Robust Optimization and Simulation of Complex Stochastic Systems
Poster Session Abstracts
Saturday, September 14, 2024

Adversarially Robust Learning with OT-Regularized Divergences
Jeremiah Birrell, Texas State University

We introduce the $ARMOR_D$ methods as novel approaches to enhancing the adversarial robustness of deep learning models. These methods are based on a new class of optimal-transport-regularized divergences, constructed via an infimal convolution between an information divergence and an optimal-transport (OT) cost. We use these as tools to enhance adversarial robustness by maximizing the expected loss over a neighborhood of distributions, a technique known as distributionally robust optimization (DRO). Viewed as a tool for constructing adversarial samples, our method allows samples to be both transported, according to the OT cost, and re-weighted, according to the information divergence; the addition of a principled and dynamical adversarial re-weighting on top of adversarial sample transport is the key innovation of $ARMOR_D$. $ARMOR_D$ can be viewed as a generalization of the best-performing loss functions and OT costs in the adversarial training literature; we demonstrate this flexibility by using $ARMOR_D$ to augment the UDR, TRADES, and MART methods and obtain improved performance on CIFAR-10 image recognition.

Fractal opinions among interacting agents
Fei Cao, University of Massachusetts Amherst

We investigate an opinion model consisting of a large group of interacting agents, whose opinions are represented as numbers in [−1, 1]. At each update time, two random agents are selected, and the opinion of the first agent is updated based on the opinion of the second (the "persuader"). We derive the mean-field kinetic equation describing the large population limit of the model, and we provide several quantitative results establishing convergence to the unique equilibrium distribution. Surprisingly, in some range of the model parameters, the support of the equilibrium distribution exhibits a fractal structure. This provides a new mathematical description for the so-called opinion fragmentation phenomenon.

Symmetry-preserving machine learning
Ziyu Chen, University of Massachusetts Amherst

In this poster, I will present the task of learning a group-invariant distribution. The first part is devoted to the improved convergence of the empirical estimations of some types of probability divergences with known group symmetry. The second part showcase the application of the first part to group-invariant GANs in theory and with a simple numerical example.

Reinforcement Learning for the Optimal Dividend Problem
Thejani Gamage, University of Massachusetts, Amherst

We study the optimal dividend problem with the restricted dividend rate, both under the continuous time diffusion model and the well-known "Cramer-Lundberg" model. Unlike the standard literature, we shall particularly be interested in the case with unspecified model parameters, so that the optimal control cannot be explicitly determined. To approximate the optimal strategy, we use methods from the Reinforcement Learning (RL) literature, specifically, the method of solving the corresponding RL-type entropy-regularized exploratory control problem, which randomizes the control actions, and balances the levels of exploitation and exploration. We use a policy improvement argument, along with policy evaluation devices to construct approximating sequences of the optimal strategy. We present some numerical results using different parametrization families for the cost functional, to illustrate the effectiveness of the approximation schemes and to discuss possible methodologies to improve the effectiveness of Policy Evaluation methodologies.

Stochastic Modeling of Sieving and Adsorption in Membrane Filtration
Binan Gu, Worcester Polytechnic Institute

Membrane filtration of fluids is ubiquitous, and there is increasing interest from industrial practitioners in mathematical models that capture the complex nature of the process. Previous theoretical efforts include studies on material features of the membrane structure, and their connection to fluid mechanical properties and filtering efficiency. In this work, we model membrane filtration in a network of pores with simultaneous adsorption and sieving, the two fouling mechanisms typically observed during the early stages of commercial filtration applications. In comparison, first-principle partial differential equations model adsorptive fouling and species transport in the continuum in each pore, whereas sieving particles are assumed to follow a discrete Poisson arrival process. Our goal is to understand the individual influence of each fouling mode and highlight the effect of their coupling on the performance of pore-radius graded banded filters. Our results suggest that sieving alters the convexity of the flux decline due to the discrete nature of pore blockage and flux removal. Moreover, the difference between sieving particle sizes and the initial pore radius in each band plays a crucial role in indicating the onset and disappearance of sieving-adsorption competition. Lastly, we demonstrate a phase transition in filter lifetime as a function of sieving particle arrival frequency.

Combining Wasserstein-1 and Wasserstein-2 proximals: robust manifold learning via well-posed generative flows
Hyemin Gu, University of Massachusetts - Amherst

We formulate the formulation of well-posed continuous-time generative flows for learning distributions that are supported on low-dimensional manifolds through Wasserstein proximal regularizations of f-divergences. Wasserstein-1 proximal operators regularize f-divergences so that singular distributions can be compared. Meanwhile, Wasserstein-2 proximal operators regularize the paths of the generative flows by adding an optimal transport cost, i.e., a kinetic energy penalization. Via mean-field game theory, it is shown that the \emph{combination} of the two proximals is critical for formulating well-posed generative flows. Generative flows can be analyzed through optimality conditions of a mean-field game (MFG), a system of a backward Hamilton-Jacobi (HJ) and a forward continuity partial differential equations (PDEs) whose solution characterizes the optimal generative flow. For learning distributions that are supported on low-dimensional manifolds, the MFG theory shows that the Wasserstein-1 proximal, which addresses the HJ terminal condition, and the Wasserstein-2 proximal, which addresses the HJ dynamics, are both necessary for the corresponding backward-forward PDE system to be well-defined and have a unique solution with provably \emph{linear} flow trajectories. This implies that the corresponding generative flow is also unique and can therefore be learned in a robust manner even for learning high-dimensional distributions supported on low-dimensional manifolds. The generative flows are learned through \emph{adversarial training} of continuous-time flows, which bypasses the need for reverse simulation. Numerical experiments demonstrate the efficacy of our approach for generating high-dimensional images without the need of autoencoders or specialized architectures for the normalizing flow. Our work demonstrates how ideas from the analysis of multi-agent systems contribute to the design and understanding of generative flows.

Mimicking Theorems for Volterra Processes and Applications to Interacting Particle Systems
Kevin Hu, Brown University

Mimicking theorems allow one to write an Ito process as the solution to a stochastic differential equation. In this work we prove mimicking theorems for stochastic processes driven Volterra processes such as fractional Brownian motion and obtain quantitative propagation of chaos for mean-field systems driven by fractional noise. This is joint work with Kavita Ramanan and William Salkeld.

Control policies for sharing networks which achieve HGI performance
Dane Johnson, Elon University

Sharing networks are processing systems in which jobs can require simultaneous processing from multiple resources. This “sharing” feature means that control policies, which allocate resource capacity to job types with the goal of minimizing a linear holding cost, must account for “blocking” where resources are unable to utilize their full capacity despite the presence of jobs which require their processing for completion. The search for effective controls can be simplified by considering a sequence of networks under a diffusion scaling and focusing on asymptotic performance. This scaling leads to a Brownian control problem characterization of asymptotic performance, and one resulting (generally suboptimal) performance standard is the Hierarchical Greedy Ideal (HGI). HGI performance is characterized by avoiding unnecessary resource idle time to minimize resource workloads while simultaneously remaining in queue length configurations that provide the minimal holding cost for their corresponding workloads. We created simple form threshold control policies which achieve this HGI benchmark.

A Dynamic Likelihood Approach to Filtering for Advection-Diffusion Dynamics
Johannes Krotz, Notre Dame

A Bayesian data assimilation scheme is formulated for advection-dominated advective and diffusive evolutionary problems, based upon the Dynamic Likelihood (DLF) approach to filtering. The DLF was developed specifically for hyperbolic problems -waves-, and in this paper, it is extended via a split step formulation, to handle advection-diffusion problems. In the dynamic likelihood approach, observations and their statistics are used to propagate probabilities along characteristics, evolving the likelihood in time. The estimate posterior thus inherits phase information. For advection-diffusion the advective part of the time evolution is handled on the basis of observations alone, while the diffusive part is informed through the model as well as observations. We expect, and indeed show here, that in advection-dominated problems, the DLF approach produces better estimates than other assimilation approaches, particularly when the observations are sparse and have low uncertainty. The added computational expense of the method is cubic in the total number of observations over time, which is on the same order of magnitude as a standard Kalman filter and can be mitigated by bounding the number of forward propagated observations, discarding the least informative data.

Strong Data Processing Inequalities in Φ-Divergence and Φ-Mutual Information for the Langevin Dynamics and Algorithm
Siddharth Mitra, Yale University

We present bounds on the convergence of Φ-divergence (mixing time) and Φ-mutual information along popular Markov chains for continuous space sampling -- the Langevin dynamics in continuous time, and the unadjusted Langevin algorithm in discrete time. Our main approach is via strong data processing inequalities, and we show how to bound the contraction coefficients along these Markov chains, leading to fast convergence of the Φ-divergence and Φ-mutual information. We also briefly mention extensions to other Markov chains and the use of other techniques such as reverse transport inequalities.

The Stationary Behavior of a Diffusion Limit for Shortest Remaining Processing Time Queues
Amber Puha, CSU San Marcos

We characterize the stationary behavior of diffusion limits for shortest remaining processing time queues with heavy tailed processing time distributions. Here the diffusion limit is a measure valued process that arises as the limit of SRPT measure valued state descriptors that at each time put a unit mass at the remaining processing time of each job in system. The scaling for this limit is nonstandard and processing time distribution dependent, and the form of the diffusion limit depends on the tail behavior of the processing time distribution. For light tailed processing times, the measure valued diffusion limit is a point mass at one with the total mass fluctuating in accordance with the workload process diffusion limit (Ji and Puha (2024)). As such, the stationary behavior is straight forward to derive. In the case of heavier tailed processing time distributions, the diffusion limit is a measure valued process whose mass spreads out according to a random profile (Banerjee, Budhiraja and Puha (2022)). This random profile is determined by a random field W^* on $\mathbb{R}_+^2$, where the indices have a natural interpretation as time and space with the spatial variable being associated with the limiting rescaled remaining processing times. We determine the stationary behavior of W^* (i.e., the limit as time tends to infinity) by specifying the form of its finite dimensional joint distributions. We use this to explicitly compute the 2-dimensional joint distribution and covariance. Then, we leverage the stationary behavior of W^* to specify the form of the stationary measure valued diffusion limit, which in particular gives rise to the stationary limiting queue length. While we can compute the first couple of moments of stationary limiting queue length, its distribution appears to be complex in nature, which is in contrast to first-come-first-serve and many other service disciplines.

Joint Work with Sixian Jin (CSUSM) and Marvin Pena (CSUSM)

Control of Plasma through an external electric field
Yukun Yue, University of Wisconsin-Madison

Plasma instabilities are a major concern in plasma science, for applications ranging from particle accelerators to nuclear fusion reactors. In this work, we consider the possibility of controlling such instabilities by adding an external electric field to the Vlasov--Poisson equations. Our approach to determining the external electric field is based on conducting a linear analysis of the resulting equations. We show that it is possible to select external electric fields that completely suppress the plasma instabilities present in the system when the equilibrium distribution and the perturbation are known. In fact, the proposed strategy returns the plasma to its equilibrium at a rate that is faster than exponential in time if the Fourier transform of the initial data decays super-exponentially with respect to the Fourier variable corresponding to velocity. We further perform numerical simulations of the nonlinear two-stream and bump-on-tail instabilities to verify our theory and to compare the different strategies that we propose in this work.

Wasserstein proximal operators describe score-based generative models and resolve memorization
Benjamin Zhang, Brown University

We focus on the fundamental mathematical structure of score-based generative models (SGMs). We formulate SGMs in terms of the Wasserstein proximal operator (WPO) and demonstrate that, via mean-field games (MFGs), the WPO formulation reveals mathematical structure that describes the inductive bias of diffusion and score-based models. In particular, MFGs yield optimality conditions in the form of a pair of coupled PDEs: a forward-controlled Fokker-Planck (FP) equation, and a backward Hamilton-Jacobi-Bellman (HJB) equation. Via a Cole-Hopf transformation and taking advantage of the fact that the cross-entropy can be related to a linear functional of the density, we show that the HJB equation is an uncontrolled FP equation. Next, with the mathematical structure at hand, we present an interpretable kernel-based model for the score function which dramatically improves the performance of SGMs in terms of training samples and training time. The WPO-informed kernel model is explicitly constructed to avoid the recently studied memorization effects of score-based generative models. The mathematical form of the new kernel-based models in combination with the use of the terminal condition of the MFG reveals new explanations for the manifold learning and generalization properties of SGMs, and provides a resolution to their memorization effects. Our mathematically informed kernel-based model suggests new scalable bespoke neural network architectures for high-dimensional applications.