to Table of Contents