Abstracts

Contents

Updated: 3 January 2025

Gianluca Ceruti
University of Innsbruck, Austria

Tree Tensor Network Operators for Long-Range Pairwise Interactions

---

In this contribution, we introduce a novel approach for representing and applying linear operators within the Tree Tensor Networks (TTNs) formalism, focusing on the specific challenges posed by long-range interactions in quantum many-body systems. Traditional methods often rely on matrix product operators, which can be limited in capturing the complexity of Hamiltonians with distant particle interactions. Building on foundational work by Lin, Tong, and others, we develop a framework that leverages the hierarchical low-rank structure of interaction matrices to construct Tree Tensor Network Operators (TTNOs) that are both compact and efficient. Our approach compresses the interaction matrix into a hierarchically semi-separable form, achieving significant reductions in computational complexity. Numerical experiments on quantum spin models illustrate the potential of TTNOs as a complementary tool to matrix product state techniques, representing a step forward in scalable quantum system simulations and other high-dimensional computational applications.

Back to Contents Page →

Shaoning Han
National University of Singapore, Singapore

Real-Time Solution of Mixed-Integer Quadratic Programs Using Decision Diagrams

We consider mixed-integer quadratic optimization problems with banded matrices and indicator variables. These problems arise pervasively in statistical inference problems with time-series data, where the banded matrix captures the temporal relationship of the underlying process. In particular, the problem studied arises in monitoring problems, where the decision-maker wants to detect changes or anomalies. We propose to solve these problems using decision diagrams. In particular we show how to exploit the temporal dependencies to construct diagrams with size polynomial in the number of decision variables. We also describe how to construct the convex hull of the set under study from the decision diagrams, and how to deploy the method online to solve the problems in milliseconds via a shortest path algorithm.

Back to Contents Page →

Anran Hu
Columbia University, USA

Optimization and Learning for Mean-Field Games via Occupation Measure

Understanding how large groups of agents interact is critical in fields like economics, finance, and engineering. However, directly analyzing N-player games becomes infeasible as the number of agents grows. Mean-field games (MFGs) provide a powerful solution by approximating these complex systems, allowing us to study the collective behavior of a large population as the limit of the N-player game when N is very large. This approximation significantly simplifies analysis while retaining key insights about the system.

In this talk, we present recent advancements at the intersection of MFGs and multi-agent reinforcement learning (MARL). First, we introduce Mean-Field Occupation Measure Optimization (MF-OMO), a novel framework that reinterprets Nash equilibria in discrete-time MFGs as a single optimization problem over occupation measures. This reformulation avoids restrictive assumptions often required by traditional methods and enables the use of standard optimization algorithms to find multiple equilibria. We also extend this approach to continuous-time finite-state MFGs, broadening its applicability.

Building on the concept of occupation measures, we then propose Mean-Field Occupation Measure Learning (MF-OML), the first fully polynomial online reinforcement learning algorithm capable of finding approximate Nash equilibria in large population games beyond variants of zero-sum and potential games. We establish regret bounds for the N-player games that can be approximated by MFGs under monotonicity conditions.

Together, these advancements provide a comprehensive approach to characterizing and solving Nash equilibria in complex multi-agent environments.

Back to Contents Page →

Qi Lei
New York University, USA

Efficient and Distribution-aware Model and Data Pruning

In the foundation model paradigm, pruning data and model sizes without compromising performance is essential for efficiency, energy, and memory savings. This talk presents two approaches to data and model pruning, each using distribution-aware metrics for efficient, single-shot selection of data samples or model components. In the first part, I introduce Sketchy Moment Matching (SkMM), a scalable data selection method for finetuning that balances variance and bias by matching the second moment of the original data distribution within a sketched low-dimensional space. SkMM enables fast, theoretically sound data selection while retaining critical distributional properties. In the second part, I discuss a structured pruning method for large language models that preserves model outputs. This method leverages a depth-2 pruning structure and input-distribution-aware metrics to achieve efficient model pruning without retraining. Together, these methods illustrate how distribution-aware metrics and greedy selection can yield effective pruning strategies that preserve model performance and reduce computational costs across various machine learning tasks.

Back to Contents Page →

Eitan Levin
California Institute of Technology, USA

Any-dimensional optimization

Optimization problems in many applications arise as sequences indexed by dimension. Indeed, graph problems are defined for graphs of all sizes, regularizers for inverse problems are defined for signals and images of any size, and (quantum) information problems are defined for distributions over any number of (qu)bits. We present a framework to systematically study such sequences of optimization problems, fit tractable data-driven relaxations, and derive lower bounds on their limiting values. We do so by observing that these problems are described in a “free” way, which makes it obvious how to instantiate them in any dimension. We explain how such free descriptions arise from the phenomenon of representation stability, and study their properties by considering the relations between problems across dimensions and their symmetries.

Back to Contents Page →

Cheuk Ting Li
The Chinese University of Hong Kong, Hong Kong SAR of China

Channel Simulation and Lossy Compression

Channel simulation concerns the setting where an encoder observes an input X, and would like to use the minimal amount of communication to a decoder in order to allow the decoder to output Y following a given conditional probability distribution P(Y—X). We discuss some recent developments on channel simulation schemes via rejection sampling, Poisson processes and dithered quantization, as well as their applications to tasks involving lossy compression and privacy.

Back to Contents Page →

Tianjiao Li
Georgia Institute of Technology, USA

Universal parameter-free methods for convex and nonconvex optimization

We present a novel class of projected gradient (PG) methods for minimizing a smooth but not necessarily convex function over a convex compact set. We first provide a novel analysis of the "vanilla" PG method, achieving the best-known iteration complexity for finding an approximate stationary point of the problem. We then develop an "auto-conditioned" projected gradient (AC-PG) variant that achieves the same iteration complexity without requiring the input of the Lipschitz constant of the gradient or any line search procedure. We then generalize the PG methods to the stochastic setting, by proposing a stochastic projected gradient (SPG) method and a variance-reduced stochastic gradient (VR-SPG) method, achieving new complexity bounds in different oracle settings. We also present auto-conditioned stepsize policies for both stochastic PG methods and establish comparable convergence guarantees.

Back to Contents Page →

Atsushi Nitanda
*A*STAR, Singapore*

Propagation of Chaos for Mean-field Neural Networks and its Application to Model Merging

---

Mean-field Langevin dynamics (MFLD) minimizes an entropy-regularized nonlinear convex functional defined over the space of probability distributions. MFLD has gained attention due to its connection with noisy gradient descent for mean-field two-layer neural networks. Unlike standard Langevin dynamics, the nonlinearity of the objective functional induces particle interactions, necessitating multiple particles to approximate its mean-field limit. Recent works have demonstrated the uniform-in-time propagation of chaos (PoC) for MFLD, showing that the gap between the particle system and the mean-field limit uniformly shrinks over time as the number of particles increases. In this talk, I'll talk about recent PoC results independent of logarithmic Sobolev inequality (LSI) constants which can exponentially deteriorate with the regularization coefficient. I'll also introduce the PoC-based model merging method for neural networks.

Back to Contents Page →

Kunlun Qi
University of Minnesota-Twin Cities, USA

On the Kinetic Description of Objective Molecular Dynamics (OMD): multiscale modeling, numerics, data-driven applications.

In the first part of this talk, a multiscale modeling framework for objective molecular dynamics (OMD), a reduced molecular dynamics approach with inherent symmetries, will be presented. This hierarchical framework bridges OMD with statistical kinetic equations and macroscopic hydrodynamic models. In the kinetic regime, we identify two distinct interaction scalings, leading to either a mean-field-type or Boltzmann-type equation. At the macroscopic level, we derive reduced Euler and Navier-Stokes systems through a detailed asymptotic analysis. The second part of the talk introduces a fast spectral method for numerically solving the derived kinetic equations, supported by convergence analysis and numerical simulations to confirm its effectiveness. If time permits, I will discuss our recent progress in applying data-driven methodologies to kinetic theory.

Back to Contents Page →

Andreas Søjmark
London School of Economics and Political Science, UK

Stochastic Stefan problems in finance: from contagion to supercooling

In this talk, I will discuss a stochastic version of the supercooled Stefan problem which appears naturally as a mean field description of a financial system with contagion and common exposures. Corresponding to a temporary break-down of this financial system, we will see that there is finite time blow-up with strictly positive probability if part of the initial profile exceeds a critical value. On the other hand, the system evolves continuously for all time when the initial profile is everywhere below this value. In the former case, we derive a global weak formulation that incorporates blow-ups, by allowing for jump discontinuities to appear while respecting conservation of energy. Finally, we single out a minimal solution whose discontinuities are shown to represent the natural resolution of emerging instabilities with respect to an infinitesimal external shock to the system.

Back to Contents Page →

Tang Tianyun
National University of Singapore, Singapore

Exploring chordal sparsity in semidefinite programming with sparse plus low-rank data matrices

Semidefinite programming (SDP) problems are challenging to solve because of their high dimensionality. However, solving sparse SDP problems with small tree-width are known to be relatively easier because: (1) they can be decomposed into smaller multi-block SDP problems through chordal conversion; (2) they have low-rank optimal solutions. In this paper, we study more general SDP problems whose coefficient matrices have sparse plus low-rank (SPLR) structure. We develop a unified framework to convert such problems into sparse SDP problems with bounded tree-width. Based on this, we derive rank bounds for SDP problems with SPLR structure, which are tight in the worst case.

Back to Contents Page →

Gabriele Visentin
ETH Zürich, Switzerland

Calibration to market-implied risk measures

In incomplete market models the interval of arbitrage-free prices for a contingent claim is typically very wide, with different prices corresponding to radically different hedging risks. The classical approach to model calibration consists in estimating a single martingale measure for the underlying asset by matching the observed mid prices of some calibration instruments (typically a selection of ATM call/put options) and using the resulting measure as a linear pricing rule for pricing new claims. This approach does not explicitly take into account market-implied risk preferences and, when used to price exotic or bespoke contingent claims, it may yield prices that are far from realistic market valuations and hedging positions that can only be liquidated at a loss.

In this work we present an alternative approach to model calibration, in which we directly estimate the market-implied risk measure that corresponds to the observed bid and ask prices of the calibration instruments. This approach is related to the literature on good deal bounds and is based on machine learning techniques, which combine deep hedging under convex risk measures with adversarial machine learning. We present empirical results on synthetic and real datasets and investigate the stability of the market-implied risk measures in time.

Back to Contents Page →

Guanyi Wang
National University of Singapore, Singapore

Solving Sparse \& High-Dimensional-Output Regression via Compression

Multi-Output Regression (MOR) has been widely used in scientific data analysis for decision-making. Unlike traditional regression models, MOR aims to simultaneously predict multiple real-valued outputs given an input. However, the increasing dimensionality of the outputs poses significant challenges regarding interpretability and computational scalability for modern MOR applications. As a first step to address these challenges, this paper proposes a Sparse & High-dimensional-Output REgression (SHORE) model by incorporating additional sparsity requirements to resolve the output interpretability, and then designs a computationally efficient twostage optimization framework capable of solving SHORE with provable accuracy via compression on outputs. Theoretically, we show that the proposed framework is computationally scalable while maintaining the same order of training loss and prediction loss before-and-after compression under arbitrary or relatively weak sample set conditions. Empirically, numerical results further validate the theoretical findings, showcasing the efficiency and accuracy of the proposed framework.

Back to Contents Page →

Peng Wang
University of Michigan, USA

Understanding Distribution Learning of Diffusion Models via Low-Dimensional Modeling.

Recent empirical studies have demonstrated that diffusion models can effectively learn the image distribution and generate new samples. Remarkably, these models can achieve this even with a small number of training samples despite a large image dimension, circumventing the curse of dimensionality. In this work, we provide theoretical insights into this phenomenon by leveraging key empirical observations: (i) the low intrinsic dimensionality of image datasets and (ii) the low-rank property of the denoising autoencoder in trained diffusion models. These observations motivate us to assume the underlying data distribution as a mixture of low-rank Gaussians and to parameterize the denoising autoencoder as a low-rank model. With these setups, we rigorously show that optimizing the training loss of diffusion models is equivalent to solving the canonical subspace clustering problem over the training samples. This insight carries practical implications for training and controlling diffusion models. Specifically, it allows us to characterize precisely the minimal number of samples necessary for learning correctly the low-rank data support, shedding light on the phase transition from memorization to generalization. Moreover, we empirically establish a correspondence between the subspaces and the semantic representations of image data, facilitating image editing. We validate these results with corroborated experimental results on both simulated distributions and image datasets.

Back to Contents Page →

Wang Wenjia
Hong Kong University of Science and Technology, China

Near-Optimal Regret Guarantee of CE Heuristic for Online Linear Programming

---

Certainty equivalent (CE) is a classical heuristic building on the intuition of replacing uncertainties with their average values, and making accept/reject decisions accordingly. Despite being the workhorse algorithm for online linear programming, the performance guarantee of CE measured by the additive regret against the prophet benchmark, remains largely unclear. Existing results typically rely on strong technical conditions called non-degeneracy conditions, which are not imposed on problem primitives, hard to intuit and very restrictive.

In this work, we show the (near) optimality of certainty equivalent (CE) for a fairly general class of instances, specified only by natural and mild assumptions on problem primitives, in particular, boundedness and smoothness on the (conditional) distribution of the reward. For centain class of instances, we prove that CE achieves an $O((\log T)^2)$ regret, which is near optimal (with an additional $\log T$ factor). En route characterizing the regret scaling, we borrow techniques such as the peeling device from the empirical processes and statistical theory, and establish concentration analysis of the solution to an optimization problem with large-size i.i.d. input, which may be of independent theoretical interest.

Back to Contents Page →

Marko Hans Weber
National University of Singapore, Singapore

Hedging: Holding Stocks, Trading Bonds

In an economy with random growth, several long-lived agents with heterogeneous risk-aversions, time-preferences, and income streams make consumption and investment decisions, trading stocks and a long-term bond, and borrowing from and lending to each other. We find in closed form equilibrium stock prices, interest rates, consumption, and trading policies. Agents do not trade stocks, although their returns are time-varying and predictable. Agents dynamically trade the long-term bond in response to growth shocks.

Back to Contents Page →

Renyuan Xu
New York University, USA

Generative diffusion models: optimization, generalization and fine-tuning

Recently, generative diffusion models have outperformed previous architectures, such as GANs, in generating high-quality synthetic data, setting a new standard for generative AI. A key component of these models is learning the associated Stein's score function. Though diffusion models have demonstrated practical success, their theoretical foundations are far from mature, especially regarding whether gradient-based algorithms can provably learn the score function. In this talk, I will present a suite of nonasymptotic theory aimed at understanding the data generation process in diffusion models and the accuracy of score estimation. Our analysis addresses both the optimization and generalization aspects of the learning process, establishing a novel connection to supervised learning and neural tangent kernels.

Building on these theoretical insights, another key challenge arises when fine-tuning pretrained diffusion models for specific tasks or datasets to improve performance. Fine-tuning requires refining the generated outputs based on particular conditions or human preferences while leveraging prior knowledge from the pre-trained model. In the second part of the talk, we formulate this fine-tuning as a stochastic control problem, establishing its well-definedness through the Dynamic Programming Principle and proving convergence for an iterative Bellman scheme.

This talk is based on joint works with Yinbin Han (NYU) and Meisam Razaviyayn (USC).

Back to Contents Page →

Liu Yang
National University of Singapore, Singapore

Towards Large Scientific Learning Models with In-Context Operator Networks

Can we build a single large model for a wide range of PDE-related scientific learning tasks? Can this model generalize to new PDEs without any fine-tuning? Incontext operator learning and the corresponding model In-Context Operator Networks (ICON) represent an exploration of these questions. We showed how a single ICON model (without fine-tuning) manages 19 distinct problem types, encompassing forward and inverse ODE, PDE, and mean-field control problems, provided with appropriately designed data prompts. We show positive evidence for the questions above, through a study on 1D scalar nonlinear conservation laws, a family of PDEs with temporal evolution. In particular, we show that an ICON model trained on conservation laws with cubic flux functions can generalize well to some other flux functions of more general forms, without fine-tuning.

Back to Contents Page →

Chulhee Yun
KAIST AI, Korea

Provable Benefit of Cutout and CutMix for Feature Learning

Patch-level data augmentation techniques such as Cutout and CutMix have demonstrated significant efficacy in enhancing the performance of image-based tasks. However, a comprehensive theoretical understanding of these methods remains elusive. In this talk, I will present about a recent NeurIPS 2024 paper that studies the underlying reasons behind the success of Cutout and CutMix through the lens of feature learning. In the paper, we study two-layer neural networks trained using three distinct methods: vanilla training without augmentation, Cutout training, and CutMix training. Our analysis focuses on a feature-noise data model, which consists of several labeldependent features of varying rarity and label-independent noises of differing strengths. Our theorems demonstrate that Cutout training can learn low-frequency features that vanilla training cannot, while CutMix training can even learn rarer features that Cutout cannot capture. From this, we establish that CutMix yields the highest test accuracy among the three. Notably, our novel analysis reveals that CutMix training makes the network learn all features and noise vectors "evenly" regardless of the rarity and strength, which provides an interesting insight into understanding patch-level augmentation.

Back to Contents Page →

Wei Zhang
Zuse Institute Berlin, Germany

Mathematical aspects of deep-learning techniques for identifying collective variables of molecular dynamics

---

High-dimensional metastable molecular dynamics can often be characterized by a few features of the system, i.e. collective variables (CVs). Thanks to the rapid advance in machine learning and deep learning, various data-driven CV identification methods have been developed in recent years, allowing for accurate modeling and efficient simulation of complex molecular systems. These methods can be broadly categorized into two groups. Methods in the first group employ autoencoders to minimize the reconstruction error and can be viewed as (static) dimensionality reduction methods, whereas methods in the second group compute functions (e.g. eigenfunctions or committor) that encode dynamical properties of the underlying system. In this talk, I will discuss the mathematics of deep learning-based methods in both groups mentioned above. For the methods in the second group, in particular, I will present a unified view from the perspective of constructing optimal effective dynamics (surrogate models). I will present concrete learning algorithms and illustrate them on numerical examples.

Back to Contents Page →