siam 2024

Searchable Abstracts Document

SIAM Conference on Mathematics of Data Science

(MDS24)

October 21–25, 2024

This document was current as of October 7, 2024. Abstracts appear as submitted.

siam

3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 U.S.
Telephone: 800-447-7426 (U.S. & Canada) +1-215-382-9800 (Worldwide)
meetings@siam.org

IP1

Solving Sparse Statistical Optimization Problems by Semismooth Newton Based Proximal Point Algorithms

Large-scale optimization problems arising from data science and statistics often look for optimal solutions with certain structured sparsity properties. In this talk, we shall introduce a dual semismooth Newton based proximal point algorithm (PPDNA) to solve such problems and explain how our method can be much more efficient than various first-order methods. The key idea is to make use of the second-order sparsity of the solutions, in addition to data sparsity, to make the per-iteration cost of our second-order method to be as low as that of first-order methods. We demonstrate that by incorporating the PPDNA within an adaptive sieving framework, we can efficiently generate the solution paths of large-scale problems corresponding to a sequence of regularization parameters. We shall illustrate the high efficiency of our approach on several popular models including convex clustering, lasso, and exclusive lasso.

Kim-Chuan Toh
National University of Singapore
mattohkc@nus.edu.sg

IP2

Some Foundations and Fairness in Machine Learning

In this talk, we will address several areas of recent work centered around the themes of foundational machine learning and fairness. We will discuss recent results involving linear algebraic tools for learning, such as methods in nonnegative matrix factorization, how we might think about fairness in these contexts, and approaches to mitigate bias. We will discuss new directions including an example in large-scale optimization that allows for population subgroups to have better predictors than when treated within the population as a whole. These methods allow for natural transparency and human interpretability while still offering strong performance. In addition, we will present recent theoretical results that govern various types of behaviors in neural nets, in particular overfitting and benign overfitting. Throughout the talk, we will include example applications from collaborations with community partners.

Deanna Needell
University of California, Los Angeles, U.S.
deanna@math.ucla.edu

IP3

The Data and Science of Elections

Social choice theory is built around the idea that different voting rules algorithms for aggregating many preferences into a single decision can lead to qualitatively different outcomes. Ill pay particular attention to the case of electing representatives, as we are about to do in the United States. By examining the goals and norms of representation, and developing new tools to describe and model the data, we can begin to approach electoral reform as a design problem for contemporary democracy.

Moon Duchin
Cornell University, U.S.
mduchin@cornell.edu

IP4

The Fiction Machine

Imagine a machine that can read a story and generate a meaningful continuation, that is, one that complies with the narrative demands of the story and makes sense to a meaningful fraction of the readers. Because what is true in the world of the story needs not be true in our world, this machine cannot be expected to say the truth. It only knows narrative necessity. This machine is of course an idealized model of modern AI systems, from language models and chatbots to movie generation. It is also an opportunity to formulate important questions and sometimes catch a glimpse of their answers. How can we define such a machine more rigorously? What can it compute? How does it compare to logic and mathematical reasoning? Can it be used to make inferences about our world even though its output is not constrained by what is true in our world? Such questions are not only relevant to artificial intelligence, but also useful to understand certain aspects of human intelligence and society.

Léon Bottou
Meta, U.S.
leonb@meta.com

IP5

On Parameterizing Optimal Transport with Elastic Costs

I will present in this talk an overview of the computations of optimal transport, focusing in particular on the challenge of computing OT maps using two samples from high-dimensional probability measures. After reviewing a few of the popular methods that have been explored for this task recently, including those leveraging neural architectures, I will introduce our recent work on parameterising OT problems with elastic costs, i.e. ground costs that mix the classic squared Euclidean distance with a regularizer (e.g. L1 norm). After highlighting the properties of OT maps that follow such costs, I will present a method to compute ground truth OT maps with elastic costs and also a method to learn the parameters, adaptively, of such a regularizer.

Marco Cuturi
Apple, France
cuturi@apple.com

IP6

Smartphone Privacy: How to Learn from Distributed, Private Data

Many of us use smartphones and rely on tools like autocomplete and spelling auto-correct to make using these devices more pleasant, but building these tools creates a conundrum. On the one hand, the machine-learning algorithms used to provide these features require data to learn from, but on the other hand, who among us is willing to send a carbon copy of all our text messages to device manufacturers and software vendors to provide that data? In this talk, we show a surprising paradox discovered roughly 20 years ago: it is possible to learn from user data in the aggregate, while mathematically provably maintaining privacy at the individual level. We then discuss some recent

algorithms that achieve this.

Jelani Nelson
University of California, Berkeley
minilek@berkeley.edu

IP7

The Emerging Science of Benchmarks

Benchmarks are the keystone that hold the machine learning community together. Growing as a research paradigm since the 1980s, there's much we've done with them, but little we know about them. In this talk, I will trace the rudiments of an emerging science of benchmarks through selected empirical and theoretical observations. Specifically, we'll discuss the role of annotator errors, external validity of model rankings, and the promise of multi-task benchmarks. The results in each case challenge conventional wisdom and underscore the benefits of developing a science of benchmarks.

Moritz Hardt
Max Planck Institute for Intelligent Systems
hardt@is.mpg.de

IP8

Tensor Decomposition meets Reproducing Kernel Hilbert Spaces (RKHS)

Tensor decompositions require that data live on a regular d-way grid, but many real-world datasets do not have this property. For example, time-evolving data may be measured at different intervals for different subjects and adaptive meshes in simulations are irregular by design. We can handle irregular grids by treating some modes as infinite-dimension rather than finite-dimensional; we refer to such tensors as quasi-tensors. For their decompositions, this means that we want the factors in the tensor decomposition to be smooth functions rather than vectors. This basic idea has appeared in myriad forms over the years, often using different terminology and with different applications. I will recall and build on these efforts. The result is a generic framework for incorporating continuous modes into the CP tensor decomposition. We focus on learning the infinite-dimensional modes from a reproducing kernel Hilbert space (RKHS) and present an alternating least squares algorithm that is computationally efficient. Including infinite-dimensional modes (1) enables practitioners to enforce common structural assumptions about data such as smoothness, (2) extends to situations in where the measurement times do not align by utilizing the framework of missing data, and (3) provides a more principled way to interpolate between observed points.

Tamara G. Kolda
MathSci.ai
tammy.kolda@mathsci.ai

SP1

2024 SIAG/DATA Career Prize - Mathematics in Scientific Machine Learning

Artificial intelligence (AI) and machine learning (ML) are poised to revolutionize the pace and nature of scientific discovery. The widespread adoption of AI in the sciences has the potential to integrate scientific inquiry with modes of hypothesis generation, data analysis, experimental design, and simulation, transforming our capacity to address scientific problems that currently seem insurmountable. The mathematical foundations of AI and ML are crucial for high-quality, reproducible, AI-enabled scientific research. However, blindly applying AI and ML poses significant risks, such as the rapid acceleration of the reproducibility crisis in science. In this talk, I will discuss fundamental machine learning challenges and opportunities that are particularly relevant to scientific discovery, such as emulators, generative models, and inverse problems. These problems underscore the importance of incorporating mathematical and physical models as well as numerical algorithms into ML frameworks, highlighting exciting directions for future work.

Rebecca Willett
The University of Chicago
willett@uchicago.edu

SP2

2024 SIAG/DATA Early Career Prize: Learning Matchings, Maps, and Trajectories

This talk will survey some recent advances in the statistical theory of optimal transport. Optimal transport provides a geometrical theory to describe transformations of probability distributions, making it a suitable framework for applications from machine learning to high-energy physics. We will study statistical estimators for this problem, characterizing their finite-sample behavior and obtaining distributional limits suitable for practical inference. Additionally, we will explore structural assumptions that improve the statistical and computational performance of these estimators in high dimensions.

Jonathan Niles-Weed
New York University, U.S.
jnw@cims.nyu.edu

MS1

Data-Driven Higher Order Differential Equations Inspired Graph Neural Networks

A recent innovation in Graph Neural Networks (GNNs) is the family of Differential Equation-Inspired Graph Neural Networks (DE-GNNs), which leverage principles from continuous dynamical systems to model information flow on graphs with built-in properties. However, existing DE-GNNs rely on first or second-order temporal orders. In this paper, we propose a neural extension to those pre-defined temporal dependencies. We show that our model, called TDE-GNN, can capture a wide range of temporal dynamics that go beyond typical first or second-order methods, and provide use cases where existing temporal models are challenged. We demonstrate the benefit of learning the temporal dependencies using our method rather than using pre-defined temporal dynamics on several graph benchmarks.

Moshe Eliasof
University of Cambridge
eliasof@post.bgu.ac.il

Eldad Haber
Department of Mathematics
The University of British Columbia
haber@eoas.ubc.ca

Eran Treister
Ben Gurion University of the Negev
erant@cs.bgu.ac.il

Carola-Bibiane Schönlieb
University of Cambridge
cbs31@cam.ac.uk

MS1

Structure-Preserving Solutions of Hamiltonian Systems Based on Neural Networks

The Hamiltonian formalism allows describing several dynamical systems with a conserved energy function. One of the primary additional properties of Hamiltonian systems is that they conserve a suitable symplectic form. This talk focuses on Hamiltonian systems in the linear space $\mathbb{R}^{2n}$ equipped with its canonical symplectic structure. A major challenge in developing numerical methods to approximate the solution of Hamiltonian systems, is making them preserve both the Hamiltonian energy and the symplectic structure. We develop a neural network-based approach which provides a piecewise-smooth approximation of the solutions of the Hamiltonian differential equation $\dot{x}(t) = X_H(x(t))$. Further, we propose a methodology to constrain the approximate solution to be a symplectic map while also promoting the conservation of the Hamiltonian energy $x \mapsto H(x)$. Thanks to the formalism in [Galley C.R., "Classical mechanics of nonconservative systems.", 2013], allowing us to express Hamiltonian systems with external forcing terms in $\mathbb{R}^{2n}$ as conservative ones in $\mathbb{R}^{4n}$, we also explore forced Hamiltonian systems with this strategy.

Davide Murari
Norwegian University of Science and Technology
davide.murari@ntnu.no

Priscilla Canizares-Martinez, Ferdia Sherry, Zakhar Shumaylov
University of Cambridge
pc464@cam.ac.uk, fs436@cam.ac.uk, zs334@cam.ac.uk

MS1

CoLoRA: Continuous Low-Rank Adaptation for Reduced Implicit Neural Modeling of Parameterized Partial Differential Equations

This work introduces reduced models based on Continuous Low Rank Adaptation (CoLoRA) that pre-train neural networks for a given partial differential equation and then continuously adapt low-rank weights in time to rapidly predict the evolution of solution fields at new physics parameters and new initial conditions. The adaptation can be either purely data-driven or via an equation-driven variational approach that provides Galerkin-optimal approximations. Because CoLoRA approximates solution fields locally in time, the rank of the weights can be kept small, which means that only few training trajectories are required offline so that CoLoRA is well suited for data-scarce regimes. Predictions with CoLoRA are orders of magnitude faster than with classical methods and their accuracy and parameter efficiency is higher compared to other neural network approaches.

Jules Berman
NYU
jmb1174@nyu.edu

Benjamin Peherstorfer
Courant Institute of Mathematical Sciences
New York University
pehersto@cims.nyu.edu

MS1

Data-driven Computation of Spectral Properties of Koopman Operators

Koopman operators are infinite-dimensional operators that globally linearize nonlinear dynamical systems, making their spectral information valuable for understanding dynamics. However, Koopman operators can have continuous spectra, can lack finite-dimensional invariant subspaces, and approximations can suffer from spectral pollution (spurious modes). These issues make computing the spectral properties of Koopman operators a considerable challenge. In this talk we will detail ResDMD, which rigorously computes the spectra and pseudospectra of general Koopman operators from snapshot data.

Alex Townsend
Cornell University
townsend@cornell.edu

Matthew Colbrook
University of Cambridge
mjc249@cam.ac.uk

MS2

Disentangled Adversarial Flow (DAF) for Multi-Source Network Learning with Applications in Brain Connectomes

Neuroimaging studies, such as the Adolescent Brain Cognitive Development (ABCD) and Human Connectome Project (HCP), have collected vast amounts of brain imaging data to better understand brain structures and functions. However, heterogeneous data collection methods across studies and sites, combined with limited samples in specialized cohorts like the Alzheimer's Disease Neuroimaging Initiative (ADNI), pose significant challenges when attempting collect multiple studies to enhance prediction and inference for small-scaled data. This work introduces a novel approach called Disentangled Adversarial Flow (DAF) to leverage information from large-scale, multi-source datasets and improve predictive accuracy in smaller-scale neuroimaging studies. DAF is a flow-based generative model that generates domain-invariant representations of brain networks while preserving their essential features. It employs a bidirectional-generative architecture and a kernel-based dependence measure to quantify and minimize the dependence between brain networks and their associated domain labels. Additionally, we introduce an ensemble-based DAF regression framework that utilizes a weighted, data-adaptive approach to integrate information from multiple large-scale source datasets, effectively mitigating information loss when dealing with multi-domain data. We applied the proposed method to study brain structural connectivity and its relationship with cognition in the ABCD, HCP, and ADNI datasets.

Yixin Chen
Virginia Tech
yixinc@vt.edu

Zhengwu Zhang
UNC Chapel Hill
zhengwu_zhang@unc.edu

Xin Xing
Virginia Tech

xinxing@vt.edu

Meimei Liu
Virginia Polytechnic Institute and State University
meimeiliu.stat@gmail.com

MS2

Fast Variational Inference of Latent Space Models for Dynamic Networks Using Bayesian P-Splines

Latent space models (LSMs) are often used to analyze dynamic (time-varying) networks that evolve in continuous time. Existing approaches to Bayesian inference for these models rely on Markov chain Monte Carlo algorithms, which cannot handle modern large-scale networks. To overcome this limitation, we introduce a new prior for continuous-time LSMs based on Bayesian P-splines that allows the posterior to adapt to the dimension of the latent space and the temporal variation in each latent position. We propose a stochastic variational inference algorithm to estimate the model parameters. We use stochastic optimization to subsample both dyads and observed time points to design a fast algorithm that is linear in the number of edges in the dynamic network. Furthermore, we establish non-asymptotic error bounds for point estimates derived from the variational posterior. To our knowledge, this is the first such result for Bayesian estimators of continuous-time LSMs. Lastly, we use the method to analyze a large data set of international conflicts consisting of 4,456,095 relations from 2018 to 2022.

Joshua Loyal
Florida State University
jloyal@fsu.edu

MS2

Bias-Corrected Joint Spectral Embedding for Multilayer Networks with Invariant Subspace: Entrywise Eigenvector Perturbation and Inference

In this talk, I will introduce a novel bias-corrected joint spectral embedding algorithm to estimate the invariant subspace across heterogeneous multiple networks. The proposed algorithm recursively calibrates the diagonal bias of the sum of squared network adjacency matrices by leveraging the closed-form bias formula and iteratively updates the subspace estimator using the most recent estimated bias. Correspondingly, we establish a complete recipe for the entrywise subspace estimation theory for the proposed algorithm, including a sharp entrywise subspace perturbation bound and the entrywise eigenvector central limit theorem. Leveraging these results, we settle two multiple network inference problems: the exact community detection in multilayer stochastic block models and the hypothesis testing of the equality of membership profiles in multilayer mixed membership models. Our proof relies on delicate leave-one-out and leave-two-out analyses that are specifically tailored to block-wise symmetric random matrices and a martingale argument that is of fundamental interest for the entrywise eigenvector central limit theorem.

Fangzheng Xie
Indianna University
fxie@iu.edu

MS2

Active Learning on Network via Sequential Biased Sampling

We consider the semi-supervised learning on the network that infers the response of nodes on the network by actively querying the responses from a small proportion of nodes. The problem is important in many real-world network applications where the nodes responses are expensive and scarce. In this paper, we propose an off-line sequential biased sampling method to select the nodes whose responses are informative in predicting responses on remaining nodes via the response dependency among network and node covariates. The proposed method consists of a two-stage sampling scheme to take both informativeness and representativeness into consideration in node selection.Informativeness refers to the level of response complexity that can be learned by the queried nodes, and representativeness refers to the controllability of prediction error resulting from the noises in queried responses. To achieve above two goals, our approach builds on graph spectral theory and random spectral sparsification, which also provides a theoretical framework to illustrate trade-off between informativeness and representativeness in active learning on network, akin to classical bias-variance trade-off. The proposed method is provably efficient in response prediction and robust to both covariates and response noise. The numerical experiments show that the proposed method outperforms existing network active learning methods in both regression and classification tasks. Keyword: network optimal design, graph fourier analysis, graph semi-supervised learning, spectral sparsification

Yubai Yuan
Pennsylvania State University
yvy5509@psu.edu

MS3

Nonlinear Generative Flows for Enhanced Learning of Structured Distributions

I will present a novel method for training score-based generative models which uses nonlinear noising dynamics and a variance-reduced objective function to improve learning of structured distributions. Generalizing to a nonlinear drift allows for additional structure to be incorporated into the dynamics, thus making the training better adapted to the data, e.g., in the case of multimodality or (approximate) symmetries. I will demonstrate that this method is capable of learning score-based generative models with less data and better class-balance by flexibly incorporating structure arising from the dataset.

Jeremiah Birrell
Texas State University
jbirrell@email.arizona.edu

Markos Katsoulakis, Luc Rey-Bellet, Benjamin J. Zhang
University of Massachusetts Amherst
markos@umass.edu, luc@math.umass.edu,
bjzhang@umass.edu

Wei Zhu
UMass Amherst
weizhu@gatech.edu

MS3

Equivariant Score-based Generative Models Provably Learn Distributions with Symmetries Effi-

ciently

Symmetry is ubiquitous in many real-world phenomena and tasks, such as physics, images, and molecular simulations. Empirical studies have demonstrated that incorporating symmetries into generative models can provide better generalization and sampling efficiency when the underlying data distribution has group symmetry. In this work, we provide the first theoretical analysis and guarantees of score-based generative models for learning distributions that are invariant with respect to some group symmetry. First, we rigorously demonstrate that one can learn the score of a symmetrized distribution using equivariant vector fields without data augmentations through the analysis of the optimality and equivalence of score-matching objectives. This also provides practical guidance that one does not have to augment the dataset as long as the vector field or the neural network is equivariant. Second, building on recent work on the Wasserstein-1 (W1) guarantees of SGMs and empirical estimations of probability divergences under group symmetry, we provide an improved generalization bound when the data distribution is group-invariant. Conversely, we quantify the impact of not incorporating the equivariant structure into the score parametrization, by showing that non-equivariant vector fields can lead to a worse generalization bound. This can be viewed as a type of model-form error that describes the missing structure of non-equivariant vector fields. Numerical simulations verify our theories and highlight that data augmentations cannot replace the role of equivariant vector fields.

Ziyu Chen, Markos Katsoulakis
University of Massachusetts Amherst
ziyuchen@umass.edu, markos@umass.edu

Benjamin Zhang
University of Massachusetts Amherst
U.S
bjzhang@umass.edu

MS3

Wasserstein Proximal Operators Describe Score-Based Generative Models and Resolve Memorization

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.

Siting Liu
University of California, Los Angeles, U.S.
siting6@math.ucla.edu

Benjamin J. Zhang
University of Massachusetts Amherst
bjzhang@umass.edu

Markos A. Katsoulakis
University of Massachusetts, Amherst
Dept of Mathematics and Statistics
markos@math.umass.edu

Stanley J. Osher
UCLA
sjo@math.ucla.edu

Wuchen Li
University of South Carolina
wuchen@mailbox.sc.edu

MS3

Multilevel Diffusion Models

Score-based diffusion models (SBDM) have recently emerged as state-of-the-art approaches for image generation. We develop SBDMs in the infinite-dimensional setting, that is, we model the training data as functions supported on a rectangular domain. Besides the quest for generating images at ever higher resolution, our primary motivation is to create a well-posed infinite-dimensional learning problem so that we can discretize it consistently at multiple resolution levels. We demonstrate how to overcome two shortcomings of current SBDM approaches in the infinite-dimensional setting by ensuring the well-posedness of forward and reverse processes, and derive the convergence of the approximation of multilevel training. We illustrate that approximating the score function with an operator network is beneficial for multilevel training.

Nicole Yang, Ruthotto Lars
Emory University
tianjiao.yang@emory.edu, lruthotto@emory.edu

Paul Hagemann, Gabriele Steidl, Sophie Mildenberger
TU Berlin
hagemann@math.tu-berlin.de, steidl@math.tu-berlin.de, mildenberger@tu-berlin.de

MS4

Analytic Formulas for Marginal Feature Attributions of Oblivious Decision Trees

Marginal feature attributions of any decision tree are simple function on a certain grid partition determined by the splits in the tree. We show that when the tree is oblivious (symmetric), an explicit formula in terms of internal model parameters can be derived for the marginal Shapley values at any leaf of the tree. This can be generalized to a host of other game values (e.g. Banzhaf) which we shall axiomatically characterize as well as to certain coalitional game values (e.g. the Owen value). This results in an inherently interpretable method for explaining ensembles of oblivious decision trees (e.g. CatBoost models). Finally, we point out that our results on symmetric trees can in principle

yield analytic formulas for marginal feature attributions of any decision tree.

Khashayar Filom
Discover Financial Services
khashayar.1367@gmail.com

Alexey Miroshnikov
Discover Financial Services
U.S
amiroshn@terpmail.umd.edu

Konstandinous Kotsiopoulos, Arjun Ravi Kannan
Discover Financial Services
kkotsiop@gmail.com, arjun.kannan@gmail.com

MS4

Neural Differential Equations for Medical Image Prediction and Segmentation

In this talk, we discuss recent progress in incorporating ODEs and PDEs in medical image prediction and segmentation. First, we present a PDE-guided deep learning framework to learn the underlying tumor cell dynamics influenced by radiotherapy. A two-branch neural network is designed to encode a reaction-diffusion equation with an unknown operator approximated by a neural network. Starting from pre-treatment PET images and radiation dose distributions, this model shows promising results in predicting the post-treatment PET images and the influence of the imposed radiotherapy. Second, we propose a Neural-ODE based method for interpreting the behavior of neural networks in multi-parametric medical image segmentation tasks. We characterize the continuous evolution of images with multi-modality from inputs to segmentation results using Neural ODEs. We also design an accumulative contribution curve to quantify the utilization of each modality in the learned dynamics. In a multi-parametric MRI-based glioma segmentation study, the proposed method successfully identifies key MR modalities. This method offers a new tool for optimizing inputs and enhancing the interpretability of deep learning models for multimodal image segmentation.

Hangjie Ji
North Carolina State University
hangjie_ji@ncsu.edu

Zhenyu Yang, Kyle Lafata, Scott Floyd, Fang-fang Yin, Chunhao Wang
Duke University
zhenyu.yang893@duke.edu, kyle.lafata@duke.edu, scott.floyd@duke.edu, fangfang.yin@duke.edu, chunhao.wang@duke.edu

MS4

Stability Theory of Game-Theoretic Group Feature Explanations for Machine Learning Models

In this work, we study feature attributions of Machine Learning (ML) models originating from linear game values and coalitional values defined as operators on appropriate functional spaces. The main focus is on random games based on the conditional and marginal expectations. The first part of our work formulates a stability theory for these explanation operators by establishing certain bounds for both marginal and conditional explanations. The differences between the two games are then elucidated, such as showing that the marginal explanations can become discontinuous on some naturally-designed domains, while the conditional explanations remain stable. In the second part of our work, group explanation methodologies are devised based on game values with coalition structure, where the features are grouped based on dependencies. We show analytically that grouping features this way has a stabilizing effect on the marginal operator on both group and individual levels, and allows for the unification of marginal and conditional explanations. Our results are verified in a number of numerical experiments where an information-theoretic measure of dependence is used for grouping.

Alexey Miroshnikov
Discover Financial Services
U.S
amiroshn@terpmail.umd.edu

Konstandinous Kotsiopoulos, Khashayar Filom, Arjun Ravi Kannan
Discover Financial Services
kkotsiop@gmail.com, khashayar.1367@gmail.com, arjun.kannan@gmail.com

MS4

Formal Interpretability for Machine Learning

Recent works have extended notions of feature importance to semantic concepts that are inherently interpretable to the users interacting with a black-box predictive model. Yet, precise statistical guarantees, such as false positive rate control, are needed to communicate findings transparently and to avoid unintended consequences in real-world scenarios. In this talk, we will formalize the global (i.e., over a population) and local (i.e., for a sample) statistical importance of semantic concepts for the predictions of opaque models, by means of conditional independence, which allows for rigorous testing. We use recent ideas of sequential kernelized testing (SKIT) to induce a rank of importance across concepts, and showcase the effectiveness and flexibility of our framework on synthetic datasets as well as on image classification tasks using vision-language models such as CLIP.

Jeremias Sulam
Johns Hopkins University
jsulam1@jhu.edu

MS5

Derivative-informed Neural Operator Acceleration of Geometric MCMC for Infinite-dimensional Bayesian Inverse Problems

We propose an operator learning approach to accelerate geometric Markov chain Monte Carlo (MCMC) for solving infinite-dimensional Bayesian inverse problems (BIPs). While geometric MCMC employs high-quality proposals that adapt to posterior local geometry, it requires repeated computations of gradients and Hessians of the log-likelihood, which becomes prohibitive when the parameter-to-observable (PtO) map is defined through large-scale parametric partial differential equations (PDEs). We consider a delayed-acceptance geometric MCMC method driven by a neural operator surrogate of the PtO map, where the proposal exploits fast surrogate approximations of the log-likelihood and, simultaneously, its gradient and Hessian. In this work, we present an extension of derivative-informed operator learning using samples of the PtO map and its parametric derivative. This leads to derivative-informed neural operator (DINO) surrogates

that accurately approximate the PtO map and its parametric derivative over the prior distribution at a significantly lower training cost than conventional methods. Numerical studies demonstrate that DINO-driven MCMC generates effective posterior samples 3–9 times faster than geometric MCMC and 60–97 times faster than prior geometry-based MCMC. Furthermore, the training cost of DINO surrogates breaks even compared to geometric MCMC after just 10–25 effective posterior samples.

Lianghao Cao
California Institute of Technology
lianghao@caltech.edu

Thomas O'Leary-Roseberry
The University of Texas at Austin
Oden Institute for Computational and Engineering Sciences
tom@oden.utexas.edu

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

MS5

Divide and Conquer - Improved Training Of Neural Ordinary Differential Equations for Chaotic Dynamical Systems

Forecasting high-dimensional dynamical systems is a fundamental challenge in various fields of science and engineering. Neural Ordinary Differential Equations (NODEs), which combine the power of neural networks and numerical solvers, have emerged as promising tools for forecasting complex nonlinear dynamical systems. However, conventional optimization methods become inefficient for chaotic dynamical systems, as the loss functional becomes highly non-convex with exploding gradients in time. To address this challenge, we propose a training algorithm that spilts the time domain into multiple steps with penalty against intermediate discontinuities. The proposed algorithm, denoted the Multistep Penalty Neural ODE (MP-NODE), is applied to chaotic systems such as the Lorenz, Kuramoto-Sivashinsky equation, and Kolmogorov Flow. It is observed that MP-NODE converges to a better approximation of the physical system for the same training time than standard NODEs. Furthermore, the training cost of MP algorithm is significantly smaller than that of the least-square shadowing method.

Dibyajyoti Chakraborty
Pennsylvania State University
d.chakraborty@psu.edu

Seung Whan Chung
Lawrence Livermore National Laboratory
chung28@llnl.gov

Romit Maulik
Pennsylvania State University
rmaulik@psu.edu

MS5

Fokker–Planck-Based Inverse Reinforcement Learning

Inverse Reinforcement Learning (IRL) is a compelling technique for revealing the rationale underlying the behavior of autonomous agents. IRL seeks to estimate the unknown reward function of a Markov decision process (MDP) from observed agent trajectories. While most IRL approaches require the transition function to be prescribed or learned a-priori, we present a new IRL method targeting the class of MDPs that follow the Ito dynamics without this requirement. Instead, the transition is inferred in a physics-constrained manner simultaneously with the reward functions from observed trajectories leveraging the mean-field theory described by the Fokker-Planck (FP) equation. We conjecture an isomorphism between the time-discrete FP and MDP that extends beyond the minimization of free energy (in FP) and maximization of the reward (in MDP). This isomorphism allows us to infer the potential function in FP using variational system identification, which consequently allows the evaluation of reward, transition, and policy by leveraging the conjecture. We demonstrate the effectiveness of FP-IRL by applying it to synthetic benchmarks and a biological problem of cancer cell dynamics, where the transition function is unknown.

Chengyang Huang, Siddhartha Srivastava, Xun Huan
University of Michigan
chengyah@umich.edu, sidsriva@umich.edu, xhuan@umich.edu

Krishna Garikipati
University of Southern California
garikipa@usc.edu

MS5

Neural Network Approaches for High Dimensional Parameteric Optimal Control

We present numerical approaches for deterministic, high-dimensional optimal control problems whose dynamics depend on an unknown or uncertain parameter. The objective is to amortize the solution over a set of relevant parameters in an offline stage to enable rapid decision-making and be able to react to changes in the parameter in the online stage. To tackle the curse of dimensionality arising when the state and or parameter dimension are high-dimensional, we represent the policy using neural networks. We compare two training paradigms: First, our model-driven approach leverages the dynamics and definition of the objective function to learn the value function of the parameterized optimal control problem and obtain the policy using a feedback form. Second, we use actor-critic reinforcement learning to approximate the policy in a data-driven way. Through a two-dimensional convection diffusion equation, featuring high-dimensional state and parameter spaces, we investigate the accuracy, efficiency, and scalability of both training paradigms. While both paradigms lead to a reasonable approximation of the policy, the model-driven approach is more accurate and reduces the number of PDE solves significantly.

Deepanshu Verma
Department of Mathematics
Emory University
dverma@clemson.edu

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

Bart Waanders
Sandia National Labs

bartv@sandia.gov

Nick Winovich
Sandia National Laboratories
nwinovi@sandia.gov

MS7

Variationally Consistent Hamiltonian Model Reduction

Hamiltonian systems offer a simple description of conservative dynamics which presents many challenges for model reduction, particularly when canonical position and momentum variables are separated in scale by orders of magnitude. This talk presents a novel and variationally consistent method for the model reduction of canonical Hamiltonian systems. Its distinguishing factors are (1) its ability to accommodate nearly arbitrary reduced bases, (2) its applicability in both intrusive and nonintrusive settings, and (3) its interpretable error estimate involving a projection term and a deviation-from-canonicity term, both of which must balance for accurate state approximation. Results are presented using examples from 3D solid mechanics, showing that the proposed method offers several advantages when compared to the existing state of the art.

Anthony Gruber, Irina K. Tezaur
Sandia National Laboratories
adgrube@sandia.gov, ikalash@sandia.gov

MS7

Data-Driven Estimation of Stability Guarantees for Nonlinear Dynamical Systems

Analyzing the stability of a nonlinear dynamical system is central to understanding system behavior and designing controllers, and we use a Lyapunov function for this purpose. It is possible to guarantee the stability of a system if one can find a Lyapunov function that is positive definite and decreasing over time along the orbit of the system, thus providing a sufficient condition for stability. A Lyapunov function also characterizes an estimate of the domain of attraction, which indicates the region under which the system states asymptotically converge to equilibrium. The construction of a Lyapunov function is done analytically and ad hoc for certain nonlinear systems. However, doing so for systems with high nonlinearities and different dimensions is a challenging task. To address this problem, we present a data-driven method for discovering Lyapunov functions, called Lyapunov function Inference (LyapInf). This new method fits a quadratic Lyapunov function to the state trajectory data of the dynamics via optimization, where the process of inferring a Lyapunov function is based on the non-intrusive model reduction method of Operator Inference. This method learns one of many possible Lyapunov functions that ensures stability and estimates the domain of attraction with or without access to the system model. In this work, we demonstrate this new method on several numerical examples.

Tomoki Koike, Elizabeth Qian
Georgia Institute of Technology
tkoike3@gatech.edu, eqian@gatech.edu

MS7

Greedy Construction of Quadratic Manifolds for Nonlinear Dimensionality Reduction and Nonlinear Model Reduction.

Dimensionality reduction on quadratic manifolds augments linear approximations with quadratic correction terms. Previous works rely on linear approximations given by projections onto the first few leading principal components of the training data; however, linear approximations in subspaces spanned by the leading principal components alone can miss information that are necessary for the quadratic correction terms to be efficient. In this presentation, we propose a greedy method that constructs subspaces from leading as well as later principal components so that the corresponding linear approximations can be corrected most efficiently with quadratic terms. Properties of the greedily constructed manifolds allow applying linear algebra reformulations so that the greedy method scales to data points with millions of dimensions. Numerical experiments demonstrate that an orders of magnitude higher accuracy is achieved with the greedily constructed quadratic manifolds compared to manifolds that are based on the leading principal components alone.

Paul Schwerdtner
New York University
ps5136@nyu.edu

Benjamin Peherstorfer
Courant Institute of Mathematical Sciences
New York University
pehersto@cims.nyu.edu

MS7

Learning Mechanical Systems Via a Structured AAA Algorithm

Data-driven reduced-order modeling is an essential tool in constructing high-fidelity compact models to approximate physical phenomena when explicit models, such as state-space formulations with access to internal variables, are not available yet abundant input/output data are. A popular method for this endeavor is the AAA (Adaptive Antoulas-Anderson) algorithm that constructs rational approximations by harmoniously blending interpolation and least squares (LS) fitting on a data set composed of frequency-response measurements. When derived from first principles, models typically contain differential structures that are essential for certain system properties and physical interpretation. Such structures are, for example, second-order time derivatives in the modeling of mechanical and electro-mechanical processes. However, these differential structures are typically lost when creating models from data. Based on our work on structured barycentric forms for systems with second-order structure, we propose an extension of the AAA algorithm that allows to efficiently learn dynamical systems of mechanical processes from data.

Steffen W. R. Werner
Virginia Tech
steffen.werner@vt.edu

Ion Victor Gosea
Max Planck Institute Magdeburg
gosea@mpi-magdeburg.mpg.de

Serkan Gugercin
Virginia Tech
Department of Mathematics
gugercin@vt.edu

Michael S. Ackermann
Virginia Tech
amike98@vt.edu

MS8

Emergent Bottleneck Structure in Deep Neural Nets: a Theory of Feature and Symmetry Learning

Deep Neural Networks have proven to be able to break the curse of dimensionality, and learn complex tasks on high dimensional data, such as images or text. But we still do not fully understand what makes this possible. To answer this question, I will describe the appearance of a Bottleneck structure as the number of layers in the network grows, where the network learns low-dimensional features in the middle of the network. This allows the network to identify and learn symmetries of the task it is trained on, without any prior knowledge. This could explain the success of Deep Learning on image and text tasks which feature many 'hidden' symmetries.

Arthur Jacot
New York University
arthur.jacot@nyu.edu

MS8

The Role of Sparsity in Learning With Overparameterized Deep Neural Networks

Much of the folklore that surrounds the success of deep learning is that trained neural networks can automatically adapt to the intrinsic low-dimensional structure found in real-world high-dimensional data. In this talk, we investigate this hallmark in the context of learning with overparameterized deep neural networks trained with weight decay. Our investigation relies on recent developments that show that weight decay is secretly sparsity promoting. We will discuss the development of vector-valued variation spaces, a new class of reproducing kernel Banach spaces. These spaces emerge from studying the effect of weight decay in training networks with activation functions like the ReLU. A key contribution of this work is the development of a representer theorem for the vector-valued variation spaces. This theorem establishes that vector-valued neural networks are the solutions to data-fitting problems over these spaces, where the network widths are bounded by the number of training data squared. This observation reveals that the norm associated with these spaces encourages learning of features that are useful for multiple tasks. Finally, this paper develops a connection between weight-decay regularization and the multi-task lasso problem. This connection leads to novel bounds for layer widths in deep networks that depend on the intrinsic dimensions of the training data representations. This insight yields a simple convex optimization method for deep neural network compression.

Rahul Parhi
USCD
rahul@parhi.com

MS8

Linear Layers in ReLu Networks Promote Learning Single-/Multiple-Index Models

Why do deeper neural networks tend to outperform shallow ones? In this talk I will discuss the role of depth in the simplified case where most layers have a linear activation. Despite these models having the same capacity at different depths, they do not all have the same representation cost. Specifically, minimizing the ℓ_2 penalty when training a neural network with many linear layers followed by a single ReLu layer using weight decay is equivalent to a function-space penalty that encourages the network to select a function with low mixed variation. That is, the function has limited variation in directions orthogonal to a low-dimensional subspace. This means that the trained model will approximately be a single- or multiple- index model. Our experiments show that when this active subspace structure exists in the data, adding linear layers can improve generalization and result in a network that is well-aligned with the true active subspace.

Suzanna Parkinson
University of Chicago
sueparkinson@uchicago.edu

MS8

How Do Neural Networks Learn Features from Data?

Understanding how neural networks learn features, or relevant patterns in data, for prediction is necessary for their reliable use in technological and scientific applications. We propose a unifying mechanism that characterizes feature learning in neural network architectures. Namely, we show that features learned by neural networks are captured by a statistical operator known as the average gradient outer product (AGOP). Empirically, we show that the AGOP captures features across a broad class of network architectures including convolutional networks and large language models. Moreover, we use AGOP to enable feature learning in general machine learning models through an algorithm we call Recursive Feature Machine (RFM). We show that RFM automatically identifies sparse subsets of features relevant for prediction and explicitly connects feature learning in neural networks with classical sparse recovery and low rank matrix factorization algorithms. Overall, this line of work advances our fundamental understanding of how neural networks extract features from data, leading to the development of novel, interpretable, and effective models for use in scientific applications.

Adityanarayanan Radhakrishnan
Harvard University
aradha@seas.harvard.edu

MS9

Open Problems in the Microscopic Deformation of Solids and Their Relation to Random Fields

Deformation in metals is produced by the motion of vast collections of line defects called dislocations. In order to approach this problem, a continuum theory of dislocation dynamics is needed which describes the transport of continuous densities of dislocations. The system of continuum vector density fields which constitute such models are the result of a coarse-graining process; the precise positions of the dislocations are lost by smearing the lines in space. In this presentation, the dislocation density field will be discussed as two random processes. First, the lost coarse-graining information, which affects interactions between dislocations, is considered via the average two-point behavior of the dislocations: dislocation correlation functions. Second, the two-point behavior of the continuum dislocation density field itself will be discussed. This partition of the two-point behavior of the system into short- and

long-range components has implications for the physics of dislocation transport as well as the macroscopic response of the crystal. Some applications which will be discussed are 1) short-range dislocation interaction forces, 2) stochastic contributions to the local stress field, 3) detection of pattern formation, and 4) the X-ray diffraction response of dislocated crystals. Paths forward on these problems in terms of random fields will be discussed.

Joseph P. Anderson
Carthage College
Department of Physics and Astronomy
janderson19@carthage.edu

Anter El-Azab
Purdue University
aelazab@purdue.edu

MS9

Mbius Inversion Meets Tensors: Sampling by Edgeworth Series

Spatial statisticians, physicists, and data scientists all make Gaussian approximations of non-Gaussian phenomena of interest. Statistical inference in Gaussian random variables is highly tractable as operations like conditioning, evaluation of the (log-)likelihood, and sampling all involve numerical computation with the covariance matrix which can be accelerated with the Cholesky factor. We generalize these results to non-Gaussian random variables through the formalism of measure transport. The Cholesky factor becomes the Knothe-Rosenblatt rearrangement, a unique transport map whose Jacobian is lower triangular with positive diagonal. In the non-Gaussian case, we cannot hope to match only the mean and covariance, as in Gaussians. Instead, we will use higher-order cumulant tensors to characterize probability distributions, which can be estimated directly from samples. Expansions based on cumulants such as the Edgeworth series and Cornish-Fisher expansion give simple polynomial expressions for approximate densities, log-densities, and transport maps through Mbius inversion. We can also estimate functionals of the distribution such as entropy directly from samples.

Stephen Huan
Georgia Tech
shuan@gatech.edu

Andreas Robertson
Georgia Institute of Technology
arobertson38

Youssef M. Marzouk
Massachusetts Institute of Technology
ymarz@mit.edu

Florian Schaefer
Georgia Institute of Technology
florian.schaefer@cc.gatech.edu

MS9

Generative Modeling of Conditional Spatial Distributions via Autoregressive Gaussian Processes

In many applications, including climate-model emulation and calibration, there is a need to learn the conditional distribution of a high-dimensional spatial field given a covariate vector, based on a small number of training samples. We propose a nonparametric Bayesian method that decomposes this challenging conditional density estimation task into a large series of univariate autoregressions that we model using heteroskedastic Gaussian processes with carefully chosen prior parameterizations. We describe scalable variational inference based on stochastic gradient descent. The resulting generative model can be used to sample from the learned distribution or transform existing fields as a function of the covariate vector. We provide numerical illustrations and comparisons on simulated data and showcase our method for emulating the distribution encoded by a climate model as a function of its input parameters.

Matthias Katzfuss
University of WisconsinMadison
katzfuss@gmail.com

MS9

Local-Global Decompositions: Data-Scarce and Stable Deep Generative Models for Turning Sparse Experiments into Big Datasets in Materials Science

Data science techniques have demonstrated the potential to help address some of our hardest problems. However, when data is limited often the case in engineering applications, we see dramatic decreases in performance. This talk presents the foundation of a framework for generating large, statistically realistic synthetic datasets for materials informatics from sparse, high fidelity experiments. The primary focus is a proposed family of generative models Local Global Decomposition (LGD) models which allow us to synthesize material microstructures a spatial field which dictates how engineering materials behave conditioned on physically important microstructure statistics. The LGD framework is a proposed decomposition of the probability distribution which conceptually generates a microstructure. The decomposition combines a global, Gaussian distribution, which incorporates the desired microstructure statistics, with a local, learned perturbation to correct the higher order statistics. The learned perturbation is approximated using a deep diffusion model which, critically, we demonstrate can be trained with a single training image. Equally importantly, the stable, Gaussian term allows the model to extrapolate to microstructure statistics outside of its training data. I will conclude the presentation by briefly demonstrating an application of the complete framework for generating the first large microstructure dataset and discuss outlooks and important directions.

Andreas Robertson
Sandia National Laboratories
aerober@sandia.gov

MS10

Accelerating Convergence of Score-Based Diffusion Models, Provably

Diffusion models, which convert noise into new data instances by learning to reverse a Markov diffusion process, have become a cornerstone in contemporary generative modeling. While their practical power has now been widely recognized, theoretical underpinnings for mainstream samplers remain far from mature. Additionally, despite a flurry of recent activities towards speeding up diffusion-based samplers in practice, convergence theory for acceleration techniques remains severely limited. In this talk, we first present a new suite of non-asymptotic theory towards understanding the popular DDIM (or probability flow ODE) sampler in discrete time, which significantly improves upon

prior convergence guarantees for this sampler. We then design training-free algorithms that provably accelerate the DDIM and DDPM samplers, which leverage insights from higher-order approximation and share similar intuitions as popular high-order ODE solvers like DPM-Solver-2. Our non-asymptotic theory accommodates L2-accurate score estimates, and does not require log-concavity or smoothness on the target distribution.

Yuxin Chen
University of Pennsylvania
yuxinc@wharton.upenn.edu

MS10

Score Approximation, Estimation and Distribution Recovery of Diffusion Models on Low-Dimensional Data

Diffusion models achieve state-of-the-art performance in various generation tasks. However, their theoretical foundations fall far behind. This paper studies score approximation, estimation, and distribution recovery of diffusion models, when data are supported on an unknown low-dimensional linear subspace. Our result provides sample complexity bounds for distribution estimation using diffusion models. We show that with a properly chosen neural network architecture, the score function can be both accurately approximated and efficiently estimated. Further, the generated distribution based on the estimated score function captures the data geometric structures and converges to a close vicinity of the data distribution. The convergence rate depends on subspace dimension, implying that diffusion models can circumvent the curse of data ambient dimensionality.

Minshuo Chen
Princeton University
minshuochen@princeton.edu

MS10

The Emergence of Reproducibility and Generalizability in Diffusion Models

We reveal an intriguing and prevalent phenomenon of diffusion models which we term as "consistent model reproducibility': given the same starting noise input and a deterministic sampler, different diffusion models often yield remarkably similar outputs while they generate new samples. We demonstrate this phenomenon through comprehensive experiments and theoretical studies, implying that different diffusion models consistently reach the same data distribution and scoring function regardless of frameworks, model architectures, or training procedures. More strikingly, our further investigation implies that diffusion models are learning distinct distributions affected by the training data size and model capacity, so that the model reproducibility manifests in two distinct training regimes with phase transition: (i) "memorization regime", where the diffusion model overfits to the training data distribution, and (ii) "generalization regime", where the model learns the underlying data distribution and generate new samples with finite training data. Finally, our results have strong practical implications regarding training efficiency, model privacy, and controllable generation of diffusion models, and our work raises numerous intriguing theoretical questions for future investigation.

Qing Qu, Peng Wang
University of Michigan
qingqu@umich.edu, pengwa@umich.edu

MS11

Learning Interaction Kernels for Particle Systems on Networks

We consider stochastic systems of interacting particles on networks, which are commonly used for modeling across the sciences. Oftentimes the laws of interaction between the agents are quite simple, for example they depend only on pairwise interactions, and only on pairwise distance of the pair of states in each interaction. We consider the following inference problem for a system of interacting particles or agents: given only observed trajectories of the agents in the system, can we learn both the unknown network and the unknown laws of interactions? We cast this as an inverse problem, discuss when this problem is well-posed, construct estimators for the interaction kernels with provably good statistically and computational properties. This is joint work with F. Lu and X. Wang.

Mauro Maggioni, Fei Lu, Xiong Wang
Johns Hopkins University
mauromaggionijhu@icloud.com, flu15@jh.edu, xiong_wang@jhu.edu

MS11

A Bilevel Optimization Approach for Inverse Mean-Field Games

Mean field game (MFG) problems analyze the strategic movements of a large number of similar rational agents seeking to minimize their costs. However, in many practical applications, the cost function of MFGs may not be available, rendering the associated agent dynamics unavailable. In this talk, I will discuss our recent work on learning dynamics guided by MFGs. We begin by studying a low-dimensional setting using conventional discretization methods. We propose a bilevel optimization formulation for learning dynamics guided by MFGs with unknown obstacles and metrics. We also establish local unique identifiability results and design an alternating gradient algorithm with convergence analysis. Furthermore, we extend our proposed bi-level method to a deep learning-based algorithm by bridging the trajectory representation of MFG with a special type of deep generative model known as normalizing flows. Our numerical experiments demonstrate the efficacy of the proposed methods. This is a joint work with Quan Xiao, Bill Huang, Tianyi Chen and Rongjie Lai.

Jiajia Yu
Duke University
jiajia.yu@duke.edu

Quan Xiao, Han Huang, Tianyi Chen
Rensselaer Polytechnic Institute
xiaoq5@rpi.edu, huangh14@rpi.edu, chent18@rpi.edu

Rongjie Lai
Purdue University
lairj@purdue.edu

MS12

Evasion Paths in Mobile Sensor Networks

Suppose ball-shaped sensors are scattered in a bounded domain. Unfortunately the sensors don't know their locations

(they're not equipped with GPS), and instead only measure which sensors overlap each other. Can you use this connectivity data to determine if the sensors cover the entire domain? I will explain how tools from topology allow you to address this coverage problem. Suppose now that the sensors are moving; an evasion path exists if a moving intruder can avoid overlapping with any sensor. Can you use the time-varying connectivity data of the sensor network to decide if an evasion path exists? Interestingly, there is no method that gives an if-and-only-if condition for the existence of an evasion path, but I will advertise follow-up questions that remain open!

Henry Adams
University of Florida
henry.adams@ufl.edu

MS12

A Compositional Framework for First-Order Optimization

Optimization decomposition methods are a fundamental tool to develop distributed solution algorithms for large scale optimization problems arising in data science. We present an algebraic framework for hierarchically composing optimization problems defined on hypergraphs and automatically generating distributed solution algorithms that respect this hierarchical structure. The central abstractions of our framework are operads, operad algebras, and algebra morphisms, which formalize notions of syntax, semantics, and structure preserving semantic transformations. These abstractions allow us to formally relate composite optimization problems to the distributed algorithms that solve them. Specifically, we show that certain classes of optimization problems form operad algebras, and a collection of first-order solution methods, namely gradient descent, Uzawa's algorithm, and their subgradient variants, yield algebra morphisms from these problem algebras to algebras of dynamical systems. Primal and dual decomposition are then recovered by applying these morphisms to certain classes of composite problems. We also derive a novel sufficient condition for when a problem defined by compositional data is solvable by a decomposition method. We show that the minimum cost network flow problem satisfies this condition, allowing us to derive a hierarchical dual decomposition algorithm for finding minimum cost flows on composite networks. We discuss our implementation of this framework in Julia.

Tyler Hanks
University of Florida
GATAS Lab
t.hanks@ufl.edu

Matthew Klawonn
Air Force Research Lab
Information Directorate
matthew.klawonn.2@us.af.mil

Matthew Hale
Georgia Institute of Technology
mhale30@gatech.edu

Evan Patterson
Topos Institute
evan@epatters.org

James P. Fairbanks
University of Florida
fairbanksj@ufl.edu

MS12

Mining Sheaf Theoretic Narratives

Querying time-series data or finding motifs therein is well understood for certain data types but remains challenging in the general case. For various data types or patterns sought, it is sometimes unclear what the correct definition of a temporal pattern is even if the analogous static pattern is well studied. One can look to various theories of temporal graphs, for instance, to see how time adds nuances and complexity to the study of common graph theoretic artifacts like paths. In this talk we recapitulate recent work done by collaborators to define temporal data via a sheaf theoretic construct that is agnostic to the object of study, captures richer structure than mere sequences of data, and leads to natural temporal versions of static patterns, among other benefits. We further show how this sheaf theoretic definition of temporal data can be leveraged for finding/counting occurrences of temporal patterns within a time-series by way of a homomorphism search algorithm. We will discuss empirical results of experiments grounded in Air Force planning applications.

Matthew Klawonn
Air Force Research Lab
Information Directorate
matthew.klawonn.2@us.af.mil

James P. Fairbanks
University of Florida
fairbanksj@ufl.edu

MS12

Session Overview and The Synthetic Theory of Probability and Statistics

Probability theory and statistics are usually conceived analytically, grounded in the standard measure-theoretic foundation. In recent years, significant progress has made toward a synthetic account of probability and statistics that treats many of the common definitions and theorems in an axiomatic and purely algebraic setting. The synthetic theory is based on Markov categories and other category-theoretic ideas. Besides widening the scope of probability theory to nonstandard settings, the synthetic theory can be used to give a formal semantics to probabilistic programs, a structuralist account of statistical models and the relationships between them, and a language for specifying statistical models richer than the familiar graphical models. In this talk, we give an introduction to the synthetic theory of probability and statistics.

Evan Patterson
Topos Institute
evan@epatters.org

MS13

Compact Hessian Estimates for Data-Fitting

For minimization problems without 2nd derivative information, methods that estimate Hessian matrices can be very effective. However, conventional techniques generate dense matrices that are prohibitive for large problems. Limited-memory compact representations express the dense arrays in terms of a low rank representation and have become the state-of-the-art for software implemen-

tations on large deterministic problems. We develop new compact representations that are parameterized by a choice of vectors and that reduce to existing well known formulas for special choices. We demonstrate effectiveness of the compact representations for data fitting problems such as tensor factorizations and nonlinear regressions.

Johannes Brust
Arizona State University
johannes.brust@asu.edu

MS13

An Extension of the Tensor Randomized Kaczmarz Algorithm to Problems with Factorized Operator or Data

Solving large-scale systems of linear equations or linear regressions has vital applications in nearly every field of data-driven science. Recently, Kaczmarz-type methods have been proposed for a variety of tensor linear systems and regression problems. In this talk, I will present convergence results on a variant of the randomized Kaczmarz method for tensor regression problems where the data tensor is factorized by means of the t-product.

Paulina Hoyos
University of Texas Austin
paulinah@utexas.edu

MS13

Randomized Gauss-Seidel and Column-Slice-Action Methods for Tensor Problems

In this talk, we delve into the realm of randomized iterative methods for matrix-vector and matrix-matrix regression, honing in on extending the Gauss-Seidel method to tensor regression under the t-product. While Kaczmarz-type methods garner attention in tensor regression, column-action methods remain relatively unexplored in the literature. Our primary objective is to discuss column-slice-action iterative methods for linear tensor problems, particularly crucial in scenarios where row slice sizes exceed active memory capacity or data is naturally organized into column-slice components. We present a series of numerical experiments demonstrating the efficacy of our proposed methods, showcasing their utility in tasks such as image deblurring.

Alejandra Castillo
Oregon State University
castiale@ oregonstate.edu

Jamie Haddock
Harvey Mudd College
jhaddock@g.hmc.edu

Iryna Hartsock
H. Lee Moffitt Cancer Center & Research Institute (US)
iryna.hartsock@moffit.org

Paulina Hoyos-Restrepo
The University of Texas at Austin
paulinah@utexas.edu

Lara Kassab
Colorado State University
lara.kassab@colostate.edu

Alona Kryshchenko
California State University of Channel Islands
alona.kryshchenko@csuci.edu

Kamila Larripa
Humboldt State University
kamila.larripa@gmail.com

Deanna Needell
University of California, Los Angeles, U.S.
deanna@math.ucla.edu

Shambhavi Suryanarayanan
Princeton University
shambhavi.suryanarayanan@gmail.com

Karamatou Yacoubou-Djima
Wellesley College
ky105@wellesley.edu

MS13

Optimal Matrix-Mimetic Tensor Algebras via Variable Projection

Many data are naturally represented as multiway arrays or tensors, and as a result, multilinear data analysis tools have revolutionized feature extraction and data compression. Despite the success of tensor-based approaches, fundamental linear algebra properties often break down in higher dimensions. Recent advances in matrix-mimetic tensor algebra in have made it possible to preserve linear algebraic properties and, as a result, to obtain optimal representations of multiway data. Matrix-mimeticity arises from interpreting tensors as t-linear operators, which in turn are parameterized by invertible linear transformations. The choice of transformation is critical to representation quality, and thus far, has been made heuristically. In this talk, we will learn data-dependent, orthogonal transformations by leveraging the optimality of matrix-mimetic representations. In particular, we will exploit the coupling between transformations and optimal tensor representations using variable projection. We will highlight the efficacy of our proposed approach on image compression and reduced order modeling tasks.

Elizabeth Newman, Katherine Keegan
Emory University
elizabeth.newman@emory.edu, kkeega3@emory.edu

MS14

Emergent Low-Dimensional Subspaces in High-Dimensional Stochastic Gradient Descent

It has been empirically observed that the spectrum of neural network Hessians after training have a bulk concentrated near zero, and a few outlier eigenvalues. Moreover, the eigenspaces associated to these outliers have been associated to a low-dimensional subspace in which most of the training occurs, and this implicit low-dimensional structure has been used as a heuristic for the success of high-dimensional classification. We will describe recent rigorous results in this direction for the Hessian spectrum over the course of the training by SGD in high-dimensional classification tasks with one and two-layer networks. We focus on the separation of outlier eigenvalues from the bulk, and subsequent crystallization of the outlier eigenvectors. Based on joint work with Ben Arous, Huang, and Jagan-

nath.

Reza Gheissari
Northwestern University
gheissari@northwestern.edu

MS14

Machine Learning, Q Functions, and Arnold Tongues for the Synchronization of Stochastic Oscillators

Large-scale brain oscillations may reflect the synchronous behavior of neuron populations, yet the mechanisms underlying collective neuron dynamics are not well understood. One approach is modeling neuron populations as systems of oscillators: ODEs with stable limit-cycle solutions. However, neural activity is noisy, necessitating the study of coupled stochastic oscillators. Koopman operator spectral methods provide a universal description of stochastic oscillators [Perez et al 2023 PNAS]. However, the structure of the low-lying spectra for coupled stochastic oscillators remains an important open question. It is challenging to solve the related PDE for the Koopman eigenmodes; standard methods suffer from the curse of dimensionality and often yield inadequate results in dimensions $n \geq 4$. Recently, machine learning (ML) methods have proven effective for solving high-dimensional PDEs [Zhai et al 2022 PMLR]. Here, we derive a novel ML-based PDE method to compute the Koopman eigenmodes of coupled stochastic oscillators, which is effective for both weakly and strongly coupled systems. We complement our numerical methods with precise analytic statements describing the change in the low-lying spectra of weakly coupled oscillators. Moreover, we observe a non-smooth bifurcation in the low-lying eigenvalues of symmetrically coupled stochastic oscillators, which we relate to synchronization regimes analogous to Arnold tongues for deterministic coupled oscillators.

Maxwell Kreider
Case Western University
mbk62@case.edu

MS14

Data-Driven Prediction of Single-Cell Fate: Large Deviations and Extremal Events

Large deviation theory concerns the asymptotics of exponentially rare events. Such events become important when their impact (e.g. on the accuracy of estimator) is exponentially large, thereby cancelling out their rarity. This talk will provide a brief background on large deviation theory, before surveying some examples of inference problems where large deviations need to be considered, including my own recent work on importance sampling methods to predict how gene expression profiles respond to antibiotics.

Ethan Levien
Dartmouth College
ethanlevien@gmail.com

MS14

Normalization Effects and Scaling Limits for Deep Neural Networks: Towards Statistically Robust Models

We study the effect of normalization on the layers of deep neural networks. A given layer i with N_i hidden units is allowed to be normalized by $1/N_i^{\gamma_i}$ with $\gamma_i \in [1/2, 1]$ and we study the effect of the choice of the γ_i on the statistical behavior of the neural networks output (such as variance) as well as on the test accuracy on the MNIST and CIFAR10 data sets. We find that in terms of variance of the neural networks output and test accuracy the best choice is to choose the γ_is to be equal to one, which is the mean-field scaling. We also find that this is particularly true for the outer layer, in that the neural networks behavior is more sensitive in the scaling of the outer layer as opposed to the scaling of the inner layers. The mechanism for the mathematical analysis is an asymptotic expansion for the neural networks output and corresponding mean field analysis. An important practical consequence of the analysis is that it provides a systematic and mathematically informed way to choose the learning rate hyperparameters. Such a choice guarantees that the neural network behaves in a statistically robust way as the N_i's grow to infinity.

Konstantinos Spiliopoulos
Boston University
Department of Mathematics and Statistics
kspiliop@bu.edu

MS15

Auditing Fairness under Unobserved Confounding

Inequity in resource allocation has been well-documented in many domains, such as healthcare. Causal measures of equity / fairness seek to isolate biases in allocation that are not explained by other factors, such as underlying need. However, these fairness measures require the (strong) assumption that we observe all relevant indicators of need, an assumption that rarely holds in practice. For instance, if resources are allocated based on indicators of need that are not recorded in our data ("unobserved confounders"), we may understate (or overstate) the amount of inequity. In this talk, I will present work demonstrating that we can still give informative bounds on certain causal measures of fairness, even while relaxing (or even eliminating) the assumption that all relevant indicators of need are observed. We use the fact that in many real-world settings (e.g., the release of a new treatment) we have data from prior to any allocation, which can be used to derive unbiased estimates of need. This result is of immediate practical interest: we can audit unfair outcomes of existing decision-making systems in a principled manner. For instance, in a real-world study of Paxlovid allocation, we show that observed racial inequity cannot be explained by unobserved confounders of the same strength as important observed covariates. This talk is based on the following paper (AISTATS 2024): https://arxiv.org/abs/2403.14713

Michael Oberst
Johns Hopkins University
moberst@jhu.edu

MS15

Protected Test-Time Adaptation via Testing by Betting

This talk presents a novel approach for test-time adaptation via online self-training, consisting of two components. First, I'll introduce a statistical framework that detects distribution shifts in the classifier's entropy values obtained on a stream of unlabeled samples. Then, Ill show how to devise an online adaptation mechanism that utilizes the evidence of distribution shifts captured by the detection tool to dynamically update the classifier's parameters. The resulting adaptation process drives the distribution of test

entropy values obtained from the self-trained classifier to match those of the source domain, building invariance to distribution shifts. This approach combines concepts in betting martingales, online learning, and optimal transport. Experimental results demonstrate that this approach improves test-time accuracy under distribution shifts while maintaining accuracy and calibration in their absence, outperforming leading entropy minimization methods across various scenarios.

Yaniv Romano
CS and ECE Departments, Technion
Israel
yromano@technion.ac.il

MS15

Collective Outlier Detection and Enumeration with Conformalized Closed Testing

This talk presents a flexible distribution-free method for collective outlier detection and enumeration, designed for situations in which the presence of outliers can be detected powerfully even though their precise identification may be challenging due to the sparsity, weakness, or elusiveness of their signals. The described method builds upon recent developments in conformal inference and integrates classical ideas from other areas, including multiple testing, rank tests, and non-parametric large-sample asymptotics. The key innovation lies in developing a principled and effective approach for automatically choosing the most appropriate machine learning classifier and two-sample testing procedure for a given data set. The performance of our method is investigated through extensive empirical demonstrations, including an analysis of the LHCO high-energy particle collision data set.

Matteo Sesia
University of Southern California
Marshall School of Business
sesia@marshall.usc.edu

MS16

The Occupation Kernel Method for Learning Vector Fields with Constraints

The occupation kernel method (OCK) has proven itself as a robust and efficient method for learning nonparametric systems of ordinary differential equations from trajectories in arbitrary dimensions. Using an implicit formulation provided by vector-valued reproducing kernel Hilbert spaces, we aim to show how the OCK method can be adapted to learn vector fields satisfying various physical constraints. In particular, by choosing an appropriate kernel, we can ensure that the learned vector fields analytically satisfy either solenoidal (divergence-free) and irrotational (curl-free) properties. We validate the proposed method through experiments on a variety of simulated and real datasets. It is shown that the added constraints often lead to better approximations in these application specific problems.

Nicholas Fisher
Portland State University
nicholfi@pdx.edu

MS16

Solving Parameterized-PDE-Constrained Inverse Problems for Flow Reconstruction and Particle Characterization

Lagrangian particle tracking is a workhorse tool for experimental research on fluid turbulence. Particles are seeded into a flow, illuminated, imaged, and tracked to obtain Lagrangian trajectories, The sparse trajectories, a.k.a. tracks, are reconstructed to obtain 4D Eulerian flow states, which comprise velocity, pressure, and other fields. Data assimilation (DA) is typically used to synthesize the tracks with the governing partial differential equations (PDEs) i.e., the NavierStokes equations (NSE) to enhance the accuracy of velocity data and infer pressure. Particle motion is described by the MaxeyRiley equation (MRE), but the particles are usually assumed to be ideal tracers that faithfully follow the flow, in which case the MRE has a trivial solution, namely, the carrier fluid velocity along each particle streamline. In some cases, however, particles are too dense, too light, or too large to follow the flow, and the MRE must be included in the DA procedure. This talk reviews a neural DA method for reconstructing rich 4D flow states from inertial particles by solving the NSE and MRE, subject to a set of inertial Lagrangian tracks. Further, the MRE is parameterized in terms of the particle size and density, which are inferred for each individual particle as part of the reconstruction procedure. This is the first demonstration of flow reconstruction based on the coupled NSE and MRE. Information content of the inertial tracks is assessed.

Samuel Grauer
Penn State University
Dept. of Mechanical Engineering
sgrauer@psu.edu

MS16

Statistical Mechanics of Dynamical System Identification

Recovering dynamical equations from observed noisy data is the central challenge of system identification. We develop a statistical mechanical approach to analyze sparse equation discovery algorithms, which typically balance data fit and parsimony via hyperparameter tuning. In this framework, statistical mechanics offers tools to analyze the interplay between complexity and fitness, in analogy to that done between entropy and energy. To establish this analogy, we define the hyperparameter optimization procedure as a two-level Bayesian inference problem that separates variable selection from coefficient values and enables the computation of the posterior parameter distribution in closed form. A key advantage of employing statistical mechanical concepts, such as free energy and the partition function, is in the quantification of uncertainty, especially in the low-data limit that is frequently encountered in real-world applications. As the data volume increases, our approach mirrors the thermodynamic limit, leading to distinct sparsity- and noise-induced phase transitions that delineate correct from incorrect identification. This perspective of sparse equation discovery is versatile and can be adapted to various other equation discovery algorithms. https://arxiv.org/abs/2403.01723

Andrei A. Klishin
University of Washington
aklishin@uw.edu

MS16

An Operator Theoretic Approach to Resolving In-

verse Problems

Koopman operators have been used throughout the past decade to resolve inverse problems in Dynamical Systems theory. This combination of Hilbert spaces, Koopman operators, and the full state observable can be abstracted in such a way to apply to a broader range of inverse problems. In this talk we will discuss this general framework, and present several generalizations.

Joel Rosenfeld
University of South Florida
Dept. of Math and Stats
rosenfeldj@usf.edu

MS17

Combining Diffusion Models and SMC for (Posterior) Sampling

We provide connections between stochastic optimal control, diffusion models, and posterior sampling. This allows us to solve inverse problems using recently developed sampling methods based on controlled diffusion processes. We further show how to improve performance by combining these approaches with SMC methods and ideas from reinforcement learning. Finally, we evaluate our proposed algorithms on several inverse problems.

Julius Berner
California Institute of Technology
jberner@caltech.edu

MS17

Generative Models for Inverse Design of Multi-Physics Devices: Challenges and Opportunities

This talk will outline some of our recent work in Generative Models for Inverse Design, which is the use of ML-based models to predict the optima of various design optimization problems. Examples include primarily those expressed as PDE-Constrained Optimization problems in Aerodynamics, Conjugate Heat Transfer, and Heat Conduction. These problems are difficult to solve and to predict optima for because they are generally non-convex (i.e., possess multiple optima) and are also often multi-objective (i.e., we need to generate not just a single design but rather samples along a Pareto Frontier). We will discuss some of these challenges as well as recent work we have done to address them using concepts from Optimal Transport (e.g., Entropic GAN models) and diffusion models as well as a new form of regularization in Autoencoders which we call *Least Volume* that possesses useful empirical and theoretical properties for understanding underlying data manifolds upon which useful designs reside. We will also comment on practical considerations that occur in settings with training sample size or quality is limited and how this intersects with model choice.

Qiuyi Chen
University of Maryland
qchen88@umd.edu

Mark Fuge
Maryland University
fuge@umd.edu

MS17

Diffusion Models for Solving Probabilistic Inverse Problems

Conditional generative models aim to sample from a conditional distribution given samples from the joint distribution, which makes them particularly suitable as tools for solving inverse problems in a Bayesian setting, especially when solving the forward problem is efficient. Thus, a significant body of work exists on utilizing conditional generative models for solving inverse problems. Diffusion modelsa type of generative modelare responsible for significant advances in generative artificial intelligence. This work explores the use of conditional score-based diffusion models for solving physics-based inverse problems. Using multiple challenging applications in computational mechanics involving synthetic and experimental data, this work explores the efficacy of conditional score-based diffusion models in solving large-scale physics-based inverse problems. It also contrasts their performance against conditional generative adversarial networks, another class of popular generative models. It also explores the effect of various sampling and model-related hyperparameters on performance, and the behavior of these models in different limits.

Assad A. Oberai
University of Southern California
aoberai@usc.edu

Agnimitra Dasgupta
Department of Civil and Environmental Engineering
University of Southern California
adasgupt@usc.edu

Javier Murgoitio-Esandi, Harisankar Ramaswamy
University of Southern California
murgoiti@usc.edu, hramaswa@usc.edu

MS17

Discovering Unknown Nonautonomous Stochastic Dynamics from Noisy Data

We present a numerical framework for learning unknown nonautonomous stochastic differential equations. The key ingredient is to transfer the original nonautonomous system into a piecewise stochastic parametric system that is locally time-invariant. This is accomplished by locally parameterizing the time-dependent external inputs over a set of discrete time instances. The resulting system is driven by a parametric stochastic flow map (sFM), which is approximated by our proposed method. A conditional normalizing flow model is devised to learn the local sFM. The final constructed parametric stochastic flow map then defines a stochastic evolution model that is a weak approximation, in terms of distribution, of the unknown nonautonomous stochastic system. A comprehensive set of numerical examples are presented to demonstrate the effectiveness of the proposed sFM method for various types of stochastic systems.

Yuan Chen
Ohio State University
chen.11050@buckeyemail.osu.edu

Zhongshu Xu
The Ohio State University
xu.4202@osu.edu

Dongbin Xiu
Ohio State University

xiu.16@osu.edu

MS18

Physics-guided Full Waveform Inversion Using Encoder-Solver Convolutional Neural Networks

Full Waveform Inversion (FWI) is an inverse problem for estimating wave velocity distribution in a given domain, based on observed data measurements on the boundaries. The inversion process is computationally demanding because we are required to solve multiple forward problems, either in time or frequency domains, to simulate data that is then iteratively fit to the observed data. We consider FWI in the frequency domain, where the Helmholtz equation is used as a forward model, and its repeated solution is the main computational bottleneck of the inversion process. To ease this cost, we integrate a learning process of an encoder-solver preconditioner that is based on convolutional neural networks (CNNs). The encoder-solver is trained to effectively precondition the discretized Helmholtz operator and given velocity medium parameters. Then, by re-training the CNN between the iterations of the optimization process, the encoder-solver is adapted to the iteratively evolving velocity medium as part of the inversion. Without retraining, the performance of the solver deteriorates as the medium changes. Using our light retraining procedures, we obtain the forward simulations effectively throughout the whole process. We demonstrate our approach to solving FWI problems using 2D geophysical models with high-frequency data.

Matan Goren
Ben-Gurion University of the Negev
gorenm@post.bgu.ac.il

Eran Treister
Ben Gurion University of the Negev
erant@cs.bgu.ac.il

MS18

DeepONet-Based Preconditioning

We introduce a new class of hybrid preconditioners for solving parametric linear systems of equations. The proposed preconditioners are constructed by hybridizing the deep operator network, namely DeepONet, with standard iterative methods. Exploiting the spectral bias, DeepONet-based components are harnessed to address low-frequency error components, while conventional iterative methods are employed to mitigate high-frequency error components. Our preconditioning framework utilizes the basis functions extracted from pre-trained DeepONet to construct a map to a smaller subspace, in which the low-frequency component of the error can be effectively eliminated. Our numerical results demonstrate that the proposed approach enhances the convergence of Krylov methods by a large margin compared to standard non-hybrid preconditioning strategies. Moreover, the proposed hybrid preconditioners exhibit robustness across a wide range of model parameters and problem resolutions.

Alena Kopanicakova
Brown University
alena.kopanicakova@usi.ch

George E. Karniadakis
Brown University
Division of Applied Mathematics
george_karniadakis@brown.edu

MS18

Multiscale Neural Networks for Approximating Green's Functions

Solving partial differential equations (PDEs) using neural networks (NNs) has been widely applied in the fields of physics, biology, and engineering. Learning Green's functions is an effective method for solving PDEs with the same differential operator. However, Green's functions are challenging to learn due to their poor regularity, which necessitates larger NN sizes and longer training times. In this paper, we focus on learning Green's functions using multiscale NNs. Through theoretical analysis using multiscale Barron space methods and experimental validation, we demonstrate that the multiscale approach reduces the required neural network size and accelerates training speed.

Rui Peng Li
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
li50@llnl.gov

Yuanzhe Xi
Department of Mathematics
Emory University
yuanzhe.xi@emory.edu

Tianshi Xu
Emory University
tianshi.xu@emory.edu

Yahong Yang
Penn State University
yxy5498@psu.edu

Wenrui Hao
Pennsylvania State University
wxh64@psu.edu

MS19

Linearized Optimal Transport: Theory and Applications

We focus on the problem of classifying distributions and point clouds via the linear optimal transport embedding. This is a method for embedding distributions into a Hilbert space with guarantees on when this embedding is nearly isometric to Wasserstein distance. In this talk, we'll discuss the underlying theory for both distributions on Rd and discrete distributions on graphs. We'll also discuss statistical guarantees for approximating LOT embeddings from finite samples, both using traditional linear programming tools and using input convex neural networks. Finally, we will demonstrate the benefits of this geometric perspective in small data problems where there are only a handful of distributions in the training data.

Alexander Cloninger
University of California, San Diego, USA
acloninger@ucsd.edu

MS19

Ricci Curvature Tensor Based Volumetric Segmentation

Existing curvature used in the regularization term of the

level set model only focuses on 1D curvature and 2D curvature. However, for 3D image segmentation, a proper regularization term should be a well-defined 3D curvature energy. This is the first paper to introduce a regularization energy that incorporates 3D scalar curvature for 3D image segmentation, inspired by the Einstein-Hilbert functional. To derive its Euler-Lagrange equation, we employ a two-step gradient descent strategy, alternatively updating the level set function and its gradient. The paper also establishes the existence and uniqueness of the viscosity solution for the proposed model. Experimental results demonstrate that our proposed model outperforms other state-of-the-art models in 3D image segmentation.

Jisui Huang
Capital Normal University, China
University of Liverpool, UK
jisui@liverpool.ac.uk

Ke Chen
University of Strathclyde
k.chen@strath.ac.uk

Emil Saucan
ORT Braude College, Israel
semil@braude.ac.il

Na Lei
Dalian University of Technology
nalei@dlut.edu.cn

MS19

Function Approximation with Kernel Operators on Vector Bundles

In this talk, we explore kernel smoothing operators on vector bundles and their applications. Given a vector bundle P on a base manifold M, and given a connection on this vector bundle, we look at the generalized notion of the Laplacian operator with the associated kernel operator that acts on sections of P. We argue that this more general notion of smoothing operators extends the concept of smoothing to a much broader range of applications.

Mohammad Latifi, Dimitrios Giannakis
Dartmouth College
Mohammad.Javad.Latifi.Jebelli@dartmouth.edu, Dimitrios.Giannakis@dartmouth.edu

MS19

Geometric Invariants for Textures Classification

We propose a discrete geometric approach to the intelligence and classification of textures in images, with special emphasis on natural ones. We first make appeal to a number of discrete notions of curvature eminently suited for this task, namely both the graph and the full Forman-Ricci curvatures, the Ollivier-Ricci curvature and the Menger curvature measure. Furthermore, we consider a different type of geometric network measure, inspired by the early work of Duffin on electrical networks and stemming from Complex Function Theory, namely the network modulus. In addition, we propose a number of metric invariants introduced by Grove and Markvorsen that encode the essential global geometry of the given structure. Combining these geometric measures with comparison methods developed originally for the study of networks, we are able to distinguish and classify various types of textures, with a special focus on stochastic textures.

Emil Saucan
ORT Braude College, Israel
semil@braude.ac.il

MS20

Revisiting Zeroth-Order Optimization for Memory-Efficient Llm Fine-Tuning: A Benchmark

In the evolving landscape of natural language processing (NLP), fine-tuning pre-trained Large Language Models (LLMs) with first-order (FO) optimizers like SGD and Adam has become standard. Yet, as LLMs grow in size, the substantial memory overhead from back-propagation (BP) for FO gradient computation presents a significant challenge. Addressing this issue is crucial, especially for applications like on-device training where memory efficiency is paramount. This paper proposes a shift towards BP-free, zeroth-order (ZO) optimization as a solution for reducing memory costs during LLM fine-tuning, building on the initial concept introduced by MeZO. Unlike traditional ZO-SGD methods, our work expands the exploration to a wider array of ZO optimization techniques, through a comprehensive, first-of-its-kind benchmarking study across five LLM families (Roberta, OPT, LLaMA, Vicuna, Mistral), three task complexities, and five fine-tuning schemes. Our study unveils previously overlooked optimization principles, highlighting the importance of task alignment, the role of the forward gradient method, and the balance between algorithm complexity and fine-tuning performance. We further introduce novel enhancements to ZO optimization, including block-wise descent, hybrid training, and gradient sparsity. Our study offers a promising direction for achieving further memory-efficient LLM fine-tuning.

Tianlong Chen
The University of North Carolina at Chapel Hill
tianlong@cs.unc.edu

MS20

Towards Constituting Mathematical Structures for Learning to Optimize

Learning to Optimize (L2O), a technique that utilizes machine learning to learn an optimization algorithm automatically from data, has gained arising attention in recent years. A generic L2O approach parameterizes the iterative update rule and learns the update direction as a black-box network. While the generic approach is widely applicable, the learned model can overfit and may not generalize well to out-of-distribution test sets. In this paper, we derive the basic mathematical conditions that successful update rules commonly satisfy. Consequently, we propose a novel L2O model with a mathematics-inspired structure that is broadly applicable and generalized well to out-of-distribution problems. Numerical simulations validate our theoretical findings and demonstrate the superior empirical performance of the proposed L2O model.

Jialin Liu
University of Central Florida
jialin.liu@ucf.edu

MS20

From Large Language Models and Optimization to Decision Optimization CoPilot: A Research Mani-

festo

Significantly simplifying the creation of optimization models for real-world business problems has long been a major goal in applying mathematical optimization more widely to important business and societal decisions. The recent capabilities of Large Language Models (LLMs) present a timely opportunity to achieve this goal. Therefore, we propose research at the intersection of LLMs and optimization to create a Decision Optimization CoPilot (DOCP) - an AI tool designed to assist any decision maker, interacting in natural language to grasp the business problem, subsequently formulating and solving the corresponding optimization model. This paper outlines our DOCP vision and identifies several fundamental requirements for its implementation. We describe the state of the art through a literature survey and experiments using ChatGPT. We show that a) LLMs already provide substantial novel capabilities relevant to a DOCP, and b) major research challenges remain to be addressed. We also propose possible research directions to overcome these gaps. We also see this work as a call to action to bring together the LLM and optimization communities to pursue our vision, thereby enabling much more widespread improved decision-making.

Farzaneh Mirzazadeh
MIT-IBM Watson AI Lab
USA
farzaneh@ibm.com

MS20

Large Language Models As Optimizers

Optimization is ubiquitous. While derivative-based algorithms have been powerful tools for various problems, the absence of gradient imposes challenges on many real-world applications. In this work, we propose Optimization by PROmpting (OPRO), a simple and effective approach to leverage large language models (LLMs) as optimizers, where the optimization task is described in natural language. In each optimization step, the LLM generates new solutions from the prompt that contains previously generated solutions with their values, then the new solutions are evaluated and added to the prompt for the next optimization step. We first showcase OPRO on linear regression and traveling salesman problems, then move on to our main application in prompt optimization, where the goal is to find instructions that maximize the task accuracy. With a variety of LLMs, we demonstrate that the best prompts optimized by OPRO outperform human-designed prompts by up to 8% on GSM8K, and by up to 50% on Big-Bench Hard tasks. Code at https://github.com/google-deepmind/opro.

Chengrun Yang
Google DeepMind
chengrunyang@gmail.com

MS21

Stochastic-Gradient-Based Algorithms for Solving Nonlinearly Constrained Optimization Problems

I will present recent work by my research group on the design and analysis of stochastic-gradient-based algorithms for solving supervised learning problems that incorporate prior information through hard constraints. Our methods handle the constraints with Newton-based techniques, as in state-of-the-art sequential quadratic optimization and interior-point methods, and handle large-scale data-driven objective functions with stochastic-gradient-based ideas. Our experiments reveal superior performance by incorporating a projection-based variant of the popular Adam diagonal scaling technique.

Frank E. Curtis
Industrial and Systems Engineering
Lehigh University
frank.e.curtis@gmail.com

Qi Wang
Lehigh University
qiw420@lehigh.edu

MS21

Enabling Efficient PDE Constrained Optimization Under Uncertainty Using Derivative-Informed Neural Operators

We present a framework for PDE-constrained optimization under uncertainty (OUU) using derivative-informed neural operators (DINOs). OUU problems are often orders of magnitude more expensive to solve compared to their deterministic counterparts due to the need to evaluate statistical/risk measures by stochastic integration, requiring a large number PDE solves. To this end, we propose a neural operator surrogate for the underlying PDE, which is trained on Frèchet derivatives of the solution operator. This ensures that the neural operator has accurate derivatives with respect to the optimization variable thereby improving its fitness for OUU tasks. We present some supporting theoretical results and demonstrate the performance of our method over a suite of numerical experiments, showcasing how DINOs can be used to accurately solve OUU problems in a sample efficient manner.

Dingcheng Luo
Oden Institute for Computational Engineering and Sciences,
The University of Texas at Austin
dc.luo@utexas.edu

Xindi Gong
Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin
xindi.gong@utexas.edu

Thomas O'Leary-Roseberry
The University of Texas at Austin
Oden Institute for Computational and Engineering Sciences
tom@oden.utexas.edu

Peng Chen
Georgia Institute of Technology
pchen402@gatech.edu

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

MS21

Data-Driven Discovery of a Constrained Weno Reconstruction

In this talk, we propose a data-driven approach to learn a weighted essentially non-oscillatory (WENO) reconstruc-

tion, which is constrained to satisfy important constraints. Among these is the so called sign property that is essential for the construction of high-order entropy stable finite difference schemes to solve hyperbolic conservation laws. A strong imposition of consistency, sign-property and accuracy conditions (in smooth regions) on the WENO weights leads to a convex polygonal selection region for these weights. Our strategy involves training a light-weight neural network to find the optimal WENO weights within this region that leads to desirable shock-capturing properties. Solutions to conservation laws are not required to train our network, making our approach is model agnostic. Once trained, the same network can be used to solve any conservation law. We present several numerical results to demonstrate a significant improvement over existing variants of WENO with the sign-property.

Deep Ray
University of Maryland, College Park
deepray@umd.edu

Philip Charles
Maryland University
charlesp@umd.edu

MS21

σ-Anls: An Effective Training Method for Neural Networks with Sigmoidal Activation

We propose a novel training method for multivariate two-layer neural networks of sigmoidal-type activation functions. The core mechanism stems from Active Neuron Least Squares (ANLS), the method we developed for rectified activation functions. It augments the standard gradient descent direction by including search vectors chosen to exploit locality and nonlocality. σ-ANLS follows a similar procedure to ANLS. However, the underlying principle of the search vector generation is completely different. Numerical examples are provided that demonstrate the effectiveness of ANLS compared with existing first- and second-order optimization algorithms on various learning tasks ranging from function approximation, operator learning to solving PDEs.

Yeonjong Shin
North Carolina State University
yeonjong_shin@ncsu.edu

Mark Ainsworth
Division of Applied Mathematics
Brown University
Mark_Ainsworth@brown.edu

MS22

Solving Infinite-Dimensional Inverse Problems with Score-Based Diffusion Models

Diffusion models have recently emerged as a powerful framework for generative modeling. They consist of a forward process that perturbs input data with Gaussian noise and a reverse process that learns a score function to generate samples by denoising. Despite their tremendous success, they are mostly formulated on finite-dimensional Euclidean spaces, excluding their application to domains such as inverse problems where the data consist of functions. In this presentation, we first introduce a framework called Denoising Diffusion Operators (DDOs) for sampling distributions in function space with probabilistic diffusion models. In DDOs, the forward process perturbs input functions using a Gaussian process and learns an appropriate score function in infinite dimensions. We show that our discretized algorithm generates accurate samples at a fixed cost that is independent of the data resolution. Second, we propose a model for super-resolution that does not require the low-frequency content of the inputs and outputs to match. Our model relies on optimal transport to debias the low-resolution data and diffusion models for conditional sampling. We numerically verify that our improvements capture the statistics of high-resolution fluid dynamics problems and popular imaging datasets.

Ricardo Baptista
Caltech
rsb@caltech.edu

Nikola Kovachki
Nvidia
nkovachki@nvidia.com

Leonardo Zepeda-Núñez
University of Wisconsin-Madison
zepedanunez@wisc.edu

Zhong Yi Wan
Google
wanzy@google.com

MS22

Wasserstein Proximals Stabilize Training of Generative Models and Learn Manifolds

In this talk, I will delve into the crucial role of Wasserstein proximals in enhancing the training of generative models. When generative models learn data distributions, they typically encounter two primary challenges: training instability and effectively representing low dimensional manifolds in high dimensional spaces. By reformulating the learning losses of generators as Mean Field Game(MFG) problems, we show that traditional approaches like Normalizing Flows (NFs) and generative adversarial networks (GANs) are ill-posed. This inadequacy leads to training instability for NFs and GANs. However, advanced models that incorporate the Wasserstein-2 proximal through a running cost have proven to be effective at mitigating this issue. On the other hand, capturing low dimensional structure in high dimensional spaces is related to the numerical tractability of the terminal condition in the mean-field game. Unlike NFs, f-GANs address this challenge by introducing Wasserstein-1 proximal in the loss function. Wasserstein proximals offer a mathematical insight into generative model training techniques, and guide a more robust and efficient training approach for generator models. We demonstrate the use of Wasserstein proximal methods to stabilize the training of flow-based generative models.

Hyemin Gu, Benjamin J. Zhang
University of Massachusetts Amherst
gu@math.umass.edu, bjzhang@umass.edu

Markos A. Katsoulakis
University of Massachusetts, Amherst
Dept of Mathematics and Statistics
markos@math.umass.edu

Luc Rey-Bellet
University of Massachusetts Amherst

luc@math.umass.edu

MS22

Diffusion Models for Functions

We present a generalization of score-based diffusion models to function space by perturbing functional data via a Gaussian process at multiple scales. We obtain an appropriate notion of score by defining densities with respect to Guassian measures and generalize denoising score matching. We then define the generative process by integrating a function-valued Langevin dynamic. We show that the corresponding discretized algorithm generates accurate samples at a fixed cost that is independent of the data discretization. Furthermore, we show that by using a Gaussian process reference, we are able to train models which are equivariant with respect to domain diffeomorphisms. This allows us to generate temporally consistent videos by warping the noise in time with a model trained only on natural images.

Nikola Kovachki
Nvidia
nkovachki@nvidia.com

MS22

Quantum State Generation with Structure-Preserving Diffusion Model

This article considers the generative modeling of the states of quantum systems, and an approach based on denoising diffusion model is proposed. The key contribution is an algorithmic innovation that respects the physical nature of quantum states. More precisely, the commonly used density matrix representation of mixed-state has to be complex-valued Hermitian, positive semi-definite, and trace one. Generic diffusion models, or other generative methods, may not be able to generate data that strictly satisfy these structural constraints, even if all training data do. To develop a machine learning algorithm that has physics hard-wired in, we leverage the recent development of Mirror Diffusion Model and design a mirror map, to enable strict structure-preserving generation. Both unconditional generation and conditional generation via classifier-free guidance are experimentally demonstrated efficacious, the latter enabling the design of new quantum states when generated on unseen labels.

Yuchen Zhu, Tianrong Chen
Georgia Institute of Technology
yzhu738@gatech.edu ¡yzhu738@gatech.edu¿,
tchen429@gatech.edu

Evangelos Theodorou
Aerospace Engineering, Georgia Tech
evangelos.theodorou@gatech.edu

Xie Chen
California Institute of Technology
xiechen@caltech.edu

Molei Tao, Molei Tao
School of Mathematics, Georgia Institute of Technology
mtao@gatech.edu, mtao@gatech.edu

MS23

Nonlinear Control Algorithms and Large-Time Behavior of Dissipative Differential Equations

One of the fundamental challenges of accurate simulation of turbulent flows is that initial data is often incomplete, which for said flows is a strong impediment to accurate modeling due to sensitive dependence on initial conditions. A continuous data assimilation method proposed by Azouani, Olson, and Titi in 2014 introduced a linear feedback control term to dissipative systems, proving directly how the maximal effectiveness of DA is dependent on the specific physics of the system being investigated. In this talk, we will focus on the insights of a nonlinear variation of the AOT algorithm and distinguish the clear connections to the large-time behavior of physical systems.

Elizabeth Carlson
California Institute of Technology
elizcar@caltech.edu

MS23

Recent Progress in Using Generative AI Methods for Ensemble Data Assimilation

Generative AI methods have shown impressive performance in image and video generation tasks. Their probabilistic nature serves as an appealing framework for extending many well established branches of data science. One prominent example is the field of ensemble data assimilation (DA), which is widely used in the geosciences to quantify the initial condition uncertainty in numerical models. The Ensemble Kalman Filter (EnKF) is one of the most popular ensemble DA algorithms with demonstrated stability in high-dimensional settings. However, the Gaussian assumptions made in its formulation lead to biases that eventually deteriorate the model prediction skill. This talk will highlight two recent DA frameworks which utilize generative AI tools to relax the Gaussian assumptions in the EnKF. Specifically, the Ensemble Conjugate Transform Filter (ECTF) generalizes the EnKF to arbitrarily complex distributions based on ideas from normalizing flows, while the Ensemble Score Filter (EnSF) samples the Bayesian posterior with training-free diffusion models. Both methods take advantage of the appealing idea that target distributions in the DA problem can be obtained as successive transformations of much simpler base distributions like the Gaussian. Following a brief mathematical introduction, I will describe some recent progress in applying these new DA algorithms to challenging toy models in the geosciences.

Hristo G. Chipilski
Florida State University, Department of Scientific Computing
hchipilski@fsu.edu

MS23

Preserving Linear Invariants in Ensemble Filtering Method

Formulating dynamical models for physical phenomena is essential for understanding the interplay between the different mechanisms, predicting the evolution of physical states, and developing effective control strategies. However, a dynamical model alone is often insufficient to address these fundamental tasks, as it suffers from model errors and uncertainties. One common remedy is to rely on data assimilation. Ensemble filters sequentially assimilate observations by updating a set of samples over time. They operate in two steps: a forecast step and an analysis step. For accurate and robust predictions of dynamical systems,

discrete solutions must preserve their critical invariants. While modern numerical solvers satisfy these invariants, existing invariant-preserving analysis steps are limited to Gaussian settings and are often not compatible with classical regularization techniques. The present work focuses on preserving linear invariants. Using tools from measure transport theory, we introduce a generic class of nonlinear ensemble filters that automatically preserve desired linear invariants in non-Gaussian filtering problems.

Mathieu Le Provost
Massachusetts Institute of Technology
n/a

MS23

Improving the Likelihood Score in Ensemble Score Filter for Nonlinear Data Assimilation with Partial Observations

Recently, the Ensemble Score Filter (EnSF) was introduced to tackle high-dimensional nonlinear data assimilation(DA) problems. The primary limitation of EnSF is its tendency to underestimate the correlation structure within the filtering distribution, rendering it less effective for DA problems with sparse observations. This issue arises because the updates to unobserved dimensions rely heavily on their correlation with observed dimensions. To overcome this, we composite the state dynamics with the observation function in the Bayesian sampling step of EnSF. This composition creates non-zero gradients and updates to the unobserved dimensions without the need for a strong prior correlation structure. Moreover, we employ automatic differentiation and likelihood bias correction techniques to enhance the efficiency of Bayesian sampling in EnSF. Extensive numerical experiments demonstrate that our method maintains accuracy and effectiveness even under conditions of sparse observation.

Zezhong Zhang
Oak Ridge National Laboratory
zhangz2@ornl.gov

MS24

Priors for Efficient Three-Dimensional Imaging

In this talk, we explore appropriate priors for the challenging problem of 3D image reconstruction from indirect measurements. We apply the proposed techniques for synthetic aperture radar (SAR) imaging.

Victor Churchill
Trinity College
victor.churchill@trincoll.edu

MS24

Predicting in Latent Space with Wasserstein Dependence-Maximizing Autoencoders

Wasserstein dependence is an optimal transport-based analog of mutual information that has been gaining traction in both statistical inference and representation learning research. In this talk, I'll show that (sliced) Wasserstein dependence maximization serves as an effective and computationally efficient training objective for autoencoders, significantly reorganizing the latent space as compared to a Gaussian prior. The proposed method additionally admits a theoretically rigorous Bayesian interpretation that helps clarify the extent to which statistics computed in the latent space can be decoded into accurate statistical estimates in the original, high-dimensional data space.

Keenan Eikenberry
Dartmouth College
keenan.j.eikenberry@dartmouth.edu

Lizuo Liu
Southern Methodist University
lizuo.liu@dartmouth.edu

Yoonsang Lee
Dartmouth College
Department of Mathematics
yoonsang.lee@dartmouth.edu

MS24

A Data-Driven Statistical-Stochastic Model for Multiscale Turbulent Systems

The capability of using imperfect stochastic and statistical reduced-order models to capture key statistical features in multiscale nonlinear dynamical systems is investigated. A systematic framework is proposed using a high-order statistical closure enabling accurate prediction of leading-order statistical moments and probability density functions in multiscale complex turbulent systems. A new efficient ensemble forecast algorithm is developed dealing with the nonlinear multiscale coupling mechanism as a characteristic feature in high-dimensional turbulent systems. To address challenges associated with closely coupled spatio-temporal scales in turbulent states and expensive large ensemble simulation for high-dimensional complex systems, we introduce efficient computational strategies using the so-called random batch method. It is demonstrated that crucial principal statistical quantities in the most important large scales can be captured efficiently with accuracy using the new reduced-order model in various dynamical regimes of the flow field with distinct statistical structures. Finally, the proposed model is applied for a wide range of problems in uncertainty quantification, data assimilation, and control. Associated poster: A Statistical-Stochastic Reduced-Order Model for Multiscale Turbulent Systems

Di Qi
Purdue University
qidi@purdue.edu

Jian-guo Liu
Duke University
jian-guo.liu@duke.edu

MS24

Large Deviation-Informed Sampling for Rare Events in High Dimensions and Connection to Bayesian Inversion

Rare and extreme events like hurricanes, energy grid blackouts, dam breaks, earthquakes, and pandemics are infrequent but have severe consequences. Because estimating the probability of such events can inform strategies that mitigate their effects, scientists must develop methods to study the distribution tail of these occurrences. However, calculating small probabilities is hard, particularly when involving complex dynamics and high-dimensional random variables. In this talk, I will discuss our proposed method for the accurate estimation of rare event or failure probabilities for expensive-to-evaluate numerical models in high dimensions, and the connection between this method and the

Bayesian inverse problems. The proposed approach combines ideas from large deviation theory and adaptive importance sampling. The importance sampler uses a cross-entropy method to find an optimal Gaussian biasing distribution, and reuses all samples made throughout the process for both, the target probability estimation and for updating the biasing distributions. Large deviation theory is used to find a good initial biasing distribution through the solution of an optimization problem. Additionally, it is used to identify a low-dimensional subspace that is most informative of the rare event probability. We compare the method with a state-of-the-art cross-entropy-based importance sampling scheme using examples including a tsunami problem.

Shanyin Tong
Columbia University
st3503@columbia.edu

Georg Stadler
Courant Institute for Mathematical Sciences
New York University
stadler@cims.nyu.edu

MS25

Continual Learning in Lithium-Ion Battery Management Systems for Electric Vehicle Safety

Intelligent battery management systems (BMS) with advanced algorithms are crucial for health-conscious decision-making and the safe operation of lithium-ion batteries. The computational complexity of the PDE-based electrochemical-thermal-aging (ETA) model limits its onboard BMS implementation. Recent advancements in physics-influenced neural networks (PINN) offer a solution by avoiding discretization when solving PDEs and compensating for uncertainties through data-driven learning. However, the PINN models may fail with unseen new data under accelerated degradation and internal fault. This presentation discusses a novel approach to embedding the BMS with continual learning capability, leveraging PINN and dynamical system theory. The PINN approximates the solution to the ETA dynamics, which combines single-particle, thermal, and aging models. We introduce an online learning mechanism by considering the PINN as a dynamical system. We define a novel loss function to train the last layer of the PINN by using the error between the measured data and the model output. We leverage the Fisher information matrix to avoid catastrophic forgetting while defining the loss function. Above all, we analytically guarantee the convergence of the neural network weight estimation error and the model output error using the Lyapunov theory. Simulation results are also presented to demonstrate the effectiveness of continual learning and their computational efficiency.

Avimanyu Sahoo
University of Alabama, Huntsville
avimanyu.sahoo@uah.edu

Sazzad Hossen, Saima Alam
University of Alabama in Huntsville
mh0255@uah.edu, sa0175@uah.edu

MS25

Uncertainty-Adjusted Machine Learning via Optimal Control

A typical learning process of a neural network can be described by the interaction between the data, the weights and the model architecture. In a standard regime, this interaction involves updating the weights of the neural network model with respect to a series of data batches for a fixed architecture. Multitude of decisions such as the number of layers, the activation function, learning rate, etc; are made during this process of learning. As these decisions are made apriori in the standard setup, there is an inherent uncertainty in the learning process due to the discrepancy between these decisions and the nature of the data. In this talk, we will demonstrate that this uncertainty could be quantified and further leveraged to improve performance in a neural network training process. Towards this end, we will enunciate a novel optimal control-driven mathematical formulation where the dynamics of learning are represented by a differential equation. Subsequently, we will elucidate, how to quantify this uncertainty during learning?" and provide insights into the effect of the this uncertainty. We will end this talk with insights on how to use uncertainty to correct the behavior of the graph neural network while performing anomaly detection?"

Krishnan Raghavan
Mathematics and Computer Science
Argonne National Laboratory
kraghavan@anl.gov

MS25

Randomized Algorithms for Bayesian Inversion and Data Acquisition in Predictive Digital Twins

A digital twin couples computational models with a physical counterpart to create a system that is dynamically updated through bidirectional data flows as conditions change. Data Assimilation and Optimal Experimental Design (OED) provide a systematic means of updating the computational model and acquiring information as the physical system evolves. This talk will describe scalable preconditioners and solvers for Bayesian inversion using different randomization techniques. The proposed techniques are amenable to parallelization and drastically reduce the required number of model evaluations. We also develop theoretical guarantees on the condition number. Additionally, the talk will describe connections between OED for Bayesian linear inverse problems and the column subset selection problem (CSSP) in matrix approximation and derive bounds, both lower and upper, for the D-optimality criterion via CSSP for the independent and colored noise cases. We demonstrate the performance and effectiveness of our methodology on a variety of test problems such as Burgers and quasi-geostrophic equations.

Vishwas Rao
Argonne National Laboratory
vhebbur@anl.gov

MS25

Dynamic Pattern Formation through Distribution Control of Moment Kernelized Population Systems

Targeted coordination of large-scale populations of dynamical systems for tuning patterns of their measurement data is an emerging and essential task in science and engineering. Its application domains span across numerous disciplines, ranging from motion planning of robot swarms in robotics and synchronization of rhythmic networks in network science to excitation of nuclear spin ensembles in quantum science. However, these dynamic pattern formation tasks are challenged by the lack of a principled for-

mulation and the large population size of the systems. In this work, we propose a distribution control formulation for dynamic pattern formation. Specifically, by leveraging the technique of displacement interpolation in optimal transport theory, a dynamic pattern formation task is formulated as an optimal control problem over the space of probability distributions. We then develop the moment kernel transform, which enables the representation of the evolution of the measurement data pattern of a population system in terms of a dynamical system, referred to as a moment system, defined on a reproducing kernel Hilbert space. To control the moment system by using the data, we propose a deep neural network architecture, composed of multiple recurrent neural network layers, which are trained by using the transfer learning technique. The applicability and performance of the developed dynamic pattern formation framework are demonstrated using examples from practical applications.

Wei Zhang, Yi-Hsuan Shih, Jr-Shin Li
Washington University in St. Louis
wei.zhang@wustl.edu, shihy@wustl.edu, jsli@wustl.edu

MS26

Guided Data Exploration with (Semi-)supervised Manifold Learning

Modern challenges in exploratory data analysis, especially in biomedical applications involving single cell data, give rise to representation learning techniques that aim to capture intrinsic data geometry, while separating it from data distribution that is typically biased by data availability and collection artifacts, thus allowing discovery of rare subpopulations and sparse transitions between meta-stable states. A common approach in this area is the construction of a data-driven diffusion geometry that both captures intrinsic structure in data and provides a generalization of Fourier harmonics on it, combining tools and perspectives from a range of fields including manifold learning, graph signal processing, and harmonic analysis. However, most methods following this paradigm rely on unsupervised learning, under the assumption that the target phenomena of interest will form the dominant emergent patterns in the data, uncovered by the extracted representation. While this is the case in certain controlled experiment conditions, such property cannot be guaranteed in many observational services settings. As an alternative, here we present a semi-supervised approach that leverages annotations and meta information that often accompanies collected data, in order to guide the data geometry to accentuate task-informed structures in the learned representation. This approach will be demonstrated in data exploration tasks including visualization and multimodal data fusion.

Guy Wolf
Université de Montréal
guy.wolf@umontreal.ca

Adrien Aumon
Mila, Canada
adrien.aumon@mila.quebec

MS26

Modeling Single-cell and Spatial Transcriptomics Data Using Optimal Transport

Single-cell and spatial transcriptomics data examines high-throughput gene expression profiles at fine resolutions providing an unprecedented opportunity to elucidate the underlying complex biological processes. Optimal transport has proven to be an effective tool for various applications with such data, such as multi-omics integration. In this talk, we will discuss several optimal transport variants motivated by the biological applications, where there are detailed application-specific constraints, multiple distribution species, and multiple embedding spaces of the same system. We will illustrate the applications of these tools for addressing multi-compatible molecular species in cell-cell communication analysis and devising coherent trajectories of the same biological system from multi-omics datasets.

Zixuan Cang
Department of Mathematics
North Carolina State University
zcang@ncsu.edu

MS26

Statistical Inference on Shapes and Grayscale Images Via Euler Characteristic

In this talk, we illuminate methods grounded in the Euler characteristic for statistical inferences on binary and grayscale images. We commence with the analysis of shapes/binary images through the lens of the smooth Euler characteristic transform (SECT). Here, we lay out the distributional properties of the SECT. Based on these foundations, we propose algorithms for hypothesis testing on random shapes. Numerical experiments further underscore our mathematical derivations and demonstrate the performance of our proposed algorithms. Shifting our focus, we delve into the analysis of grayscale images utilizing the Euler-Radon transform (ERT). The mathematical foundations of the ERT are developed using Euler calculus. Building upon this, we generalize the SECT-based algorithms to the statistical inference on grayscale images. Our proposed algorithms apply to analyze databases in geometric morphometrics and medical imaging. Collectively, this talk serves as a confluence of myriad fields: algebraic and computational topology, probability theory and stochastic processes, Sobolev spaces and functional analysis, analysis of variance for functional data, and geometric morphometrics.

Kun Meng
Brown University
kun_meng@brown.edu

MS26

Trajectory Inference in Wasserstein Space

Capturing data from dynamic processes through cross-sectional measurements is seen in fields from computer graphics to robot path planning and cell trajectory inference. This inherently involves the challenge of understanding and reconstructing the continuous trajectory of these processes from discrete data points, for which interpolation and approximation plays a crucial role. In this work, we propose a method to compute measure-valued B-splines in the Wasserstein space through consecutive averaging. Our method can carry out approximation with high precision and at a chosen level of refinement, including the ability to accurately infer trajectories in scenarios where particles undergo splitting (division) over time. We rigorously evaluate our method using simulated cell data characterized by bifurcations and merges, comparing its performance against both state-of-the-art trajectory inference techniques and other interpolation methods. The results of our work not only underscore the effectiveness of our method in address-

ing the complexities of inferring trajectories in dynamic processes but also highlight its proficiency in performing spline interpolation that respects the inherent geometric properties of the data.

Caroline Moosmueller
University of North Carolina at Chapel Hill
cmoosm@unc.edu

MS27

Optimization in Wasserstein Space via Interacting Particle Systems

Motivated by variational inference problems, we propose a zero-order optimization algorithm based on measure-valued particles interacting via a dynamics of consensus type. The evolution exploits the Riemannian structure of the 2-Wasserstein space and is described by a system of measure differential inclusions. We will discuss possible numerical implementations, and also consider restriction of the dynamics to the Bures-Wasserstein sub-manifold.

Giacomo Borghi
RWTH Aachen University
Department of Mathematics
g.borghi@hw.ac.uk

MS27

Ensemble Kalman Methods for Non-Linear Filtering: A Mean-Field Perspective

The ensemble Kalman methodology is an innovative and flexible set of tools which can be used for both state estimation in dynamical systems and parameter estimation for generic inverse problems. Despite its widespread adoption in fields of application, firm theoretical foundations are only now starting to emerge. We consider a unifying approach to algorithms that rests on transport of probability measures and mean-field stochastic dynamical systems. With the goal of developing theoretical guarantees for ensemble Kalman methods applied to non-linear problems, we discuss the error analysis of the stochastic dynamical systems arising in ensemble Kalman filtering and the associated probability measures.

Edoardo Calvello
California Institute of Technology
e.calvello@caltech.edu

MS27

Clustering Within Causal Transformer Attention Dynamics

We study Transformers that use a causal attention mechanism. By modeling token dynamics as a continuous-time interacting particle system on a sphere, we characterize limiting configurations of the system in terms of the spectrum of the value matrix. Moreover, we find that under certain conditions, particles tend to quickly group into several meta-clusters, which continue to slowly collapse into one or two clusters over a long period of time. Finally, we provide a method to predict the positions of the meta-clusters based solely on initial conditions and describe properties of those predictions.

Nikita Karagodin
Massachusetts Institute of Technology
nikitus@mit.edu

MS27

Sampling in Unit Time with Kernel Fisher-Rao Flow

We introduce a new mean-field ODE and corresponding interacting particle systems (IPS) for sampling from an unnormalized target density. The IPS are gradient-free, available in closed form, and only require the ability to sample from a reference density and compute the (unnormalized) target-to-reference density ratio. The mean-field ODE is obtained by solving a Poisson equation for a velocity field that transports samples along the geometric mixture of the two densities, $\pi_0^{1-t}\pi_1^t$, which is the path of a particular Fisher–Rao gradient flow. We employ a RKHS ansatz for the velocity field, which makes the Poisson equation tractable and enables discretization of the resulting mean-field ODE over finite samples. The mean-field ODE can be additionally be derived from a discrete-time perspective as the limit of successive linearizations of the Monge–Ampère equations within a framework known as sample-driven optimal transport. We introduce a stochastic variant of our approach and demonstrate empirically that our IPS can produce high-quality samples from varied target distributions, outperforming comparable gradient-free particle systems and competitive with gradient-based alternatives.

Aimee Maurais, Youssef M. Marzouk
Massachusetts Institute of Technology
maurais@mit.edu, ymarz@mit.edu

MS28

Transferable Neural Networks for Partial Differential Equations

Transfer learning for partial differential equations (PDEs) is to develop a pre-trained neural network that can be used to solve a wide class of PDEs. Existing transfer learning approaches require much information about the target PDEs such as its formulation and/or data of its solution for pre-training. In this work, we propose to design transferable neural feature spaces for the shallow neural networks from purely function approximation perspectives without using PDE information. The construction of the feature space involves the re-parameterization of the hidden neurons and uses auxiliary functions to tune the resulting feature space. Theoretical analysis shows the high quality of the produced feature space, i.e., uniformly distributed neurons. We use the proposed feature space as the predetermined feature space of a random feature model, and use existing least squares solvers to obtain the weights of the output layer. Extensive numerical experiments verify the outstanding performance of our method, including significantly improved transferability, e.g., using the same feature space for various PDEs with different domains and boundary conditions, and the superior accuracy, e.g., several orders of magnitude smaller mean squared error than the state of the art methods.

Zezhong Zhang
Oak Ridge National Laboratory
zhangz2@ornl.gov

Feng Bao
Florida State University
bao@math.fsu.edu

Lili Ju
University of South Carolina
Department of Mathematics
ju@math.sc.edu

Guannan Zhang
Oak Ridge National Laboratory
zhangg@ornl.gov

MS28

Understanding Contrastive Learning from Variational Analysis and Neural Network Optimization Perspectives

not available

Wonjun Lee
University of Minnesota
University of California, Los Angeles
lee01273@umn.edu

MS28

Efficient Hybrid Spatial-Temporal Operator Learning

Recent advancements in operator-type neural networks, such as Fourier Neural Operator (FNO) and Deep Operator Network (DeepONet), have shown promising results in approximating the solutions of spatial-temporal Partial Differential Equations (PDEs). However, these neural networks often entail considerable training expenses, and may not always achieve the desired accuracy required in many scientific and engineering disciplines. In this talk, we propose a new operator learning framework to address these issues. The proposed paradigm leverages the traditional wisdom from numerical PDE theory and techniques to refine the pipeline of existing operator neural networks. The novel design allows the operator neural networks to effectively tackle low-frequency errors, while the added linear layer addresses high-frequency errors. Numerical experiments on a commonly used benchmark of 2D Navier-Stokes equations demonstrate significant improvements in both computational time and accuracy, compared to existing FNO variants and traditional numerical approaches.

Yuanzhe Xi
Department of Mathematics
Emory University
yuanzhe.xi@emory.edu

Shuhao Cao
Washington University in St. Louis
scao@umkc.edu

Francesco Brard
Emory University
francesco.brarda@emory.edu

Ruipeng Li
Lawrence Livermore National Laboratory
li50@llnl.gov

MS29

Complex Networks with Complex Weights

In many studies, it is common to use binary (i.e., unweighted) edges to examine networks of entities that are either adjacent or not adjacent. Researchers have generalized such binary networks to incorporate edge weights, which allow one to encode nodenode interactions with heterogeneous intensities or frequencies (e.g., in transportation networks, supply chains, and social networks). Most such studies have considered real-valued weights, despite the fact that networks with complex weights arise in fields as diverse as quantum information, quantum chemistry, electrodynamics, rheology, and machine learning. Many of the standard network-science approaches in the study of classical systems rely on the real-valued nature of edge weights, so it is necessary to generalize them if one seeks to use them to analyze networks with complex edge weights. In this paper, we examine how standard network-analysis methods fail to capture structural features of networks with complex edge weights. We then generalize several network measures to the complex domain and show that random-walk centralities provide a useful approach to examine node importances in networks with complex weights.

Lucas Boettcher
University of Florida
Gainesville
lbottcher@ufl.edu

MS29

Small Worldness in Complex-valued Complex Networks

The application of graph theory to networks has resulted in a myriad of classical applications across biology, economics, epidemiology, sociology, and soft condensed matter physics. Recent applications such as signal transfer networks have demonstrated the need to extend common network measures to networks with complex-valued weights. In this talk, I will describe one approach to extending concepts of clustering coefficient and path length to explore small world properties in a complex network with complex-valued edge weights. Our results highlight the ways in which constructive and destructive interference can respectively enhance or degrade the small worldness of a complex network.

Cecilia Diniz Behn
Colorado School of Mines
Applied Math and Statistics
cdinizbe@mines.edu

MS29

MagNet: A Neural Network for Directed Graphs

The prevalence of graph-based data has spurred the rapid development of graph neural networks (GNNs) and related machine learning algorithms. Yet, despite the many datasets naturally modeled as directed graphs, including citation, website, and traffic networks, the vast majority of this research focuses on undirected graphs. In this paper, we propose MagNet, a GNN for directed graphs based on a complex Hermitian matrix known as the magnetic Laplacian. This matrix encodes undirected geometric structure in the magnitude of its entries and directional information in their phase. A charge parameter attunes spectral information to variation among directed cycles. We apply our network to a variety of directed graph node classification and link prediction tasks showing that MagNet performs well on all tasks and that its performance exceeds all other methods on a majority of such tasks. The underlying principles of MagNet are such that it can be adapted to other

GNN architectures.

Michael Perlmutter
Boise State University
mperlmutter@boisestate.edu

MS29

Structural Balance and Random Walks on Complex Networks with Complex Weights

Complex numbers define the relationship between entities in many situations. A canonical example would be the off-diagonal terms in a Hamiltonian matrix in quantum physics. Recent years have seen an increasing interest to extend the tools of network science when the weight of edges are complex numbers. In this talk, we focus on the case when the weight matrix is Hermitian, a reasonable assumption in many applications, and investigate both structural and dynamical properties of the networks with complex weights. Building on concepts from signed graphs, we introduce a classification of complex-weighted networks based on the notion of structural balance, and illustrate the shared spectral properties within each type. We then apply the results to characterize the dynamics of random walks on complex-weighted networks, where local consensus can be achieved asymptotically when the graph is structurally balanced, while global consensus will be obtained when it is strictly unbalanced. Finally, we explore potential applications of our findings by generalizing the notion of cut, and propose an associated spectral clustering algorithm. We also provide further characteristics of the magnetic Laplacian, associating directed networks to complex-weighted ones. The performance of the algorithm is verified on both synthetic and real networks.

Yu Tian
Nordic Institute for Theoretical Physics
yu.tian@su.se

MS30

Interpretable Fine-Tuning for Graph Neural Network Surrogate Models

This work introduces an interpretable fine-tuning strategy for mesh-based modeling in the graph neural network (GNN) framework with application to fluid dynamics forecasting. Given a pre-trained baseline surrogate model, the end result is a fine-tuned model that identifies structures in physical space (corresponding to sub-graphs) intrinsically linked to the forecasting task, while retaining baseline predictive capability. These structures are adaptively produced in the forward pass and serve as explainable links between the baseline model architecture, the optimization goal, and known problem-specific physics. Additionally, through a regularization procedure, fine-tuned GNNs can also be used to identify, during inference, graph nodes that correspond to a majority of the anticipated forecasting error, adding a novel interpretable error-tagging capability to baseline models. Ultimately, the objective is to show how augmenting pre-trained baseline GNNs with trainable adaptive pooling modules results in an interpretable fine-tuning framework for mesh-based predictive modeling. Demonstrations are performed using unstructured fluid flow data sourced from a backward-facing step configuration at high Reynolds numbers.

Shivam Barwey
Argonne National Laboratory
Transportation and Power Systems Division
sbarwey@anl.gov

Romit Maulik
Pennsylvania State University
rmaulik@psu.edu

MS30

Computational Hypergraph Discovery: A Gaussian Process Framework to Connect the Dots

Most scientific challenges can be framed as one of the following three levels of complexity of function approximation. Type 1: Approximate an unknown function given input/output data. Type 2: Given partial observations of variables and functions indexed by nodes and hyperedges of a hypergraph, approximate the unobserved variables and unknown functions. Type 3: Given an unknown hypergraph, use partial observations of the hypergraph variables to discover its structure and approximate its unknown functions. While many challenges can be framed as Type 1 and Type 2 problems, others can only be categorized as Type 3. Despite their prevalence, Type 3 challenges have been largely overlooked due to their inherent complexity. Although Gaussian Process (GP) methods are sometimes perceived as well-founded but old technology limited to Type 1 curve fitting, their scope has recently been expanded to Type 2 problems. We introduce an interpretable GP framework for Type 3 problems, targeting data-driven discovery and completion of computational hypergraphs. Our approach is based on a kernel generalization of (1) Row Echelon Form reduction from linear systems to nonlinear ones and (2) variance-based analysis. Here, variables are linked via GPs; those contributing to the highest data variance unveil the hypergraphs structure. We illustrate the scope and efficiency of the proposed approach with applications to network discovery (gene pathways, chemical, mechanical) and raw data analysis.

Theo Bourdais
Caltech
tbourdai@caltech.edu

MS30

Representing Edge Flows on Graphs via Sparse Cell Complexes

Obtaining sparse, interpretable representations of observable data is crucial in many machine learning and signal processing tasks. For data representing flows along the edges of a graph, an intuitively interpretable way to obtain such representations is to lift the graph structure to a simplicial complex: The eigenvectors of the associated Hodge-Laplacian, respectively the incidence matrices of the corresponding simplicial complex then induce a Hodge decomposition, which can be used to represent the observed data in terms of gradient, curl, and harmonic flows. In this paper, we generalize this approach to cellular complexes and introduce the flow representation learning problem, i.e., the problem of augmenting the observed graph by a set of cells, such that the eigenvectors of the associated Hodge Laplacian provide a sparse, interpretable representation of the observed edge flows on the graph. We show that this problem is NP-hard and introduce an efficient approximation algorithm for its solution. Experiments on real-world and synthetic data demonstrate that our algorithm outperforms state-of-the-art methods with respect to approximation error, while being computationally efficient.

Josef Hoppe

RWTH Aachen
Computational Network Science
hoppe@cs.rwth-aachen.de

Michael T. Schaub
RWTH Aachen University
schaub@cs.rwth-aachen.de

MS30

Graph Neural Networks and Applied Linear Algebra

Sparse matrix computations are ubiquitous in scientific computing. Given the recent interest in scientific machine learning, it is natural to ask how sparse matrix computations can leverage neural networks. Unfortunately, multi-layer perceptron (MLP) neural networks are typically not natural for either graph or sparse matrix computations. The issue lies with the fact that MLPs require fixed-sized inputs while scientific applications generally generate sparse matrices with arbitrary dimensions and a wide range of different nonzero patterns (or matrix graph vertex interactions). While convolutional neural networks could possibly address matrix graphs where all vertices have the same number of nearest neighbors, a more general approach is needed for arbitrary sparse matrices, e.g., arising from discretized partial differential equations on structured meshes. Graph neural networks (GNNs) are one such approach suitable for sparse matrices. We discuss the interplay between matrix operations and graph neural networks. In particular, we introduce graph neural networks using common linear algebra operations and express those operations in the graph neural network framework. Concrete examples are given, both where the associated functions are determined a-priori as well as cases where parameters must be learned. It is hoped that this understanding will further stimulate data-driven extensions of classical sparse linear algebra tasks.

Nicholas Moore
West Texas A&M University
College of Engineering
nmoore@wtamu.edu

MS31

Learning Neural Operators with Derivatives

Derivatives (gradient, Jacobian, Hessian, etc) play a critical role in fast and scalable algorithms for solving high-dimensional stochastic optimization problems, Bayesian inverse problems, optimal experimental design problems, etc. To enable accurate approximation of the derivatives of system outputs (states, observables, objectives) with respect to the inputs (optimization variables, uncertain parameters, experimental design variables) by neural operators, we propose to learn the neural operators with the input-output samples augmented with their derivatives. The derivatives can be efficiently computed with proper dimension reduction, randomized algorithms, and adjoints. We demonstrate that such operator learning can not only enhance the accuracy of the input-output map with limited training samples, but also lead to much more accurate derivative information for several class of PDE equations. We demonstrate that such derivative-informed neural operators achieve superior performance for PDE-constrained optimization, Bayesian inverse, and experimental design problems.

Peng Chen
Georgia Institute of Technology
pchen402@gatech.edu

MS31

Exploiting Low-Dimensional Structures in Deep Learning

In the past decade, deep learning has made astonishing breakthroughs in various real-world applications. It is a common belief that deep neural networks are good at learning various geometric structures hidden in data sets. One of the central interests in deep learning theory is to understand why deep neural networks are successful, and how they utilize low-dimensional data structures. In this talk, I will present some statistical learning theory of deep neural networks where data are concentrated on or near a low-dimensional manifold. The learning tasks include regression, classification, feature representation and operator learning. When data are sampled on a low-dimensional manifold, the sample complexity crucially depends on the intrinsic dimension of the manifold instead of the ambient dimension of the data. These results demonstrate that deep neural networks are adaptive to low-dimensional geometric structures of data sets.

Wenjing Liao
Georgia Institute of Technology
wliao60@gatech.edu

MS31

Conformalized-Deeponet: A Distribution-Free Framework for Uncertainty Quantification in Deep Operator Networks

In this talk, we adopt conformal prediction, a distribution-free uncertainty quantification (UQ) framework, to obtain confidence prediction intervals with coverage guarantees for Deep Operator Network (DeepONet) regression. Initially, we enhance the uncertainty quantification frameworks (B-DeepONet and Prob-DeepONet) previously proposed by the authors by using split conformal prediction. By combining conformal prediction with our Prob- and B-DeepONets, we effectively quantify uncertainty by generating rigorous confidence intervals for DeepONet prediction. Additionally, we design a novel Quantile-DeepONet that allows for a more natural use of split conformal prediction. We refer to this distribution-free effective uncertainty quantification framework as split conformal Quantile-DeepONet regression. Finally, we demonstrate the effectiveness of the proposed methods using various ordinary, partial differential equation numerical examples, and multi-fidelity learning.

Guang Lin, Christian Moya, Amir Mollaali
Purdue University
guanglin@purdue.edu, cmoyacal@purdue.edu,
amollaal@purdue.edu

Zecheng Zhang
Florida State University
zz23e@fsu.edu

Lu Lu
Yale University
lu.lu@yale.edu

MS31

Predicting Operators and Symbolic Expressions

Using Multimodal Transformers

We introduce a multi-modal model for scientific problems, named PROSE-PDE. Our model, designed for bi-modality to bi-modality learning, is a multi-operator learning approach which can predict future states of spatiotemporal systems while simultaneously recovering the underlying governing equations of the observed physical system. We focus on training distinct one-dimensional time-dependent nonlinear constant coefficient partial differential equations. In addition, we will discuss some extrapolation studies related to generalizing physical features and predicting PDE solutions whose models or data were unseen during the training. We show through numerical experiments that the utilization of the symbolic modality in our model effectively resolves the well-posedness problems with training multiple operators and thus enhances our model's predictive capabilities.

Hayden Schaeffer
Department of Mathematics
University of California, Los Angeles, U.S.
hayden@math.ucla.edu

MS32

Weak Convexity and Learned Regularization

Traditional regularisers yield theoretical guarantees, but struggle to express complicated patterns. Learned regularisers can express patterns found in data, but often lack theoretical guarantees. How can we get the best of both worlds? One recent approach has been to impose convexity on learned regularisers, via input-convex neural network architectures. However, this is a significant restriction. In this talk, I will present theoretical guarantees for inverse problems with weakly convex regularisation. Furthermore, I will show how such regularisers can be learned, by incorporating our input-weakly-convex neural network architecture into the adversarial regularisation framework. I will present results in a simple denoising problem (which suggest that such adversarial weakly convex regularisers retain the interpretability of unconstrained adversarial regularisers) and in CT reconstruction (wherein adversarial weakly convex regularisation competes with state-of-the-art methods).

Jeremy Budd
California Institute of Technology
jmbudd@caltech.edu

MS32

Thermal Radiance Fields: Regularization for Sensor Fusion

Thermal imaging has a variety of applications, from agricultural monitoring to building inspection to imaging under poor visibility, such as in low light, fog, and rain. However, reconstructing thermal scenes in 3D presents several challenges due to the comparatively lower resolution and limited features present in long-wave infrared (LWIR) images. To overcome these challenges, we propose a unified framework for scene reconstruction from a set of LWIR and RGB images, using a multispectral radiance field with specialized multispectral regularizers. We calibrate the RGB and infrared cameras with respect to each other, as a preprocessing step using a simple calibration target. We demonstrate our method on real-world sets of RGB and LWIR photographs captured from a handheld thermal camera, showing the effectiveness of our method at scene representation across the visible and thermal spectra. We show that our method is capable of thermal super-resolution, as well as visually removing obstacles to reveal objects that are occluded in either the RGB or thermal channels.

Sara Fridovich-Keil
Stanford University and Georgia Institute of Technology, U.S
sarafk@stanford.edu

MS32

Generalization in Diffusion Models Arises from Geometry-adaptive Harmonic Representation

Deep neural networks (DNNs) trained for image denoising are able to generate high- quality samples with score-based reverse diffusion algorithms. These impressive capabilities seem to imply an escape from the curse of dimensionality, but recent reports of memorization of the training set raise the question of whether these networks are learning the true continuous density of the data. Here, we show that two DNNs trained on non-overlapping subsets of a dataset learn nearly the same score function, and thus the same density, when the number of training images is large enough. In this regime of strong generalization, diffusion-generated images are distinct from the training set, and are of high visual quality, suggesting that the inductive biases of the DNNs are well-aligned with the data density. We analyze the learned denoising functions and show that the inductive biases give rise to a shrinkage operation in a basis adapted to the underlying image. Examination of these bases reveals oscillating harmonic structures along contours and in homogeneous regions. We demonstrate that trained denoisers are inductively biased towards these geometry-adaptive harmonic bases since they arise not only when the network is trained on photographic images, but also when it is trained on image classes supported on low-dimensional manifolds for which the harmonic basis is suboptimal. Finally, we show that when trained on regular image classes for which the optimal basis is known to be geometry-adaptive and harmonic, the denoising performance of the networks is near-optimal. In collaboration with: Florentin Guth, Eero Simoncelli, Stephane Mallat

Zahra Kadkhodaie
New York University
zk388@nyu.edu

MS32

On Optimal Methods for Inverse Problems with Low-Dimensional Models

The 1-norm was proven to be a good convex regularizer for the recovery of sparse vectors from under-determined linear measurements. With an appropriate measurement operator, a number of measurements of the order of the sparsity of the unknown (up to log factors) is sufficient for stable and robust recovery. Such recovery results can be generalized to more general low-dimensional model sets and (convex) regularizers. These results lead to the following question: to recover a given low-dimensional model set from linear measurements, what is the "best" (convex) regularizer? To approach this problem, we propose a general framework to define the notion of "best" regularizer with respect to a low-dimensional model. We propose to maximize a compliance measure that quantifies the adequacy between the regularizer and the model set. We give optimality results for sparse and low-rank models and some of

their generalizations, such as sparsity in levels and sparse plus low-rank models; with application to inverse problems such as structure-texture decomposition of images.

Yann Traonmilin
IMB, CNRS, Université de Bordeaux
yann.traonmilin@u-bordeaux.fr

MS33

Diffusions for Learning Gaussian Mixtures

I will describe recent progress on providing rigorous convergence guarantees for diffusion models, which constitute the backbone of large-scale real-world generative models such as DALLE 3 and Sora. At the beginning, I will survey some recent results showing that these models can efficiently sample from essentially any realistic data distribution, even ones which are highly non-log-concave, provided one can perform score estimation. In the main part of the talk, I will apply this machinery to the classic question of learning Gaussian mixture models. Prior methods based on the method of moments incur doubly exponential dependence on the number of components. I will present a new algorithm based on diffusions and piecewise polynomial regression that circumvents this and achieves runtime that qualitatively matches existing computational lower bounds.

Sitan Chen
Harvard
sitan@seas.harvard.edu

Vasilis Kontonis, Kulin Shah
UT Austin
vkonton@gmail.com, kulin.shah98@gmail.com

MS33

Zeroth-Order Sampling Methods for Non-Log-Concave Distributions: Alleviating Metastability by Denoising Diffusion

This paper considers the problem of sampling from nonlogconcave distribution, based on queries of its unnormalized density. It first describes a framework, Diffusion Monte Carlo (DMC), based on the simulation of a denoising diffusion process with its score function approximated by a generic Monte Carlo estimator. DMC is an oracle-based meta-algorithm, where its oracle is the assumed access to samples that generate a Monte Carlo score estimator. Then we provide an implementation of this oracle, based on rejection sampling, and this turns DMC into a true algorithm, termed Zeroth-Order Diffusion Monte Carlo (ZOD-MC). We provide convergence analyses by first constructing a general framework, i.e. a performance guarantee for DMC, without assuming the target distribution to be log-concave or satisfying any isoperimetric inequality. Then we prove that ZOD-MC admits an inverse polynomial dependence on the desired sampling accuracy, albeit still suffering from the curse of dimensionality. Consequently, for low dimensional distributions, ZOD-MC is a very efficient sampler, with performance exceeding latest samplers, including also-denoising-diffusion-based RDMC and RS-DMC. Last, we experimentally demonstrate the insensitivity of ZOD-MC to increasingly higher barriers between modes or discontinuity in non-convex potential.

Ye He, Kevin Rojas
Georgia Institute of Technology
yhe367@gatech.edu, kevin.rojas@gatech.edu

Molei Tao, Molei Tao
School of Mathematics, Georgia Institute of Technology
mtao@gatech.edu, mtao@gatech.edu

MS33

Provable Sampling from Multimodal Distributions: From Markov Chains to Diffusion Models

We consider two settings where we obtain provable guarantees for sampling from multimodal distributions, given access to the unnormalized density or gradient log-pdf (score function). Our results are general and apply to both Langevin and Glauber dynamics. In the first setting, for a distribution with k modes, we show that given $\mathrm{poly}(k/\varepsilon)$ samples from the distribution, we can obtain an arbitrary number of samples that are ε-close in TV distance from the target distribution. This is robust to L^2 error in the score function estimate, and hence applies to the setting of score-based generative models (diffusion models), where the score function is learned from samples. Here, our results show that for distributions with bounded number of modes, learning the "vanilla" score is sufficient for efficient sampling. In the second setting, we do not assume we have samples from the underlying distribution, but rather "warm starts" for each of the modes. Under some technical conditions on the modes, we show that an algorithm based on the Annealed Leap-Point Sampler (ALPS) combining simulated tempering and teleporting allows efficient polynomial-time sampling from the target distribution.

Holden Lee
Johns Hopkins University
hlee283@jhu.edu

MS33

Optimal Score Estimation via Empirical Bayes Smoothing

We study the problem of estimating the score function of an unknown probability distribution ρ^* from n independent and identically distributed observations in d dimensions. Assuming that ρ^* is subgaussian and has a Lipschitz-continuous score function s^*, we establish the optimal rate of $\tilde{\Theta}(n^{-\frac{2}{d+4}})$ for this estimation problem under the loss function $\|\hat{s} - s^*\|^2_{L^2(\rho^*)}$ that is commonly used in the score matching literature, highlighting the curse of dimensionality where sample complexity for accurate score estimation grows exponentially with the dimension d. Leveraging key insights in empirical Bayes theory as well as a new convergence rate of smoothed empirical distribution in Hellinger distance, we show that a regularized score estimator based on a Gaussian kernel attains this rate, shown optimal by a matching minimax lower bound. We also discuss the implication of our theory on the sample complexity of score-based generative models.

Andre Wibisono
Yale University
andre.wibisono@yale.edu

Yihong Wu
Department of Statistics and Data Science
Yale University
yihong.wu@yale.edu

Kaylee Yang
Yale University

yingxi.yang@yale.edu

MS34

Deep Learning for Early Warning Signals in Cardiac Systems

The human heart is a complex system that can undergo a critical transition to an abnormal rhythm, known as a cardiac arrhythmia. How to predict or assess the risk of cardiac arrhythmia in individual patients with heart disease is not clear. In this talk, I will demonstrate how deep learning can be combined with mathematical models of the heart to (i) improve prediction of an arrhythmia known as alternans, and (ii) discover mechanisms that can lead to arrhythmia. We validate these methods in vitro using heart cell aggregates and monolayers. I will argue that the rapid development cardiac monitoring techniques is giving rise to exciting opportunities at the interface of cardiology and mathematics.

Thomas M. Bury
McGill University
thomas.bury@mcgill.ca

MS34

Foundation Models for Physiological Data

Physiological waveforms such as ECG are ubiquitous in intensive care (ICU) settings and are also playing an increasingly broader role in monitoring human health because of the increase in the acceptance of wearable devices to monitor and assess various health conditions on a day-to-day basis. Even though ICU waveforms are different from wearable data in terms of particular devices and body sites to collect them, our research and others have shown that deep neural networks trained with ICU waveforms perform well in analyzing wearable data largely ICU waveforms are enriched with clinical information captured in EHR offering an opportunity to unlock a suite of physiological (e.g., estimation of blood pressure) and biological (bloodless test to estimate plasma potassium) measurement capabilities that can be realizable from using wearable data. We propose foundation models (FMs) are a compelling technique to realize this vision. These models are trained with massive amount of data in a self-supervised way to learn versatile representation that can be fine-tuned to be performant in downstream tasks. If successful, they can be applied to process wearable data to provide ICU-level monitoring of physiological and biological functions in ambulatory settings. In this talk, I will share our ongoing investigations of different model architectures and learning algorithms in training these FMs and their applications in unlocking new physiological and biological measurement capacities.

Xiao Hu
Emory
Nell Hodgson Woodruff School of Nursing
xiao.hu@emory.edu

MS34

Early Dynamical Signs of Multiple Sclerosis and Temperature Dysregulation.

Early diagnosis of multiple sclerosis is difficult since demyelination is patchy, affecting only a few axons. We first present older work on a highly sensitive diagnosis of optic neuritis (a common first sign of MS) using infrared pupillography. We then report our recently proposed framework that predicts the form of the compound action potential down an axon bundle for a given probability of demyelination. Our findings articulate why action potential jitter and compound action potential dispersion can serve as potential markers of weak and sporadic demyelination. We further present a CNN-based analysis of spinal cord fiber tract images obtained from CARS spectroscopy in mice with ECE, an experimentally induced form of MS. Our classifier accurately predicts the clinical score from the spinal images. Effects of temperature on MS are also discussed. The talk will then move to human temperature management during exercise as a probe of fitness, aging factors, and diabetes. A state-of-the-art calorimeter quantifies internal heat and entropy production, along with external heat and entropy dissipation. Continual monitoring of heat and entropy reveals imbalances related to the aforementioned conditions. Data science applied to the measured physiological variables could detect minute defects in this management, and better ground our 2015 statistical study of a large array of nonlinear time series metrics that enabled earlier detection of sepsis onset in patients.

Andre Longtin
University of Ottawa
Physics Department
alongtin@uottawa.ca

MS34

DNB-Based Gene Intervention Design for Early Medical Treatment

The Dynamical Network Bioarker (DNB) theory provides an effective way to detect early-warning signals of critical transitions to disease stage in gene network systems only with gene expression's High-Dimension Low-Sample-Size (HDLSS) data. It is essential to consider how to prevent the disease at an early stage, which is called early medical treatment. Towards giving a theoretical foundation for early medical treatment with HDLSS data, this talk introduces a computational method to learn the approximate pole placement for re-stabilizing gene networks, which essentially stops the progress of the disease. The proposed method extracts information about a gene network system from HDLSS data analysis. It utilizes the extracted information to design the input placement and feedback gain for approximate pole placement, which specifies which gene should be intervened in and how we should intervene in them. In the proposed method, only the diagonal elements of the system matrix need to be adjusted, which reduces the complexity of realizing the re-stabilization and makes the proposed method more practical. We will take an overview of the advanced and most recent results in the above approach, the future prospects, and the open challenges.

Xun Shen
Osaka University
xun.shen.whu@gmail.com

MS35

Deep Learning of Diffeomorphisms with Application to Shape Analysis

In this talk, we discuss structure preservation and deep learning with applications to shape analysis. This is a framework for treating complex data and obtaining metrics on spaces of data. Examples are spaces of unparametrized curves, time-signals, surfaces, and images. A computationally demanding task for estimating distances between

shapes, e.g. in object recognition, is the computation of optimal reparametrizations. This is an optimization problem on the infinite-dimensional group of orientation-preserving diffeomorphisms. We approximate diffeomorphisms with neural networks and use the optimal control and dynamical systems point of view to deep learning. We will discuss useful geometric properties in this context e.g. reparametrization invariance of the distance function, invertibility, and contractivity of the neural networks. We will consider the theory and applications of these ideas.

Elena Celledoni
Norwegian University of Science and Technology
elena.celledoni@ntnu.no

MS35

Enhancing Subsurface Surrogate Efficiency: Integrating Improved Neural Operators with Graph Convolutional Networks

This study introduces the Improved Neural Operator (INO) framework [1] with graph, a novel approach designed to enhance the efficiency of Reduced Order Models (ROMs) in large-scale subsurface modeling. Due to their computational complexity, Full Order Models (FOMs) face computational challenges in field-scale parameter estimation and real-time control. INO addresses this by dividing the computational domain into smaller subdomains during training, significantly reducing computational demands without sacrificing accuracy. The methodology was tested on the Illinois Basin - Decatur Project, showing less than 1% error in pressure predictions with a small dataset. Despite its effectiveness, the INO model's subsampling strategy has limitations in capturing global phenomena within geological fields. To overcome this, we integrate Graph Convolutional Networks (GCNs) with the INO, enhancing its ability to represent complex subsurface environments comprehensively. This integration marks a significant advancement in geological storage modeling, offering a scalable, accurate, and computationally efficient solution for the industry.

Teeratorn Kadeethum
Sandia National Laboratories
tkadeet@sandia.gov

Birendra Jha
Mork Family Dept. of Chemical Eng. and Materials Sci.
University of Southern California
bjha@usc.edu

Stephen Verzi
Sandia National Laboratories
sjverzi@sandia.gov

John D. Jakeman
Sandia National Labs
jdjakem@sandia.gov

Hongkyu Yoon
Geoscience Research and Applications
Sandia National Laboratories
hyoon@sandia.gov

MS35

Data Driven Modeling of Unknown Systems with Deep Neural Networks

We present a framework of predictive modeling of unknown systems from measurement data. The method is designed to discover/approximate the unknown evolution operator, i.e., flow map, behind the data. Deep neural network (DNN) is employed to construct such an approximation. Once an accurate DNN model for the evolution operator is constructed, it serves as a predictive model for the unknown system and enables us to conduct system analysis. We demonstrate that flow map learning (FML) approach is applicable for modeling a wide class of problems, including dynamical systems, systems with missing variables and hidden parameters, as well as partial differential equations (PDEs).

Dongbin Xiu
Ohio State University
xiu.16@osu.edu

MS35

Training-Free Diffusion Models for Supervised Learning of Generative Models in Density Estimation

We present a supervised learning framework of training generative models for density estimation. Generative models are usually considered as unsupervised learning models. Despite the success of the generative models, there are several issues. To enable supervised learning in generative models, we utilize the score-based diffusion model to generate labeled data. Unlike existing diffusion models that train neural networks to learn the score function, we develop a training-free score estimation method. This approach uses mini-batch-based MC estimators to directly approximate the score function at any spatial-temporal location in solving an ODE, corresponding to the reverse-time SDE. This approach can offer both high accuracy and substantial time savings in NN training. Once the labeled data are generated, we can train a simple fully connected neural network to learn the generative model in the supervised manner. Compared with existing normalizing flow models, our method does not require to use reversible NNs and avoids the computation of the Jacobian. Compared with existing diffusion models, our method does not need to solve the reverse-time SDE to generate new samples. As a result, the sampling efficiency is significantly improved. We demonstrate the performance of our method by applying it to real data from the UCI repository, as well as Bayesian inference problems associated with the terrestrial ecosystem models in E3SM.

Guannan Zhang
Oak Ridge National Laboratory
zhangg@ornl.gov

MS36

Learning Efficient Representations of Environmental Priors in Working Memory

Experience shapes our expectations and helps us learn the structure of the environment. Inference models render such learning as a gradual refinement of the observers estimate of the environmental prior. For instance, when retaining an estimate of an objects features in working memory, learned priors may bias the estimate in the direction of common feature values. Humans display such biases when retaining color estimates on short time intervals. We propose that these systematic biases emerge from modulation of

synaptic connectivity in a neural circuit based on the experienced stimulus history, shaping the persistent and collective neural activity that encodes the stimulus estimate. Resulting neural activity attractors are aligned to common stimulus values. Using recently published human response data from a delayed-estimation task in which stimuli (colors) were drawn from a heterogeneous distribution that did not necessarily correspond with reported population biases, we confirm that most subjects response distributions are better described by experience-dependent learning models than by models with fixed biases. This work suggests systematic limitations in working memory reflect efficient representations of inferred environmental structure, providing new insights into how humans integrate environmental knowledge into their cognitive strategies.

Tahra Eissa
University of Colorado Boulder
tahra.eissa@colorado.edu

Zachary P. Kilpatrick
University of Colorado, Boulder
zpkilpat@colorado.edu

MS36

Data Augmented Machine Learning to Improve Detection of Epilepsy Biomarkers in Brain Voltage Recordings

Epilepsy is a major neurological disorder characterized by recurrent, spontaneous seizures. For patients with drug-resistant epilepsy, treatments include neurostimulation or surgical removal of the epileptogenic zone (EZ), the brain region responsible for seizure generation. Precise targeting of the EZ requires reliable biomarkers. Spike ripples - high-frequency oscillations that co-occur with large amplitude epileptic discharges - have gained prominence as a candidate biomarker. However, spike ripple detection remains a challenge. The gold-standard approach requires an expert manually visualize and interpret brain voltage recordings, which limits reproducibility and high-throughput analysis. Addressing these limitations requires more objective, efficient, and automated methods for spike ripple detection, including approaches that utilize deep neural networks. Despite advancements, dataset heterogeneity and scarcity severely limit machine learning performance. Our study explores long-short term memory (LSTM) neural network architectures for spike ripple detection, leveraging data augmentation to improve classifier performance. We highlight the potential of combining training on augmented and in vivo data for enhanced spike ripple detection and ultimately improving diagnostic accuracy in epilepsy treatment.

Emily Schlafly, Daniel Carbonero
Boston University
eds2@bu.edu, dcarbo@bu.edu

Catherine Chu
Harvard Medical School
Massachusetts General Hospital
cjchu@mgh.harvard.edu

Mark Kramer
Boston University
mak@math.bu.edu

MS36

Spatiotemporal Dynamics in Neural Systems: from Data to Mathematical Models and Computation

Neurons in cortex are connected in intricate patterns, with local- and long-range connections and distance-dependent time delays for transmitting signals. In recent work, we have found that spontaneous and stimulus-driven waves of neural activity travel over these networks, modulating both spiking activity and sensory processing as they pass. Understanding how the networks of cortex generate these sophisticated dynamics, however, remains an open problem. This is due, in part, to the fact that connecting the specific structure of networks to the resulting nonlinear dynamics is a difficult problem in general. This becomes even more challenging when considering axonal time delays, as we need to consider systems with many time delays. In this talk, I will present results connecting the structure of individual networks to the resulting dynamics in systems of nonlinear Kuramoto oscillators. We introduce a complex-valued approach that allows linking the precise structure of connections in the network to the spatiotemporal patterns that will occur in individual simulations. This approach allows understanding these activity patterns through an analytical form related to the eigenspectrum of the graph adjacency matrix. This, in turn, leads to predictions for the precise traveling wave patterns that will emerge in these systems. Finally, I will present recent work using this approach to understand computation with spatiotemporal dynamics.

Lyle Muller
Department of Mathematics
Western University
lmuller2@uwo.ca

MS36

Data Science Meets Neuroscience: An Overview

The decade following the launch of the NIH Brain Initiative has seen an explosion in the collection, curation, and sharing of data from a wide array of neural systems. This talk will introduce the special session as well as highlight a new initiative: Applying data-driven Koopman operator theoretic approaches, together with machine learning techniques, to study both fully and partially observed networks of coupled stochastic oscillators, as a framework for analysis of human EEG data.

Peter J. Thomas
Case Western Reserve University
pjthomas@case.edu

Rodica Curtu
University of Iowa
Department of Mathematics
rodica-curtu@uiowa.edu

Yao Li
University of Massachusetts Amherst
Department of Mathematics and Statistics
liyao@umass.edu

Pake Melland
Oregon Institute of Technology

pake.melland@oit.edu

MS37

Deterministic and Stochastic Moving Anchor Extragradient Algorithms for Structured Saddlepoint Problems

Our work introduces a moving anchor technique to extragradient algorithms for smooth structured minimax problems. First, our moving anchor technique is introduced into the original algorithmic anchoring framework known as EAG. We match the optimal order of convergence in terms of worst-case complexity on the squared gradient norm, $O(1/k^2)$. As many problems of practical interest are nonconvex-nonconcave, the recently developed FEG class of algorithms brings order-optimal anchoring methods developed within EAG to certain nonconvex-nonconcave problem settings. We introduce the moving anchor methods to the FEG class of algorithms and again obtain order-optimal complexity results. Extensions include a preconditioned version of our algorithms, as well as newly developed stochastic moving anchor methods for convex-concave problems. In both convex-concave and nonconvex-nonconcave settings, a variety of numerical examples demonstrate the efficacy and flexibility of the moving anchor framework over its fixed-anchor counterparts. Future directions of the moving anchor framework and other applications are also discussed.

James Alcala
University of Southern California, U.S.
jameska30@gmail.com

Yat Tin Chow
University of California, Riverside
yattinc@ucr.edu

Mahesh Sunkula
Purdue University
msunkula@purdue.edu

MS37

Mean-Field Control Barrier Functions: A Framework for Real-Time Swarm Control

Control Barrier Functions (CBFs) are an effective methodology to assure safety and performative efficacy in real-time control applications such as power systems, resource allocation, autonomous vehicles, robotics, etc. This approach ensures safety independently of the high-level tasks that may have been pre-planned off-line. For example, CBFs can be used to guarantee that a self-driving car will remain in its lane. However, when the number of agents is large, computation of CBFs can suffer from the curse of dimensionality in the multi-agent setting. In this work, we present Mean-field Control Barrier Functions (MF-CBFs), which extends the CBF framework to the mean-field (or swarm control) setting. The core idea is to model swarms as probability measures in the state space and build control barrier functions in the space of probability measures. Similar to traditional CBFs, we derive safety constraints on the (distributed) controls but now relying on the differential calculus in the space of probability measures. Our numerical experiments show the effectiveness of MF-CBFs applied to swarm tracking and avoidance.

Samy Wu Fung
Colorado School of Mines
swufung@mines.edu

MS37

Decoding Mean Field Games from Population and Environment Observations by Gaussian Processes

This talk presents a Gaussian Process (GP) framework, a non-parametric technique widely acknowledged for regression and classification tasks, to address inverse problems in mean field games (MFGs). By leveraging GPs, we aim to recover agents' strategic actions and the environment's configurations from partial and noisy observations of the population of agents and the setup of the environment. Our method is a probabilistic tool to infer the behaviors of agents in MFGs from data in scenarios where the comprehensive dataset is either inaccessible or contaminated by noises.

Xianjin Yang
California Institute of Technology
yxjmath@caltech.edu

Jinyan Guo
National University of Singapore
e0708165@u.nus.edu

Chenchen Mou
City University of Hong Kong
chencmou@cityu.edu.hk

Chao Zhou
Department of Mathematics
National University of Singapore
matzc@nus.edu.sg

MS37

A Mean-Field Games Laboratory for Generative Modeling

We demonstrate the versatility of mean-field games (MFGs) as a mathematical framework for explaining, enhancing, and designing generative models. We establish connections between MFGs and major classes of flow and diffusion-based generative models by deriving continuous-time normalizing flows, score-based models, and Wasserstein gradient flows through different choices of particle dynamics and cost functions. Furthermore, we study the mathematical structure and properties of each generative model by examining their associated MFG's optimality condition, which consist of a set of coupled forward-backward nonlinear partial differential equations. The optimality conditions of MFGs also allow us to introduce HJB regularizers for enhanced training of a broad class of generative models. We present this framework as an MFG laboratory which serves as a platform for revealing new avenues of experimentation and invention of generative models.

Benjamin J. Zhang
University of Massachusetts Amherst
bjzhang@umass.edu

Markos A. Katsoulakis
University of Massachusetts, Amherst
Dept of Mathematics and Statistics

markos@math.umass.edu

MS38

Topological Embedding of Brain Networks

We explore the dynamic behavior of brain networks through the lens of topological data analysis, centering on a recently developed embedding method known as the Topological Phase Diagram (TPD). This method leverages the robustness of persistent homology to capture and visualize the dynamic evolution of brain networks. By employing the Wasserstein distance as a metric, TPD facilitates a meaningful low-dimensional representation of high-dimensional topological data, allowing for the orthogonal projection of 0D and 1D topological features. This embedding not only quantifies the dynamics of individual human brain networks during resting states but also enhances the comparability across subjects and conditions without the need for temporal synchronization. The method is applied to understanding the dynamics of brain networks of 101 patients with temporal lobe epilepsy (TLE), revealing that while TLE networks exhibit increased rigidity in their 0D topology, they display more rapid fluctuations in the 1D topology compared to those of healthy controls. The talk is partially based on arXiv:2201.00087 (PLOS Computational Biology).

Moo K. Chung
Department of Biostatistics and Medical Informatics
University of Wisconsin-Madison
mkchung@wisc.edu

MS38

Persistent Directed Flag Laplacian

Topological data analysis (TDA) has had enormous success in science and engineering in the past decade. Persistent topological Laplacians (PTLs) overcome some limitations of persistent homology, a key technique in TDA, and provide substantial insight to the behavior of various geometric and topological objects. This work extends PTLs to directed flag complexes, which are an exciting generalization to flag complexes, also known as clique complexes, that arise naturally in many situations. We introduce the directed flag Laplacian and show that the proposed persistent directed flag Laplacian (PDFL) is a distinct way of analyzing these flag complexes. Example calculations are provided to demonstrate the potential of the proposed PDFL in real world applications.

Benjamin D. Jones
Michigan State University
jones657@msu.edu

MS38

Emerging Mathematics and Geometric Deep Learning for Drug Design

The intersection of mathematics and artificial intelligence (AI) has ushered in a new era of drug design, offering unprecedented accuracy and efficiency in identifying potential drug candidates. This talk focuses on mathematics representation learning and geometric deep learning, rapidly advancing areas within machine learning and data mining that specialize in processing graph-structured and 3D data. Our exploration includes the latest developments in differential geometry, persistent spectral graphs enhanced with geometric deep learning, and advanced large language models. These tools have proven instrumental in characterizing biomolecular and molecular interactions. A standout feature of our approach is its scalability, which accommodates diverse molecular representations, and its robustness, especially when handling low-quality data. These strengths have catapulted our models to the forefront, as evidenced by our top-tier performance in the small molecular properties benchmarks, protein-protein interactions, and the D3R grand challenges. This presentation will offer insights into the future trajectory of drug design, where the synergy of mathematics and AI is set to redefine the boundaries of what is achievable in pharmaceutical research.

Duc Nguyen
University of Kentucky
ducnguyen@uky.edu

MS38

Persistent Dirac of Paths on Digraphs and Hypergraphs

Significant progress has been made in advancing the theoretical underpinnings of topological data analysis (TDA) by conceptualizing and refining the Dirac operator, a fundamental mathematical construct used in analyzing topological signals and molecular representations. However, the current methodologies mostly rely on traditional frameworks such as Vietoris-Rips complexes and alpha complexes. This academic pursuit aims to introduce a novel methodological approach by integrating Dirac operators with path homology within the framework of TDA, with a specific emphasis on elucidating molecular structures. Path homology is a theoretical construct that extends the domain of graph homology to accommodate directed graphs and hypergraphs. Thus, it provides a more comprehensive analytical toolkit conducive to interrogating structures endowed with directional attributes. The proposed methodological framework involves extracting features from such complex structures, which are examined by various exemplary instances. This work seeks to contribute to the ongoing discourse within TDA by advancing methodologies that enhance our understanding and analytical capabilities concerning intricate molecular architectures. Thus, it aims to foster broader applicability and robustness in data-driven investigations within computational biology and related disciplines.

Faisal Suwayyid
Michigan State University
suwayyid@msu.edu

MS39

Bandit Learning-Based Exploration and Exploitation for Multi-Level and Multi-Fidelity Scientific Computing under Budget Constraints

Modern simulation-based scientific models are complex and multi-faceted, involving computationally demanding physics-based modeling and discretization, reliable and robust data assimilation, and an accurate accounting for uncertainty in the face of unknown model parameter values and/or genuine stochasticity. To meet such demands, many simulations of real-world systems often involve combinations of reduced order models of various fidelities, but these competing models often have a query cost versus accuracy payoff that is opaque. This in turns complicates the problem of how to allocate computational resources that collect data from these models. To address this problem, we discuss the application of exploration-exploitation meta-

algorithms from bandit learning under computational budget constraints: An exploration phase is devoted to learning about model relationships and interactions, followed by an exploitation phase that uses information learned in exploration to make decisions about optimal model selection and subsequently provides full distributional information of stochastic outputs. This framework leads to flexible procedures that automatically determine how to collect data from ensembles of models. We will identify existing theoretical guarantees for such procedures along with promising and impactful directions for new analysis and algorithm development.

Akil Narayan
University of Utah
akil@sci.utah.edu

MS39

Generalizing Bayesian Optimization to Accelerate Scientific Discovery

Bayesian optimization (BO) is a popular method for efficiently inferring optima of an expensive black-box function via a sequence of queries. Existing information-theoretic BO procedures aim to make queries that most reduce the uncertainty about optima, where the uncertainty is captured by Shannon entropy. However, many tasks in the physical sciences—including advanced materials design and optimization of scientific machines—often require going beyond simple optimization to meet more complex or specialized goals. In this talk, we discuss two frameworks that extend the scope of information-theoretic BO. First, we consider a generalization of Shannon entropy from work in statistical decision theory (DeGroot 1962, Rao 1984), which contains a broad class of uncertainty measures parameterized by a problem-specific loss function corresponding to a downstream task; we show that this entropy yields a flexible family of acquisition functions that can be customized for use in novel optimization settings. We then describe an information-based framework in which a user can describe a custom optimization goal via an algorithm; this allows users to automatically convert a complex, targeted experimental goal into an adaptive data collection strategy. We show how these methods allow practitioners to more-easily develop procedures to collect valuable data for a given task, and demonstrate their use in real-world design and optimization procedures being carried out in national labs.

Willie Neiswanger
Stanford University
willie.neiswanger@gmail.com

MS39

Optimal Sampling for Least Squares Approximation with General Dictionaries

We consider the problem of approximating an unknown function in a nonlinear model class from point evaluations. When obtaining these point evaluations is costly, minimising the required sample size becomes crucial. Recently, an increasing focus has been on employing adaptive sampling strategies to achieve this. These strategies are based on linear spaces related to the nonlinear model class, for which the optimal sampling measures are known. However, the resulting optimal sampling measures depend on an orthonormal basis of the linear space, which is known rarely. Consequently, sampling from these measures is challenging in practice. We present a sampling strategy that iteratively refines an estimate of the optimal sampling measure by updating it based on previously drawn samples. This strategy can be performed offline and does not require evaluations of the sought function. We establish convergence and illustrate the practical performance through numerical experiments. Comparing the presented approach with standard Monte Carlo sampling demonstrates a significant reduction in the samples required to achieve a stable approximation.

Philipp Trunschke
Nantes Université
philipp.trunschke@univ-nantes.fr

MS39

A Competitive Algorithm for Agnostic Active Learning

For some hypothesis classes and input distributions, active agnostic learning needs exponentially fewer samples than passive learning; for other classes and distributions, it offers little to no improvement. The most popular algorithms for agnostic active learning express their performance in terms of a parameter called the disagreement coefficient, but it is known that these algorithms are inefficient on some inputs. We take a different approach to agnostic active learning, getting an algorithm that is competitive with the optimal algorithm for any binary hypothesis class H and distribution $\mathcal{D}_X$ over X. In particular, if any algorithm can use m^* queries to get $O(\eta)$ error, then our algorithm uses $O(m^* \log H)$ queries to get $O(\eta)$ error. Our algorithm lies in the vein of the splitting-based approach of [Dasgupta 2004], which gets a similar result for the realizable ($\eta = 0$) setting. We also show that it is NP-hard to do better than our algorithm's $O(\log H)$ overhead in general.

Eric Price
The Unviersity of Texas at Austin
ecprice@cs.utexas.edu

Yihan Zhou
University of Texas at Austin
joeyzhou@cs.utexas.edu

MS40

Exploring Low-Dimensional Data Structures by Neural Networks with Applications on Learning Operators

Autoencoders have demonstrated remarkable success in learning low-dimensional latent features from high-dimensional data across various applications. However, there are relatively few theoretical results concerning their performance. In our work, we study the chart autoencoder, which encodes data into low-dimensional latent features on a collection of charts, and we establish statistical guarantees for the generalization error of chart autoencoders. We demonstrate that chart autoencoders can effectively denoise data corrupted with normal noise and that the squared generalization error depends only on the intrinsic dimension of the dataset. Furthermore, we apply the autoencoder to model reduction and investigate Autoencoder-based Neural Networks (AENet) for operator learning. We establish a mathematical and statistical estimation theory that analyzes the generalization error of AENet, showing that the sample complexity of training AENet is intricately linked to the intrinsic dimension of the process being modeled. Our results are validated by

numerical experiments.

Hao Liu
Hong Kong Baptist University
haoliu@hkbu.edu.hk

MS40

Learning Curves for Gaussian Process Regression with Power-Law Priors and Targets

We characterize the power-law asymptotics of learning curves for Gaussian process regression (GPR) under the assumption that the eigenspectrum of the prior and the eigenexpansion coefficients of the target function follow a power law. Under similar assumptions, we leverage the equivalence between GPR and kernel ridge regression (KRR) to show the generalization error of KRR. Infinitely wide neural networks can be related to GPR with respect to the neural network GP kernel and the neural tangent kernel, which in several cases is known to have a power-law spectrum. Hence our methods can be applied to study the generalization error of infinitely wide neural networks. The talk is based on work with Hui Jin and Pradeep Banerjee.

Guido F. Montufar
UCLA
guidomontufar@gmail.com

MS40

Kernel Regression on Data with Low Intrinsic Dimensionality: the Optimal Adaptive Posterior Contraction Rate

Bayesian regression problems are widely encountered in various applications. We focus on the scenario when data samples (predictors) have low-dimensional structures, and theoretically study the posterior contraction rate under a Gaussian process (GP) prior. The covariance kernel is the Gaussian kernel in the ambient space, which can be computed from Euclidean distances among data samples. When the data domain is a d-dimensional manifold, we prove the optimal posterior contraction rate as $O(n^{-s/(2s+d)})$, up to a logarithmic factor, when the true regression function is in an s-Hölder class intrinsically defined. In addition, we propose a new Bayes kernel bandwidth estimator using count statistics and a k-nearest neighbor (KNN) algorithm, which allows to compute the GP prior without knowing the intrinsic dimensionality. The theory extends to more general low-dimensional sets where intrinsic dimensionality is characterized by covering number.

Xiuyuan Cheng
Duke University
xiuyuan.cheng@duke.edu

Nan Wu
The University of Texas at Dallas
nan.wu@duke.edu

MS40

Statistical Guarantees of Learning under Invariance

In this talk, we will discuss how (and to what extent) incorporating intrinsic structures, such as group symmetries, into learning tasks can enhance data efficiency. In particular, we will explore the precise reduction in sample complexitynamely, the required number of samples to effectively "learn" the targetachieved by preserving such structural characteristics within the model. Intriguingly, the findings reveal that the outcomes are not as intuitive as one might expect and, in certain instances, prove to be counterintuitive.

Wei Zhu
UMass Amherst
weizhu@gatech.edu

MS41

Data Representation with Optimal Transport

Finding mathematical formulas to represent data (such as functions, vectors, measures, etc.) has long been of interest to mathematicians and various sciences. In this presentation, we will describe an emerging non-linear framework for data representation based on transporting or matching one measure, distribution, or function to another, utilizing the theory of Optimal Transport (OT). The crucial point lies in understanding that the space of probability measures, endowed with the Wasserstein distance, is not only a metric and geodesic space but also a Riemannian manifold: We will consider a reference measure, usually the uniform measure, and the tangent space at that measure, and we will then translate the Euclidean distance in this linear space to the Wasserstein manifold, resulting in a new metric (the Linear Optimal Transport distance). To define this distance, we will introduce an embedding or transformation based on OT, known as the Linear Optimal Transport Embedding. Using probability measures defined on the unit circle, we will demonstrate how these tools work through various experiments, including image retrieval, interpolation, and classification. Finally, we will explore the limitations posed by balanced mass in classical OT, delving into Partial Optimal Transport and Unbalanced Optimal Transport. This talk will cover aspects of joint works with Y. Bai, S. Kolouri, I. Medri, G. Rohde, and the MINT Lab at Vanderbilt University.

Rocío Diaz Martin
Vanderbilt University
rocio.p.diaz.martin@vanderbilt.edu

MS41

Manifold Learning in Wasserstein Space

We will discuss the theoretical foundations for manifold learning algorithms in the space of absolutely continuous probability measures on compact, convex subsets of R^d, metrized by the quadratic Wasserstein distance. We begin by introducing a natural construction of submanifolds M equipped with a metric given by the geodesic restriction of the Wasserstein metric to the manifold, d_M. In contrast to other constructions, these submanifolds are not necessarily flat, but still allow for local linearizations in a similar fashion to Riemannian submanifolds of R^d . We then show how the latent manifold structure can be learned from discrete samples and pairwise extrinsic Wasserstein distances. In particular, we show that the metric space (M, d_M) can be asymptotically recovered in the sense of GromovWasserstein from a graph over the discrete samples with edge weights being the Wasserstein distance between the samples. In addition, we demonstrate how the tangent space at a sample can be asymptotically recovered via an analogue of local principal component analysis. We will illustrate the theory with explicit constructions of submanifolds and

some numerical experiments of tangent space recovery.

Keaton Hamm
University of Texas at Arlington
keaton.hamm@uta.edu

MS41

Approximation and Registration with Slice-Matching Transport

Iterative slice-matching procedures are efficient schemes for transferring a source measure to a target measure, especially in high dimensions. Given the computational challenges of optimal transport, understanding the effectiveness of slice-matching maps (one-step of such schemes) as potential alternatives is important, especially under structural assumptions between the measures. In this talk, we delve into the exact recovery of translations and scalings, as well as the approximate recovery of perturbations of such transformations. We demonstrate an invariance property with respect to the source measure, an equivariance property with respect to the target measure, and Lipschitz continuity concerning the slicing directions. We also provide a quantitative perspective on how slice-matching procedures encode special affine transformations in their approximations through the study of basic registration problems. This talk is based on joint work with Caroline Moosmueller.

Shiying Li
University of North Carolina - Chapel Hill
shiyl@unc.edu

MS41

Towards Predictive Modeling in Wasserstein Spaces: a Signal Transformation Approach

In contrast to traditional signal detection and estimation theory used in communications technology, for example, where control over the data (signal) generation process allows us precise knowledge of the geometry encapsulating the data (i.e. curvature, Fisher information criteria) and allows us to define optimal performance (e.g. Cramer-Rao bounds), the relative lack of of understanding regarding geometry of datasets often doesn't allow us to understand when (for which samples) a particular ML solution will work, or fail. This talk will demonstrate how adopting different metrics and geometries can have a large impact in the solutions provided by ML approaches. In particular, we show that Wasserstein metric spaces can render ML approaches significantly more accurate and robust especially in applications dealing with data emanating from transport-related processes. We demonstrate how the application of simple statistical regression methods in Wasserstein space, via a novel mathematical transformation approach, can significantly outperform modern deep learning methods and may provide a path towards accurate, verifiable, robust, and understandable predictive models for a variety of signal and image modeling and classification applications.

Gustavo Rhode
University of Virginia
gustavo@virginia.edu

MS42

Non-Euclidean Generative Modeling

I will describe how to construct a diffusion generative model that allows users to, given training data that follow any latent statistical distribution in some manifold, generate more data exactly in the same space that follow the same distribution. This is based on a sequence of methodological developments, starting from manifold optimization, to sampling, and finally to generative modeling.

Molei Tao, Molei Tao
School of Mathematics, Georgia Institute of Technology
mtao@gatech.edu, mtao@gatech.edu

MS43

Operator Learning for the Petrov Galerkin Variational Form

In this talk, we propose an operator learning framework whose structure is motivated by the Petrov-Galerkin formulation of the underlying partial differential equation (PDE). This operator network, termed as VarMiON-PG, consists of a basis sub-network (which can be predetermined or learned) for the trial space, and a second basis sub-network for the weighting function space. The latter is trained to be "optimal with respect to the choice of the underlying norm in which the error is measured. We demonstrate the performance of VarMiON-PG on advection-dominated flows while handling multiple input functions. The learned basis functions are visualized and compared to the (known) optimal basis functions associated with a given trial function space for the PDE model.

Philip Charles
Maryland University
charlesp@umd.edu

Deep Ray
University of Maryland, College Park
deepray@umd.edu

Assad A. Oberai
University of Southern California
aoberai@usc.edu

Thomas Hughes
UT Austin
hughes@oden.utexas.edu

Michael Abdelmalik
Eindhoven University of Technology
m.abdel.malik@tue.nl

MS43

DeepMartNet - A Martingale Based Deep Neural Network Algorithm for Eigenvalue/bvp Problems of Pdes and Optimal Stochastic Controls

In this talk, we will present a deep neural network (DNN) learning algorithm for solving high dimensional Eigenvalue (EV) and boundary value problems (BVPs) for elliptic operators and initial BVPs (IBVPs) of quasi-linear parabolic equations as well as optimal stochastic controls. The method is based on the Martingale property in the stochastic representation for the eigenvalue/BVP/IBVP problems and martingale principle for optimal stochastic controls. A loss function based on the Martingale property can be used for an efficient optimization by sampling the stochastic processes associated with the elliptic operators or value process for stochastic controls. The Martingale property conforms

naturally with the stochastic gradient descent process for the DNN optimization. The proposed algorithm can be used for eigenvalue problems and BVPs and IBVPs with Dirichlet, Neumann, and Robin boundaries in bounded or unbounded domains and some feedback stochastic control problems. Numerical results for BVP and EV problems in high dimensions will be presented.

Qing He
-
-
andrewho@smu.edu

Daniel Margolis
Southern Methodist University
dmargolis@smu.edu

Wei Cai
Department of Mathematics
Southern Methodist University
cai@smu.edu

MS43

On the Training and Generalization of Deep Operator Networks

We present a novel training method for deep operator networks (DeepONets), one of the most popular neural network models for operators. DeepONets are constructed by two sub-networks, namely the branch and trunk networks. Typically, the two sub-networks are trained simultaneously, which amounts to solving a complex optimization problem in a high dimensional space. In addition, the nonconvex and nonlinear nature makes training very challenging. To tackle such a challenge, we propose a two-step training method that trains the trunk network first and then sequentially trains the branch network. The core mechanism is motivated by the divide-and-conquer paradigm and is the decomposition of the entire complex training task into two subtasks with reduced complexity. Therein the Gram-Schmidt orthonormalization process is introduced which significantly improves stability and generalization ability. On the theoretical side, we establish a generalization error estimate in terms of the number of training data, the width of DeepONets, and the number of input and output sensors. Numerical examples are presented to demonstrate the effectiveness of the two-step training method, including Darcy flow in heterogeneous porous media.

Sanghyun Lee
Florida State University
slee17@fsu.edu

Yeonjong Shin
North Carolina State University
yeonjong_shin@ncsu.edu

MS43

Distributed Learning and Foundation Models for Multi-Operator Learning: Extrapolation and Generalization

Single operator learning learns a single operator which maps a function to another function. In this talk we will discuss multi-operator learning (MOL) which is an extension of the single operator learning: it identify the operator and construct the operator simultaneously. We will discuss several advantages of the MOL such as the extrapolations and generalization. Particularly, we will study the multi-modality foundation model PROSE for science.

Zecheng Zhang
Florida State University
zz23e@fsu.edu

MS44

Predicting ALM, BD, and BF Using Supervised and P-Laplacian Based Semi-Supervised Learning

Nutritionists require accurate data on critical body characteristics such as Appendicular Lean Mass (ALM), Bone Density (BD), and Body Fat Percentage (BF) to conduct their research. Current methods to predict ALM, BD, and BF such as DEXA (Dual X-ray Absorptiom- etry) scans are not cost-effective. Previous studies have shown biomarkers such as waist circumference, arm length, and head circumference have correlations with ALM, BD, and BF. We applied semi-supervised learning algorithms to predict these three quantities using data on 40+ biomarkers of 700 patients provided by Pennington Biomedical Research Center or PBRC. Semi-supervised learning algorithms have been extensively used for classification and regression, especially when limited training data is available. An effective technique for exploiting unlabeled data in semi-supervised learning is to utilize a graph structure, which may be intrinsic to the data, or constructed based on similarities between data points. p- Laplacian is one of many graph-based semi-supervised learning techniques. This technique has only been used for classification problems, but a recent paper has shown it can also be used for regression. In this talk, we will elucidate the mathematical foundations behind p-Laplacian and demonstrate how we used it to predict ALM, BD, and BF by exploiting the similarity between the patients. We will also compare the results with those of supervised learning algorithms.

Alyssa Blount, Nadejda Drenska, Gyaneshwar Agrahari, Berend Grandt, Tristan Desoto, Atif Iqbal, Kathryn Simon, Jacob Kapita
Louisiana State University
ablou14@lsu.edu, ndrenska@lsu.edu, gyan.agrahari77@gmail.com, bgrand7@lsu.edu, tdesot3@lsu.edu, aiqbal6@lsu.edu, ksimo53@lsu.edu, jkapit1@lsu.edu

MS44

Robotics-Assisted Stroke Rehabilitation with Machine Learning-Based Residual Severity Classification

Stroke therapy is essential to reduce impairments and improve motor movements by engaging autogenous neuroplasticity. This study uses supervised learning methods to address a clinician's autonomous classification of stroke residual severity labeled data towards improving in-home robotics-assisted stroke rehabilitation. Thirty-three stroke patients participate in in-home therapy sessions using the Motus Nova robotics rehabilitation technology to capture upper and lower body motion. The therapy session summary data is based on high-resolution movement and assistance data and clinician-informed discrete stroke residual severity labels. We demonstrate that the light gradient boosting method provides the most reliable autonomous detection of stroke severity. This method achieved an average of 94% accuracy, measured using the F1-score performance measure with 10-fold cross-validation. We show how objectively measured rehabilitation training paired with

machine learning methods can be used to identify the residual stroke severity class with efforts to enhance in-home self-guided, individualized stroke rehabilitation. As data from rehabilitation practices are often of comparable size and nature to the data collected in our study, this suggests that the light gradient boosting method should be considered a standard, more efficient tool for this analysis.

Russell Jeter
Georgia State University
russell.jeter@outlook.com

MS45

Patch Formation Driven by Stochastic Effects of Interaction Between Viruses and Virus-like Particles

Defective interfering particles (DIPs) are virus-like particles that occur naturally during virus infections. These particles are defective, lacking essential genetic materials for replication, but they can interact with the wild-type virus and potentially be used as therapeutic agents. However, the effect of DIPs on infection spread is still unclear due to complicated stochastic effects and nonlinear spatial dynamics. In this work, we develop a model with a new hybrid method to study the spatial-temporal dynamics of viruses and DIPs co-infections within hosts. We present two different scenarios of virus production and compare the results from deterministic and stochastic models to demonstrate how the stochastic effect is involved in the spatial dynamics of virus transmission. We compare the spread features of the virus in simulations and experiments, including the formation and the speed of virus spread and the emergence of stochastic patchy patterns of virus distribution. Our simulations simultaneously capture observed spatial spread features in the experimental data, including the spread rate of the virus and its patchiness. The results demonstrate that DIPs can slow down the growth of virus particles and make the spread of the virus more patchy.

Johnny Yang
Indiana University
jy53@iu.edu

Wai-Tong (Louis) Fan
Indiana University
Indiana University
waifan@iu.edu

MS45

Mobility in Classical Epidemiological Models: Final Pandemic Size and Data-Driven Inference

It is well known that many compartmental models tend to overestimate the size of a pandemic. In this talk, I will try to partially address this issue by introducing a data-informed mobility distribution into the standard compartment model. At the beginning of a pandemic, individuals with higher mobility are more likely to be infected, which leads to an overestimate of the eventual pandemic size. Then I will demonstrate how to use deep learning tools to infer the mobility distribution in a community. Further connections with opinion dynamics will also be discussed.

Yao Li
University of Massachusetts Amherst
Department of Mathematics and Statistics
liyao@umass.edu

Ning Jiang
Department of Mathematics and Statistics
University of Massachusetts Amherst
ningjiang@umass.edu

Weiqi Chu, Weiqi Chu
University of Massachusetts Amherst
weiqichu@umass.edu, weiqichu@umass.edu

MS45

Computational Modeling, Analysis and Simulation for Data-Driven Challenges in Mathematical Biology with Applications to Epidemiology

The transmission dynamics of infectious diseases such as COVID-19 has made us to re-envision how we model, analyze and simulate the spread of infectious diseases and evaluate the effectiveness of non-pharmaceutical control measures as important mechanisms for assessing the potential for sustained transmission. Incorporating human behavior into these models responding to a perceived increase of the infections in the local environment in real-time, adds another layer of complexity in these models. Also, there have also been rapid developments in employing a Physics Informed Neural Networks (PINNs) approach to estimate the model parameters such as the transmission, infection, quarantine and recovery rate using real data sets. In this work, we present modeling, analysis and simulation through Disease Informed Neural Networks (DINNs) and its application to real data modeled using non-linear differential equations. We discuss how these approaches are capable of predicting the behavior of a disease described by modified compartmental models that include parameters and variables associated with the governing differential equations describing the dynamics of the disease. Through benchmark examples, we will show how DINNs can predict optimal parameters for given datasets for a variety of applications.

Padmanabhan Seshaiyer
Mathematical Sciences, George Mason University
pseshaiy@gmu.edu

Alonso G. Ogueda
George Mason University, U.S.
aogueda@gmu.edu

MS45

Modeling the Menstrual Cycle and Dosing Regimens of Oral Hormonal Contraceptives

Menstruation is driven by hormones. Hormonal imbalances can lead to issues with menstrual cycle that affect health, wellness, and quality of life. Oral contraceptives are a leading form of birth control in the United States. The primary goal of hormonal contraceptives is to alter the menstrual cycle to prevent ovulation. Existing mathematical models are able to capture the dynamics of key hormones for normally cycling adult women with no external interference, but do not accurately capture the effects of hormonal contraceptives. Here we explore the stability of the contraceptive state achieved by oral hormonal contraceptives using a mechanistic mathematical model of the menstrual cycle.

Lihong Zhao
Kennesaw State University
U.S

lzhao12@kennesaw.edu

MS46

Machine UnLearning via Algorithmic Stability

We study the problem of machine unlearning and identify a notion of algorithmic stability, Total Variation (TV) stability, which we argue is suitable for the goal of exact unlearning. For convex risk minimization problems, we design TV-stable algorithms based on noisy Stochastic Gradient Descent (SGD) and corresponding efficient unlearning algorithms based on constructing a near-maximal coupling of Markov chains for the noisy SGD procedure. To understand the trade-offs between accuracy and unlearning efficiency, we give upper and lower bounds on excess empirical and population risk of TV stable algorithms for convex risk minimization. Our techniques generalize to arbitrary non-convex functions, and our algorithms are also differentially private.

Raman Arora
Johns Hopkins University
rarora8@jhu.edu

MS46

Privacy in Representation Learning: Measurement and Mitigation

Large representation learning models have shown enormous promise in transforming computer vision. However, a major barrier to their widespread adoption is that many of the larger models are trained on very large datasets curated from the internet which might inadvertently contain sensitive information; this can lead to privacy concerns if the model memorizes this information from its training data. In this talk, I will talk about some of our recent research in measuring memorization in large representation learning models as well as building models with rigorous privacy guarantees. For measurement, we propose a new metric called "deja vu memorization" for image and vision-language representation learning models, and show that this kind of memorization can be mitigated with suitable choices of hyperparameters. For representation learning with rigorous privacy guarantees, we develop ViP and DP-Cap – the first foundation model for computer vision with differential privacy, one of the gold standards in private data analysis in recent years, and evaluate its privacy-utility tradeoff.

Kamalika Chaudhuri
Department of CSE, UCSD
kamalika@cs.ucsd.edu

MS46

Measuring Data Leakage in Machine-Learning Models with Fisher Information

Machine-learning models contain information about the data they were trained on. This information leaks either through the model itself or through predictions made by the model. Consequently, when the training data contains sensitive attributes, assessing the amount of information leakage is paramount. We propose a method to quantify this leakage using the Fisher information of the model about the data. Unlike the worst-case a priori guarantees of differential privacy, Fisher information loss measures leakage with respect to specific examples, attributes, or subpopulations within the dataset. We motivate Fisher information loss through the Cramr-Rao bound and delineate the implied threat model. We provide efficient methods to compute Fisher information loss for output-perturbed generalized linear models and neural networks. Finally, we empirically validate Fisher information loss as a useful measure of information leakage.

Chuan Guo
FAIR, Meta
chuanguo@meta.com

MS46

Attribute Privacy: Framework and Mechanisms

Ensuring the privacy of training data is a growing concern since many machine learning models are trained on confidential and potentially sensitive data. Much attention has been devoted to methods for protecting individual privacy during analyses of large datasets. However in many settings, global properties of the dataset may also be sensitive (e.g., mortality rate in a hospital rather than presence of a particular patient in the dataset). In this work, we depart from individual privacy to initiate the study of attribute privacy, where a data owner is concerned about revealing sensitive properties of a whole dataset during analysis. We propose definitions to capture attribute privacy in two relevant cases where global attributes may need to be protected: (1) properties of a specific dataset and (2) parameters of the underlying distribution from which dataset is sampled. We also provide two efficient mechanisms for specific data distributions and one general but inefficient mechanism that satisfy attribute privacy for these settings. We base our results on a novel and non-trivial use of the Pufferfish framework to account for correlations across attributes in the data, thus addressing the challenging problem of developing Pufferfish instantiations and algorithms for general aggregate secrets that was left open by Kifer and Machanavajjhala in 2014.

Wanrong Zhang
School Of Engineering And Applied Sciences
Harvard
imwanrongz@gmail.com

MS47

A Network Science Approach to Understanding EV Charging Accessibility in the Texas Triangle Region

A robust electric vehicle (EV) charging infrastructure is essential for the widespread adoption of electric vehicles and the transition towards sustainable energy. Understanding the spatial distribution of existing infrastructures is crucial for the strategic expansion and futureproofing of the burgeoning EV market. This research presents a detailed analysis of EV charging stations within the Texas Triangle, which includes the state's four major metropolitan areasAustin, Dallas, Houston, and San Antonio. The comprehensive dataset utilized for constructing and analyzing the network graph was sourced from the Alternative Fuels Data Center, which provides extensive information on EV charging stations. We employed a graph-based method to model the distribution of charging stations, effectively capturing the complex spatial dynamics in an intuitive and computationally efficient manner. Additionally, this study examines how specific attributes of EV stations, such as operational hours, pricing, and accessibility, influence the network's capacity to meet regional charging demands. Graph neural networks (GNNs) were utilized to accurately model

the spatial distribution of charging stations, optimized using the Adam optimizer. We also explored various complementarity metrics to assess the spatial distribution of the charging station network. Finally, we conducted a temporal analysis examine how the network evolved over time.

Gyaneshwar Agrahari
Louisiana State University
gagrah1@lsu.edu

Yogesh Sapkota
Youngstown State University
yspkta16@gmail.com

MS47
Topological Neural Networks

Deep neural networks have revolutionized machine learning by leveraging vast amounts of data across various domains. While they excel in processing standard data like images and text, they face challenges when applied to structured scientific data in non-Euclidean domains. Geometric Deep Learning (GDL) extends the capabilities of deep learning to non-Euclidean domains by incorporating geometric principles. However, capturing non-local properties inherent in topological data remains a challenge. In this presentation, I will introduce higher-order structures beyond traditional graph-based approaches, known as Topological Neural Networks (TNNs). TNNs offer a deeper understanding of complex data relationships. Moreover, I will discuss potential applications of TNNs, including drug discovery and social media analysis, and propose a novel approach (TopoX) inspired by classical topological concepts and recent advancements in higher-order networks. This framework aims to provide a more sophisticated representation of data in topological domains, pushing TNNs towards new frontiers. Relevant citations include seminal works in machine learning, geometric deep learning, and topological data analysis.

Ibrahem Aljabea
Louisiana State University (Baton Rouge, LA, US)
Louisiana State University (Baton Rouge, LA, US)
ialjab2@lsu.edu

Mustafa Hajij
Santa Clara University, U.S.
mhajij@SCU.edu

MS47
Decoding Social Networks: Llm Agents and Information Flow Dynamics

Graph based structured combined with Large Language Models (LLMs), which have demonstrated their capability to simulate human conversations, provide a powerful framework for simulating and analyzing dynamic social behaviors in networked systems. We develop such a framework that simulates conversations between LLM powered Generative Agents and measure the information flow in the network using CEM (Conversation Evaluation Model). Each agent acts as a node in our network, characterized by distinct personas. These agents engage in simulated conversation based on the graph structure providing insights into how information propagates through various network topologies. Furthermore, it also reveals the inherent biases and the capabilities of LLMs in simulating realistic conversations. We analyze how different network structures and agent characteristics influence the spread and transformation of information. GNNs enhance our framework by enabling predictions of network behavior and the potential impact of various interventions. Our study opens new avenues for both theoretical research and practical applications in network science and social behavior analysis.

Bishal Lamichhane
University of Nevada, Reno
blamichhane314@gmail.com

MS48
Distributional Smoothing for Optimization

Distribution smoothing is a method for modifying a target function in an optimization problem using a set of surrogate functions that are easier to optimize. When applied to (stochastic) gradient descent, we substitute the gradient of the target function with the gradient from this more favorable class of functions. This presentation will begin with an exploration of the intuition and motivation behind distributional smoothing. Subsequently, we will delve into the convergence properties of Smooth Gradient Descent, labeled SmoothGD, and its stochastic counterpart, SmoothSGD, demonstrating that their convergence rates are no worse than those of their non-smoothed counterparts. Finally, we will provide numerical evidence from both machine learning and data science problems.

Andrew Starnes
Lirio AI Research
astarnes@lirio.com

MS48
High-Dimensional Optimization with a Novel Nonlocal Gradient

We develop a nonlocal gradient that skips small local minima and captures major structures of the loss's landscape in optimization of high-dimensional, multimodal functions. The key idea is applying Gaussian smoothing directionally and conduct 1D long-range exploration with large smoothing radius along each orthogonal direction. Gauss-Hermite quadrature rule is employed to obtain an accurate estimator of the nonlocal gradient. We provide convergence theory and demonstrate the performance of our method in several high-dimensional tests and applications.

Hoang A. Tran
Oak Ridge National Laboratory
Computer Science and Mathematics Division
tranha@ornl.gov

MS48
Implicit Methods for Deep Learning on Graphs

In this talk, I will present two implicitly defined graph neural networks (GNNs) based on finding the fixed point of equilibrium equation and parameterizing diffusion PDE on graphs. These new implicit GNNs can effectively overcome the over-smoothing issue of GNNs and learn with minimal supervision. Also, these new models enjoy computational and memory efficiency when applying operator splitting schemes and the adjoint method for their training and inference.

Bao Wang
University of Utah

bwang@math.utah.edu

MS49

Efficient Low-Dimensional Compression for Deep Overparameterized Learning and Fine-Tuning

While overparameterization in machine learning models offers great benefits in terms of optimization and generalization, it also leads to increased computational requirements as model sizes grow. In this work, we show that leveraging inherent low-dimensional structure within the model parameter updates, we can reap the benefits of overparameterization without the computational burden. In practice, we demonstrate the effectiveness of this approach for deep low-rank matrix completion as well as fine-tuning language models. For theory of deep overparameterized low-rank matrix recovery, we show that the learning dynamics of each weight matrix are confined to an invariant low-dimensional subspace. Consequently, we can construct and train compact, highly compressed factorizations possessing the same benefits as their overparameterized counterparts. For language model fine-tuning, we introduce a method called Deep LoRA, which improves the existing low-rank adaptation (LoRA) technique, leading to reduced overfitting and a simplified hyperparameter setup, all while maintaining comparable efficiency. The effectiveness of Deep LoRA is validated through its performance on natural language understanding tasks, particularly when fine-tuning with a limited number of samples.

Laura Balzano
Electrical Engineering and Computer Science
University of Michigan
girasole@umich.edu

MS49

Zeroth-Order Regularized Optimization

We consider the problem of minimizing a high-dimensional objective function, which may include a regularization term, using only (possibly noisy) evaluations of the function. Such optimization is also called derivative-free, zeroth-order, or black-box optimization. We propose a new zeroth-order regularized optimization method, dubbed ZORO. When the underlying gradient is approximately sparse at an iterate, ZORO needs very few objective function evaluations to obtain a new iterate that decreases the objective function. We achieve this with an adaptive, randomized gradient estimator, followed by an inexact proximal-gradient scheme. Under a novel approximately sparse gradient assumption and various different convex settings, we show that the (theoretical and empirical) convergence rate of ZORO is only logarithmically dependent on the problem dimension. Numerical experiments show that ZORO outperforms existing methods with similar assumptions, on both synthetic and real datasets.

HanQin Cai
University of Central Florida
hqcai@ucf.edu

Daniel McKenzie
Colorado School of Mines
dmckenzie@mines.edu

Wotao Yin
Alibaba Group US/DAMO Academy, U.S.
wotao.yin@gmail.com

Zhenliang Zhang
Xmotors AI
zhenliang.zhang@gmail.com

MS49

Least Absolute Deviation Approach to Robust Max-Affine Regression

Max-affine model has been utilized to model multivariate polynomials and a class of deep neural network. We consider robust max-affine regression via least absolute deviation resilient to adversarial corruption on a subset of training data. An iterative algorithm has been proposed to solve the resulting non-convex optimization. We present a local convergence result so that the algorithm when suitably initialized converges linearly to the ground truth parameters. The sample complexity and tolerable ratio of outliers are quantified in the non-asymptotic analysis. We also provide a set of comprehensive numerical experiments that corroborate our theoretical results.

Seonho Kim, Haitham Kanj, Kiryung Lee
The Ohio State University
kim.7604@osu.edu, kanj.7@osu.edu, kiryung@ece.osu.edu

MS49

Exponentially Convergent Algorithms for Supervised Matrix Factorization

Supervised matrix factorization (SMF) is a classical machine learning method that simultaneously seeks feature extraction and classification tasks, which are not necessarily a priori aligned objectives. Our goal is to use SMF to learn low-rank latent factors that offer interpretable, data-reconstructive, and class-discriminative features, addressing challenges posed by high-dimensional data. Training SMF model involves solving a nonconvex and possibly constrained optimization with at least three blocks of parameters. Known algorithms are either heuristic or provide weak convergence guarantees for special cases. In this paper, we provide a novel framework that 'lifts' SMF as a low-rank matrix estimation problem in a combined factor space and propose an efficient algorithm that provably converges exponentially fast to a global minimizer of the objective with arbitrary initialization under mild assumptions. Our framework applies to a wide range of SMF-type problems for multi-class classification with auxiliary features. To showcase an application, we demonstrate that our algorithm successfully identified well-known cancer-associated gene groups for various cancers.

Hanbaek Lyu
University of Wisconsin Madison
hlyu@math.wisc.edu

MS50

Partial Alignment and Enhancement of 3D Maps for Automated Atomic Building in Cryo-Em

Advances in theory and computational practice have brought Optimal Transport (OT) to the stage of delivering immediate applications in various situations, where OT provides a natural framework for tackling various problems (e.g. registration, domain adaptation) with large and geometrically structured data, e.g. in imaging sciences, biology or medicine. In this context, my group recently developed computational methods that take advantage of OT-based metrics, with a focus on studying 3D maps of

biomolecules imaged from single particle cryogenic electron microscopy (cryo-EM), and more recently Electronic Health Records data. In this talk, I will more specifically focus on cryo-EM, and present our procedure for partial alignment of 3D maps that first finds a coupling between 3D point-cloud representations associated with their so-called unbalanced Gromov Wasserstein divergence, and second, uses this coupling to find an optimal rigid body transformation. I will further present how this framework can be applied to atomic model building from density maps, coupled with a deep learning based method for enhancing maps to automate this process at low/intermediate resolution, suggesting potential applications of Partial Optimal Transport for improving Cryo-EM pipelines.

Khanh Dao Duc
Department of Mathematics
University of British Columbia
kdd@math.ubc.ca

MS50

Stereographic Spherical Sliced Wasserstein Distances

From brain image analysis to geoscience and astronomy, many applications involve spherical distributions, making their comparison highly relevant in these fields. The growing use of optimal transport (OT) theory, noted for its favorable statistical, geometrical, and topological properties, has increased interest in employing OT for comparing spherical distributions. However, spherical OT calculations remain computationally expensive. This talk introduces the Stereographic Spherical Sliced Wasserstein (S3W) distance, a new metric for comparing spherical probability measures using stereographic projection and the generalized Radon transform. This metric offers a computationally efficient and highly parallelizable method for comparing spherical probability measures. We will demonstrate several applications of the proposed metric in various machine learning contexts.

Soheil Kolouri
Vanderbilt University
soheil.kolouri@vanderbilt.edu

MS50

Alignment of Density Maps in Wasserstein Distance

In this talk we propose an algorithm for aligning three-dimensional objects when represented as density maps, motivated by applications in cryogenic electron microscopy. The algorithm is based on minimizing the 1-Wasserstein distance between the density maps after a rigid transformation. The induced loss function enjoys a more benign landscape than its Euclidean counterpart and Bayesian optimization is employed for computation. Numerical experiments show improved accuracy and efficiency over existing algorithms on the alignment of real protein molecules. In the context of aligning heterogeneous pairs, we illustrate a potential need for new distance functions.

Amit Singer, Ruiyi Yang
Princeton University
amits@math.princeton.edu, ry8311@princeton.edu

MS51

Multiscale Topology-enabled AI for Drug Discovery

Despite the success of natural language processing (NLP) models across diverse domains, their application in computational biology has been limited due to the challenge of incorporating critical 3D structural information from biological sequences, which is incompatible with traditional NLP architectures. We introduce TopoFormer, a novel model that integrates NLP with a multiscale topology technique, the Persistent Topological Hyperdigraph Laplacian (PTHL). PTHL systematically transforms complex 3D protein-ligand interactions into NLP-admissible sequences of topological invariants and homotopic shapes, enabling the capture of essential spatial interactions. TopoFormer outperforms traditional algorithms and recent deep learning approaches, achieving superior accuracy in scoring, ranking, docking, and screening tasks across multiple benchmark datasets. Beyond computational biology, this approach offers a generalizable framework for converting high-dimensional structured data into NLP-compatible formats, with potential applications in drug discovery, including combating COVID-19 and addressing drug-resistant mutations.

Dong Chen
Michigan State University
chendo11@msu.edu

MS51

Geometric and Topological Data Analysis in Biology Applications

This talk will discuss geometric and topological data analysis in biology applications and focus on the recent developments. The multiscale analysis of graph neural network and the de Rham-Hodge theory provides a unified paradigm for the evolving manifolds constructed from filtration, which induces a family of evolutionary complexes. While the present evolutionary de Rham-Hodge method can be easily applied to close manifolds, the emphasis is given to more challenging compact manifolds with 2-manifold boundaries, which require appropriate analysis and treatment of boundary conditions on differential forms to maintain proper topological properties. Meanwhile, we will discuss the the multiscale graph neural network in the modeling of biomolecules.

Jiahui Chen
Department of Mathematical Sciences
University of Arkansas
jiahuic@uark.edu

MS51

Transformer-Assisted Spectral Graph Algorithms for Predicting Scarcely Labeled and Imbalanced Molecular Data

Research in biological and molecular sciences frequently involves the investigation of a variety of molecular properties, particularly in applications such as drug discovery. Experiments to determine these properties are time-consuming and expensive and may be subject to ethical concerns. As a result, it is often difficult to obtain large molecular data sets, which many machine and deep learning models require to make accurate predictions about such properties. To address the challenge of making predictions on scarcely labeled molecular data, the authors constructed three graph-based models incorporating Merriman-Bence-Osher (MBO) techniques. In particular, this talk will focus

on one of these models, which consists of a home-made bidirectional transformer integrated with a graph-based modification of the MBO scheme. All models were validated on five benchmark molecular data sets and thoroughly compared to other competing methods, such as support vector machines, random forests, and gradient boosting decision trees, which are known for their good performance on small data sets. The various methods were additionally analyzed using residue-similarity (R-S) scores and R-S indices. Extensive computational experiments and theoretical analysis demonstrated that the authors' proposed models, particularly the transformer-assisted model, performed very well with scarcely labeled data, even when as little as 1% of the data set was labeled.

Nicole Hayes
Michigan State University
Department of Mathematics
hayesni4@msu.edu

Ekaterina Merkurjev
Department of Mathematics and CMSE
Michigan State University
merkurje@msu.edu

Guowei Wei
Department of Mathematics
Michigan State University
weig@msu.edu

MS51

Math Meets AI: Leveraging Computational Topology and AI Models to Advance Covid-19 Research

The ongoing COVID-19 pandemic has underscored the critical need for computational tools to interpret complex biological data and enhance our understanding of viral spread and treatment efficacy. This study presents a pioneering integration of computational topology and artificial intelligence (AI), showcasing a novel approach to epidemiological research. Specifically, we studied the capacities of persistent spectral graphs (PSGs) for analyzing intricate topological and geometric properties of high-dimensional biological data. Our approach begins with the formulation of persistent Laplacian matrices (PLMs), constructed through applying a dynamic filtration parameter. The harmonic spectra derived from the null spaces of these PLMs encapsulate the underlying topological features, while the non-harmonic spectra elucidate the geometry of high-dimensional datasets. During the COVID-19 pandemic, we integrated PSG, genomics, and deep learning into a Math-AI model to predict the binding free energy (BFE) changes caused by mutations in the interaction between the virus's Spike protein and the human ACE2 receptor or antibodies. Such a Math-AI model has successfully forecasted the predominance of Omicron variants one or two months ahead of their presence, offering a glimpse into a future where mathematical techniques and biology converge to combat viral threats more effectively.

Rui Wang
Simons Center for Computational Physical Chemistry
New York University
rw3594@nyu.edu

MS52

Solver-in-the-Loop with Zeroth Order Training

Using machine learning to accelerate and/or improve simulations can yield scientific discoveries that are otherwise impossible due to the prohibitively expensive nature of these simulations. Unfortunately, ML models have yielded limited success in complex scientific domains due to large data requirements. A potential solution is to exploit the abundance of available scientific knowledge. The biggest roadblock in integrating scientific knowledge with ML is to have it in a differentiable form that can be ingested in the current first-order training paradigm. In many scientific simulation applications, simulation codes either do not provide gradient information or are non-differentiable. To overcome the challenge of effortless integration of simulation codes with ML, we developed a learning framework, called DeepZero, capable of achieving high accuracy while using gradient-free training techniques. We demonstrate the effectiveness of DeepZero by performing gradient-free training of a solver-in-the-loop pipeline with the popular open-source simulation code MFEM. LLNL-ABS-863436

William Anderson, Brian Bartoldson, Youngsoo Choi, James Diffenderfer, Bhavya Kailkhura, Konstantinos Parasyris
Lawrence Livermore National Laboratory
anderson316@llnl.gov, bartoldson1@llnl.gov, choi15@llnl.gov, diffenderfer2@llnl.gov, kailkhura1@llnl.gov, parasyris1@llnl.gov

MS52

Learn2Solve: Real-Time Uncertainty Quantification for Solutions of Dynamical Systems

In this talk, we present a brief overview of the current state of the art deep learning methods for learning solutions of dynamical systems governed by PDEs. We then present a Learn2Solve approach that aims to learn well-developed numerical methods for the solutions of dynamical systems. Our approach not only learns the underlying physics governing the dynamical system, but also learns its numerical solution methods. As result, once trained our deep learning solution can generalize well for unseen scenarios such as new initial conditions, new boundary conditions, new meshes, new geometries, etc. We shall demonstrate our approach on various equations including supersonic flows with uncertainty quantification capabilities.

Tan Bui-Thanh
The University of Texas at Austin
tanbui@oden.utexas.edu

MS52

Sparse Model Selection for Oscillatory Data-sets with Hidden States

Building models for biological, chemical, and physical systems has traditionally relied on domain-specific intuition about which interactions and features strongly influence a system. Recently developed for and applied to dynamical systems, sparse optimization strategies can scan and select a subset of terms from a library that best describes data, automatically interfering potential model structures from a broad but well-defined class. I will discuss the application of Data Assimilation for Hidden Sparse Inference (DAHSI) to discover models for metabolic and temperature regulation in hibernating mammals. The application of DAHSI to realistic data sources has inspired potential improvements to the algorithm to handle low-information content data and clarify the identifiability of the discovered

models.

Niall M. Mangan
Engineering Sciences and Applied Mathematics
Northwestern University
niall.mangan@northwestern.edu

Andrew Engedal
University of Wisconsin, Madison
aengedal@wisc.edu

Jessica Lin
Berkeley
jessica.lin@berkeley.edu

Manu Jayadharan, Cody FitzGerald
Northwestern University
manu.jayadharan@northwestern.edu,
cody.fitzgerald@northwestern.edu

MS52

Interpretable Regularization-Based Nonparametric Uncertainty Quantification and Model Reduction for Dynamical Systems

Large-scale nonlinear dynamical systems often involve several interacting components, and parameters must be estimated from experimental data. While identifying key components in a complex dynamical system enables model reduction to accelerate simulations, care must be taken when accounting for the uncertainty in the parameters used to find such reduced models. In this talk, we present a data-driven method for model reduction using l1-regularization that involves minimal parameterization, that is interpretable, and that generates a family of models which trade-off model complexity for estimation error. Then, we introduce Tikhonov regularization under moment constraints as form of a maximum a posteriori (MAP) estimate for a probability density on unknown parameters induced by a suitable Bayesian nonparametric method. We show that the MAP can be computed by solving a finite-dimensional problem which, in several cases of interest, is convex, unconstrained, and has a smooth objective function. Combining both, we are able to both find interpretable reduced models, and quantify the uncertainty in their parameters. We present experimental results showing that, on one hand, our model reduction method selects reduced models with good extrapolation properties, which is an important consideration in practical applications; on the other, that using Tikhonov regularization under moment constraints may be an efficient approach to quantify the uncertainty on model parameters.

Carlos Sing-Long
Pontificia Universidad Catolica de Chile
casinglo@uc.cl

Qian Yang
University of Connecticut
qyang@uconn.edu

Evan Reed
Stanford University
evanreed@stanford.edu

MS53

Generative Modeling of Time-Dependent Densities via Optimal Transport and Projection Pursuit

Motivated by the computational difficulties incurred by popular deep learning algorithms for the generative modeling of temporal densities, we propose a cheap alternative that requires minimal hyperparameter tuning and scales favorably to high-dimensional problems. In particular, we use a projection-based optimal transport solver to join successive samples and, subsequently, use transport splines to interpolate the evolving density. When the sampling frequency is sufficiently high, the optimal maps are close to the identity and are, thus, computationally efficient to compute. Moreover, the training process is highly parallelizable as all optimal maps are independent and can, thus, be learned simultaneously. Finally, the approach is based solely on numerical linear algebra rather than minimizing a nonconvex objective function, allowing us to easily analyze and control the algorithm. We present several numerical experiments on both synthetic and real-world datasets to demonstrate the efficiency of our method. In particular, these experiments show that the proposed approach is highly competitive compared with state-of-the-art normalizing flows conditioned on time across a wide range of dimensionalities.

Jonah Botvinick-Greenhouse, Yunan Yang
Cornell University
jrb482@cornell.edu, yunan.yang@cornell.edu

Romit Maulik
Pennsylvania State University
rmaulik@psu.edu

MS53

Real-Time Bayesian Inversion of Autonomous Systems with Application to Tsunami Forecasting

Hessian-based algorithms for the solution to Bayesian inverse problems typically require many actions of the Hessian matrix on a vector. For problems with high-dimensional parameter fields or expensive-to-evaluate forward operators, a direct approach is often computationally intractable, especially in the context of real-time inversion. One way to overcome the computational bottleneck of Hessian matrix-vector multiplications in these large-scale inverse problems is to exploit the structure of the underlying operators. A particular class of operators for which we can exploit the structure very effectively are those representing autonomous systems. The evolution of such systems with respect to any given input may depend on the system's current state but does not explicitly depend on the independent variable (e.g., time). We present a scalable and computationally efficient approach for Bayesian inversion of problems involving autonomous systems. Our approach splits the computation into a precomputation ("offline") phase and a real-time inversion ("online") phase. Contrary to other methods, this approach does not employ a lower-fidelity approximation but instead uses the full discretization obtained from the PDE-based model. The method is applied to a real-time tsunami Bayesian inverse problem involving a time-invariant dynamical system. Scalability and efficiency of the implementation are demonstrated for state-of-the-art GPU-accelerated compute architectures.

Stefan Henneking
Oden Institute, UT Austin
stefan@oden.utexas.edu

Sreeram R. Venkat
University of Texas Austin

201 E 24th St, Austin, TX 78712
srvenkat@utexas.edu

Milinda Fernando
Oden Institute, UT Austin
milinda@oden.utexas.edu

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

MS53

Experimental Design for Uncertainty-aware Digital Twins

Digital twins offer a remarkable platform for harnessing two of the most transformative advancements in computer science: generative models and high-fidelity physics simulations. However, selection of field observations for updating the digital twin underscores the importance of experimental design. We introduce a novel approach by reinterpreting likelihood-based generative models and formulating a joint optimization of variational and design parameters, efficiently generating the posterior samples necessary for data assimilation and identifying optimal designs. We demonstrate these concepts through a synthetic experiment on a 50-year CO2 sequestration project. Project managers face choices about acquiring data either directly, using monitoring wells, or indirectly, through seismic observations that rely on acoustic wave PDEs. Each data acquisition entails financial costs, compelling the need to maximize the information derived from each. Our setup treats design as a sequential decision-making process, where knowledge accumulated from previous time steps informs subsequent choices, adaptively selecting the optimal design. Additionally, the challenge posed by the multimodality of dataranging from centimeter-resolution well data to ten-meter resolution seismic datais met by the principled use of learned and physics-based summary statistics. These techniques combine to address these challenges and provide an all in one solution for digital twins with experimental design.

Felix Herrmann, Rafael Orozco
Georgia Institute of Technology
felix.herrmann@gatech.edu, rorozco@gatech.edu

Abhinav Gahlot
Georgia Tech
agahlot8@gatech.edu

MS53

Ensemble Kalman Inversion for Linear Bayesian Inverse Problems

We consider the Bayesian smoothing problem of inferring the initial state of a linear dynamical system, given noisy linear output measurements after the initial time. The ensemble Kalman inversion (EKI) method is a derivative-free iterative method for estimating the posterior distribution of this inference problem. However, accuracy of EKI depends on having a large ensemble of particles, where each particle requires evolving the dynamical system. When the system is high-dimensional, the cost per particle is high, leading to EKI either having prohibitive cost or high sampling error. In this work, we use Balanced truncation for Bayesian smoothing (BTBS) to accelerate solution of the smoothing problem via EKI. BTBS is a model reduction method that adapts balanced truncation, a system-theoretic projection-based model reduction method, to the Bayesian smoothing problem. Numerical results show that reduced EKI models achieve the same accuracy as the full EKI algorithm with multiple-orders-of-magnitude reduction in computational cost.

Pavlos Stavrinides
Georgia Tech
pstavrinides3@gatech.edu

Elizabeth Qian
Georgia Institute of Technology
eqian@gatech.edu

MS54

Novel Randomized Algorithms for Low- and Full-rank Decompositions and Their Implementations in RandLAPACK

We present a pair of novel randomized algorithms: a scheme for performing QR with column pivoting and a method for performing a partial SVD. The first algorithm, called ICQRRP, carefully uses randomized sketching to accelerate both pivot decisions for the input matrix and the process of decomposing the pivoted matrix via Cholesky QR. ICQRRP is applicable to matrices of any aspect ratio. The second algorithm, called ABRIK, improves partial SVD runtime by performing a few multiplications with large blocks of random vectors. Furthermore, ABRIK overcomes the limitation of the power iteration-based Randomized SVD, related to matrices with a slowly decaying spectrum. We implement the algorithms in RandLAPACK by calling into RandBLAS and vendor-provided BLAS/LAPACK libraries. Experiments with these implementations were performed on an Intel Xeon Gold 6245 CPU. ICQRRP demonstrates two orders-of-magnitude speedup relative to LAPACKs standard function for QRCP and superior performance to the state-of-the-art alternative randomized QRCP scheme; ABRIK exhibits an order-of-magnitude speedup over the alternative partial SVD solvers.

Max Melnichenko
University of Tennessee
mmelnic1@vols.utk.edu

Riley Murray
Sandia National Laboratories
rjmurr@sandia.gov

Robert Webber
California Institute of Technology
rwebber@ucsd.edu

MS54

Hutchinson's Estimator is Bad at Kronecker Trace Estimation

We study the problem of estimating the trace of a matrix A that can only be access via Kronecker-matrix-vector products. That is, we can compute Ax for any vector x that has Kronecker structure. In particular, we study how Hutchinson's Estimator performs in this setting, proving tight rates for the number of matrix-vector products this estimator needs to find a relative error approximation to the trace of A. We find an exact expression for the variance of this estimator, show this is unavoidably exponential, and conclude with evidence that a much faster non-Hutchinson

algorithm may exist.

Raphael A. Meyer
New York University
ram900@nyu.edu

Haim Avron
Tel Aviv University
haimav@tauex.tau.ac.il

MS54

Asymptotically Free Sketching and Applications in Ridge Regression

Classical results in sketching for dimensionality reduction in machine learning assert that, provided the sketch size is sufficiently large, the original unsketched solution is recovered at a fraction of the cost. However, in many practical settings, the sketch size may be smaller than needed for these guarantees. We provide a more general asymptotic result for sketched matrix inversion that holds for any sketch size and reveal that sketching is equivalent to adding ridge regularization. We prove our results for a broad class of asymptotically free sketches encompassing the spectral profiles of most sketches used in practice. We then determine the precise effect of sketching on the generalization error of ridge regression and show that the generalized cross-validation risk estimator is consistent for sketched ensembles, enabling the efficient evaluation of unsketched ridge regression risk using only sketched data.

Daniel LeJeune
Stanford University
daniel@dlej.net

Pratik Patil
Berkeley
pratikpatil@berkeley.edu

MS55

Learning the Interaction Kernel from a Mean-Field Opinion Model

The viewpoints and opinions of individuals continually shift and disseminate through social interactions in our daily lives. In the study of opinion dynamics, researchers frequently investigate the evolution of opinions as dynamical processes occurring on various network structures, such as graphs, hypergraphs, and multi-layer networks. However, when dealing with large networks, we encounter high-dimensional systems, posing challenges in inference tasks. In this presentation, I will adopt a mean-field perspective to address this challenge and demonstrate how to infer the interaction kernel between pairwise agents based on limited partial observations. Furthermore, I will discuss the implications of handling data originating from non-Markovian dynamics.

Weiqi Chu, Weiqi Chu
University of Massachusetts Amherst
weiqichu@umass.edu, weiqichu@umass.edu

Mason A. Porter
UCLA
mason@math.ucla.edu

MS55

Provable Approximations of Multivariate Functions on Smooth Manifolds Using Deep Relu Neural Networks

The impressive capacity of deep neural networks to approximate multivariate functions, thus overcoming the curse of dimensionality, is exemplified by their effectiveness in addressing high-dimensional problems where traditional numerical solvers fail. To offer a theoretical framework elucidating this phenomenon, we analyze the approximation of Hölder functions defined on a d-dimensional smooth manifold M embedded in R^D, with $d \ll D$, using deep neural networks. Our study, consistent with recent literature, identifies neural networks as a class of structured parametric functions. We establish new uniform convergence estimates for the approximation and generalization errors by deep neural networks with ReLU activation functions, showing that such estimates do not depend on the ambient dimension D of the function but only on the lower manifold dimension d, in a precise sense. This finding represents a significant improvement over existing estimates established in the literature in a similar setting, where approximation and generalization errors were found to depend weakly on the ambient dimension D.

Demetrio Labate
University of Houston
dlabate@math.uh.edu

Ji Shi
Department of Applied Mathematics
University of Houston
jshi24@cougarnet.uh.edu

MS55

Advantages of Weak-Form Equation Learning in Multiscale Model Inference and Reduced-Order Models for Parametric PDEs

Recent work in data-driven modeling has demonstrated that a weak formulation of model equations enhances the noise robustness of a wide range of computational methods. In this presentation, we demonstrate the power of the weak form to enhance the LaSDI (Latent Space Dynamics Identification) algorithm, a recently developed data-driven reduced order modeling technique. We introduce a weak form-based version of WLaSDI (Weak-form Latent Space Dynamics Identification). WLaSDI first compresses data, then projects onto the test functions and learns the local latent space models. Notably, WLaSDI demonstrates significantly enhanced robustness to noise. WLaSDI obtains the local latent space using weak-form equation learning. Compared to the standard sparse identification of nonlinear dynamics (SINDy) used in LaSDI, the variance reduction of the weak form guarantees a robust and precise latent space recovery, hence allowing for a fast, robust, and accurate simulation.

Daniel Messenger
Department of Applied Mathematics
University of Colorado - Boulder
daniel.messenger@colorado.edu

MS55

Finite Expression Method: A Symbolic Approach for Scientific Machine Learning

Human-designed algorithms have long been fundamental in solving a variety of scientific and engineering challenges. Recently, data-driven deep learning methods have also

risen to prominence, offering innovative solutions across numerous scientific fields. While traditional algorithms excel in capturing the core aspects of specific problems, they often lack the flexibility needed for varying problem conditions due to the absence of specific data. Conversely, while data-driven approaches utilize vast datasets, they frequently fall short in domain-specific knowledge. To bridge these gaps, we introduce **FMint** (Foundation Model based on Initialization), a generative pre-trained model that synergizes the precision of human-designed algorithms with the adaptability of data-driven methods. This model is specifically engineered for high-accuracy simulation of dynamical systems. Starting from initial trajectories provided by conventional methods, FMint quickly delivers highly accurate solutions. It incorporates in-context learning and has been pre-trained on a diverse corpus of 500,000 dynamical systems, showcasing exceptional generalization across a broad spectrum of real-world applications. By effectively combining algorithmic rigor with data-driven flexibility, FMint sets the stage for the next generation of scientific foundation models, tackling complex problems with both efficiency and high accuracy.

Haizhao Yang
University of Maryland, College Park
hzyang@umd.edu

MS56

A Stochastic Maximum Principle Approach for Reinforcement Learning with Parameterized Environment.

In this work, we introduce a stochastic maximum principle (SMP) approach for solving the reinforcement learning problem with the assumption that the unknowns in the environment can be parameterized based on physics knowledge. For the development of numerical algorithms, we apply an effective online parameter estimation method as our exploration technique to estimate the environment parameter during the training procedure, and the exploitation for the optimal policy is achieved by an efficient backward action learning method for policy improvement under the SMP framework. Numerical experiments are presented to demonstrate that the SMP approach for reinforcement learning can produce reliable control policy, and the gradient descent type optimization in the SMP solver requires less training episodes compared with the standard dynamic programming principle based methods.

Siming Liang, Feng Bao
Florida State University
sliang@fsu.edu, bao@math.fsu.edu

MS56

Convergence of Policy Improvement Algorithm for Exploratory Stochastic Control Problems

In this paper we provide a simple proof for the convergence of Policy Improvement Algorithm (PIA) for reinforcement learning for a general continuous time entropy-regularized stochastic control problem. Compared to the existing results, our arguments rely largely on the connection between the parabolic PDEs and the backward SDEs, especially the representation formulae for the PDE solutions and their derivatives. Our approach is particularly effective when the diffusion coefficient is free of control, for which an iterative systems of (linear) PDEs can be obtained via PI iterative scheme, with the help of the Gibbs form of policy iteration and the Feynman-Kac formula. With the help of a representation theorem in Ma & Zhang's previous work, we can obtain the desired uniform bounds of the solution sequence of the PDEs as well as their derivatives rather easily, which is essential for the desired convergence of the PIA, especially for establishing some crucial and novel rates of convergence. Our method applied to both finite and infinite horizon cases, or in other words, both parabolic and elliptic PDEs.

Gaozhan Wang, Jianfeng Zhang, Ma Jin
University of Southern California
gaozhanw@usc.edu, jianfenz@usc.edu, jinma@usc.edu

MS56

Reinforcement Learning for Optimal Stopping Problems in Continuous Time

Optimal stopping is the problem of finding the right time to take a particular action in a stochastic system, in order to maximize an expected reward. It has applications in areas such as finance, healthcare, and statistics. In this talk, we investigate how to learn to make optimal stopping decisions in an unknown stochastic system via a novel reinforcement learning framework. In particular, in order to encourage exploration in the stochastic system, we randomize the stopping time through cumulative residual entropy, resulting in a singular control problem with special structures. We will discuss the regularity of the solution, the convergence of our proposed learning algorithm, and the algorithm performance on a real option example. This is based on joint work with Jodi Dianetti and Giorgio Ferrari (Bielefeld University).

Renyuan Xu
New York University
renyuanx@usc.edu

MS56

Phibe: A PDE-Based Bellman Equation for Continuous-Time Reinforcement Learning

In this talk, we address the problem of continuous-time reinforcement learning in scenarios where the dynamics follow a stochastic differential equation. When the underlying dynamics remain unknown, and we have access only to discrete-time information, how can we effectively conduct policy evaluation? We first highlight that the commonly used Bellman equation is not always a reliable approximation to the true value function. We then introduce PhiBE, a PDE-based Bellman equation that offers a more accurate approximation to the true value function, especially in scenarios where the underlying dynamics change slowly. Moreover, we extend PhiBE to higher orders, providing increasingly accurate approximations. Additionally, we present a model-free algorithm to solve PhiBE when only discrete-time trajectory data is available. Numerical experiments are provided to validate the theoretical guarantees we propose.

Yuhua Zhu
University of California, Los Angeles
yuhuazhu@ucla.edu

MS57

Conditional Neural Field Latent Diffusion-Based Predictive Modeling of Stochastic Spatiotemporal Physics.

This study introduces the Conditional Neural Field La-

tent Diffusion (CoNFiLD) model, a novel generative learning framework designed for rapid simulation of intricate spatiotemporal dynamics in chaotic and turbulent systems within three-dimensional irregular domains. Traditional eddy-resolved numerical simulations encounter significant limitations due to their extensive computational demands. In contrast, deep learning-based surrogate models promise efficient, data-driven solutions. However, their effectiveness is often compromised by a reliance on deterministic frameworks, which fall short in accurately capturing the chaotic and stochastic nature of turbulence. The CoNFiLD model addresses these challenges by synergistically integrating conditional neural field encoding with latent diffusion processes, enabling the memory-efficient and robust probabilistic generation of spatiotemporal turbulence under varied conditions. Leveraging Bayesian conditional sampling, the model can seamlessly adapt to a diverse range of turbulence generation scenarios without the necessity for retraining, covering applications from zero-shot full-field flow reconstruction using sparse sensor measurements to super-resolution generation and spatiotemporal flow data restoration. Comprehensive numerical experiments across a variety of inhomogeneous, anisotropic turbulent flows with irregular geometries have been conducted to demonstrate the model's versatility and efficacy.

Pan Du, Meet Parikh, Xiantao Fan, Xin-Yang Liu
University of Notre Dame
pdu@nd.edu, mparikh2@nd.edu, fanxiantao1@gmail.com, xfan2@nd.edu

Jian-Xun Wang
Dept. of Aerospace and Mechanical Engineering
University of Notre Dame
jwang33@nd.edu

MS57

Simplifying Full Waveform Inversion Via Domain-Independent Self-Supervised Learning

Deep learning has marked a significant advancement in geophysics, especially in tackling the intricate challenge of full waveform inversion (FWI). This breakthrough has enabled the effective prediction of subsurface velocity maps from seismic data. The process involves transforming seismic data into subsurface velocity maps, a task we have approached as a sophisticated form of image translation. This paper discusses a remarkable discovery: when encoders and decoders are trained independently within their specific domains through self-supervised learning, a linear relationship emerges in their latent spaces, transcending domain boundaries. This finding not only sheds light on the underlying mechanics of FWI but also elegantly unites multiple FWI datasets. These datasets can efficiently utilize a shared set of self-trained encoders and decoders, each adapted with distinct linear mappings. Building upon these insights, we introduce SimFWI, a novel methodology comprising two key steps: firstly, the independent learning of seismic encoders and velocity decoders across various datasets using masked image modeling, and secondly, the customization of a linear mapping for each dataset. Our experiments reveal that SimFWI matches the performance of traditionally trained models, which rely on paired seismic data and velocity maps, thereby opening new avenues in geophysical research.

Youzou Lin
University of North Carolina
yzlin@unc.edu

MS57

One-Shot Learning for Solution Operators of Partial Differential Equations.

Learning and solving governing equations of a physical system, represented by partial differential equations (PDEs), from data is a central challenge in a variety of areas of science and engineering. Current methods require either some prior knowledge (e.g., candidate PDE terms) to discover the PDE form, or they need a large dataset to learn a surrogate model of the PDE solution operator. Here, we propose the first solution operator learning method that only requires one PDE solution, i.e., one-shot learning. We first decompose the entire computational domain into small domains, where we learn a local solution operator, and then we find the solution of a new input function via mesh-based fixed-point iteration (FPI), meshfree local-solution-operator informed neural network (LOINN) or local-solution-operator informed neural network with correction (cLOINN). We tested our method on 7 different PDEs, including linear or nonlinear PDEs, PDEs defined on complex geometries, and PDE systems. Our method demonstrates effectiveness and generalization capabilities across these varied scenarios.

Lu Lu
Yale University
lu.lu@yale.edu

MS58

Deep Learning Models for Accelerating Phase-Field Simulations

Computational simulation of phase field dynamics can be prohibitively expensive when using standard numerical solvers. High-fidelity simulations often use very small time steps due to stability considerations, which can become a bottleneck when the target quantities of interest require predictions over long time horizons. To address this challenge, we employ machine learning-based surrogate models to help extrapolate forward in time, enabling predictions at time horizons far beyond what is achievable through traditional methods alone. Specifically, we investigate two deep learning architectures, Fourier Neural Operators (FNOs) and UNets, and train them to predict future states with much coarser time steps thus encapsulating multiple high-fidelity steps within a single surrogate evaluation. While this approach enables more rapid predictions through autoregressive evaluation of the surrogate, the incurred error is essentially uncontrolled. To alleviate this, we adopt a hybrid prediction strategy which alternates between surrogate evaluations which leap forward in time and direct numerical simulation steps which reduce errors and bring the system state back to the solution manifold. We illustrate these methods on phase-field simulations for a liquid metal dealloying system.

Cosmin Safta
Sandia National Laboratories
csafta@sandia.gov

Christophe Bonneville
Sandia National Laboratories, U.S.
cpbonne@sandia.gov

Habib N. Najm
Sandia National Laboratories

Livermore, CA, USA
hnnajm@sandia.gov

Laurent Capolungo
Los Alamos National Laboratory
laurent@lanl.gov

Nathan Bieberdorf, Mark Asta
University of California, Berkeley
nbieberdorf@berkeley.edu, mdasta@berkeley.edu

MS58

Data Driven Modeling for Stochastic Systems

We present a numerical framework for learning unknown stochastic dynamical systems using measurement data. Termed stochastic flow map learning (sFML), the new framework seeks to approximate the unknown flow map of the underlying system. Technically, it is realized by (conditional) generative models, such as Generative Adversarial networks (GANs), Autoencoders, and Normalizing Flows. Once a sFML model is trained, it serves as a stochastic evolution model that is a weak approximation, in terms of distribution, of the unknown stochastic system. It allows us to analyze the long-term system behavior under different initial conditions. A comprehensive set of numerical examples is presented to demonstrate the flexibility and effectiveness of the proposed sFML method for various types of stochastic systems. It is capable of handling systems driven by both Gaussian and non-Gaussian noises, even for jump processes, such as systems generated by Gillespie's stochastic simulations.

Yuan Chen
Ohio State University
chen.11050@buckeyemail.osu.edu

MS58

Generalization Aspects of Diffusion Models

Diffusion models, particularly score-based generative models (SGMs), have emerged as powerful tools in diverse machine learning applications, spanning from computer vision to modern language processing. In this talk, we delve into the generalization theory of SGMs, exploring their capacity for learning high-dimensional distributions. Our analysis show that SGMs achieve a dimension-free generation error bound when applied to a class of sub-Gaussian distributions characterized by certain low-complexity structures.

Yulong Lu
University of Minnesota, Twin Cities
yulonglu@umn.edu

MS58

Nonlinear Ensemble Filtering with Diffusion Models: Application to the Surface Quasi-Geostrophic Dynamics

The intersection between classical data assimilation methods and novel machine learning techniques has attracted significant interest in recent years. Here we explore another promising solution in which diffusion models are used to formulate a robust nonlinear ensemble filter for sequential data assimilation. Unlike standard machine learning methods, the proposed Ensemble Score Filter (EnSF) is completely training-free and can efficiently generate a set of analysis ensemble members. In this study, we apply the EnSF to a surface quasi-geostrophic model and compare its performance against the popular Local Ensemble Transform Kalman Filter (LETKF), which makes Gaussian assumptions on the posterior distribution. Numerical tests demonstrate that EnSF maintains stable performance in the absence of localization and for a variety of experimental settings. We find that EnSF achieves competitive performance relative to LETKF in the case of linear observations but leads to significant advantages when the state is nonlinearly observed, and the numerical model is subject to unexpected shocks. A spectral decomposition of the analysis results shows that the largest improvements over LETKF occur at large scales (small wavenumbers) where LETKF lacks sufficient ensemble spread. Overall, this initial application of EnSF to a geophysical model of intermediate complexity is very encouraging, and motivates further developments of the algorithm for more realistic problems.

Zixiang Xiong
Department of Mathematics
Florida State University
zx18@fsu.edu

MS59

Catapult Dynamics and Feature Learning in Neural Networks

I will discuss an explanation regarding the common occurrence of spikes in the training loss when neural networks are trained with stochastic gradient descent (SGD). We provide evidence that the spikes in the training loss of SGD are "catapults", an optimization phenomenon originally observed in GD with large learning rates in [Lewkowycz et al. 2020]. We empirically show that these catapults occur in a low-dimensional subspace spanned by the top eigenvectors of the tangent kernel, for both GD and SGD. Second, we posit an explanation for how catapults lead to better generalization by demonstrating that catapults promote feature learning by increasing alignment with the Average Gradient Outer Product (AGOP) of the true predictor. Furthermore, we demonstrate that a smaller batch size in SGD induces a larger number of catapults, thereby improving AGOP alignment and test performance. Joint work with Libin Zhu, Chaoyue Liu, Adityanarayanan Radhakrishnan.

Mikhail Belkin
University of California, San Diego, U.S.
mbelkin@ucsd.edu

MS59

Rectified Flow: A Straight Approach to Fast Generative Modeling and Optimal Transport

We consider the problem of learning a transport mapping between two distributions that are only observed through unpaired data points. This problem provides a unified framework for a variety of fundamental tasks in machine learning: generative modeling is about transforming an elementary (such as Gaussian) random variable to realistic data points; domain transfer concerns with transferring data points from one domain to another; optimal transport (OT) solves the more challenging problem of finding a "best" transport map that minimizes certain transport cost. Unfortunately, despite the unified view, there lacks an algorithm that can solve the transport mapping problem efficiently in all settings. The existing algorithms need to be developed case by case, and tend to be complicated

or computationally expensive. In this talk, I will show you that the problem can be addressed with a pretty simple algorithm. This algorithm, called rectified flow (RF), learns an ordinary differential equation (ODE) model to transfer between the two distributions by following straight paths as much as possible. The algorithm only requires solving a sequence of nonlinear least squares optimization problems, which guarantees to yield monotonically nonincreasing couplings w.r.t. all convex transport costs. The straight paths are special and preferred because they are the shortest paths between two points, and can be simulated exactly without time discretization, yielding computationally efficient models. RF provides a simple and clean framework of generative modeling and enables ultrafast one step generation via a reflow procedure, which is a significant speedup over traditional diffusion generative models. Moreover, with a proper modification, RF yields a tool to solve the optimal transport problems on high dimensional continuous distributions, a challenging problem for which no well accepted efficient algorithms exist.

Qiang Liu
The University of Texas at Austin
lqiang@utexas.edu

MS59

From Mixing Time to Sample Complexity: Elucidating the Design Space of Score-Based Losses

Score-based losses have emerged as a more computationally appealing alternative to maximum likelihood for fitting (probabilistic) generative models with an intractable likelihood (for example, energy-based models and diffusion models). What is gained by foregoing maximum likelihood is a tractable gradient-based training algorithm. What is lost is less clear: in particular, since maximum likelihood is asymptotically optimal in terms of statistical efficiency, how suboptimal are score-based losses? I will survey a recently developing connection relating the *statistical efficiency* of broad families of generalized score losses, to the *algorithmic efficiency* of a natural inference-time algorithm: namely, the mixing time of a suitable Markov chain using the score that can be used to draw samples from the model. This dictionary allows us to elucidate the design space for score losses with good statistical behavior, by translating techniques for speeding up Markov chain convergence (e.g., preconditioning and lifting).

Andrej Risteski
Carnegie Mellon University
Machine Learning Department
aristesk@andrew.cmu.edu

MS59

The Mechanism Behind the Implicit Biases of Large Learning Rates: Edge of Stability, Balancing, and Catapult

Large learning rates, when applied to gradient descent for nonconvex optimization, yield various implicit biases, including edge of stability, balancing, and catapult. There are a lot of theoretical works trying to analyze these phenomena, while the high level idea is still missing: it is unclear when and why these phenomena occur. In this talk, I will show that these phenomena are actually various tips of the same iceberg. They occur when the objective function of optimization has some good regularity. This regularity, together with the effect of large learning rate on guiding gradient descent from sharp regions to flatter ones, leads to the control of the largest eigenvalue of Hessian, i.e., sharpness, along the GD trajectory, which results in various phenomena. The result is based on the nontrivial convergence analysis under large learning rate on a family of nonconvex functions of various regularities without Lipschitz gradient which is usually a default assumption in nonconvex optimization. In addition, it contains the first non-asymptotic result on the rate of convergence in this circumstance. Neural network experiments will also be presented to validate this result.

Yuqing Wang, Zhenghao Xu, Tuo Zhao
Georgia Institute of Technology
ywang3398@gatech.edu, zhenghaoxu@gatech.edu, tourzhao@gatech.edu

Molei Tao, Molei Tao
School of Mathematics, Georgia Institute of Technology
mtao@gatech.edu, mtao@gatech.edu

MS61

Kernel Limit of Recurrent Neural Networks Trained on Ergodic Data Sequences

Mathematical methods are developed to characterize the asymptotics of recurrent neural networks (RNN) as the number of hidden units, data samples in the sequence, hidden state updates, and training steps simultaneously grow to infinity. In the case of an RNN with a simplified weight matrix, we prove the convergence of the RNN to the solution of an infinite-dimensional ODE coupled with the fixed point of a random algebraic equation. The analysis requires addressing several challenges which are unique to RNNs. In typical mean-field applications (e.g., feedforward neural networks), discrete updates are of magnitude O(1/N) and the number of updates is O(N). Therefore, the system can be represented as an Euler approximation of an appropriate ODE/PDE, which it will converge to as $N \to \infty$. However, the RNN hidden layer updates are O(1). Therefore, RNNs cannot be represented as a discretization of an ODE/PDE and standard mean-field techniques cannot be applied. Instead, we develop a fixed point analysis for the evolution of the RNN memory states, with convergence estimates in terms of the number of update steps and the number of hidden units. The RNN hidden layer is studied as a function in a Sobolev space, whose evolution is governed by the data sequence (a Markov chain), the parameter updates, and its dependence on the RNN hidden layer at the previous time step.

Samuel Chun Hei Lam
University of Oxford
lam@maths.ox.ac.uk

MS61

Information Gamma Calculus: Convexity Analysis for Stochastic Differential Equations

We study the Lyapunov convergence analysis for degenerate and non-reversible stochastic differential equations (SDEs). We apply the Lyapunov method to the Fokker-Planck equation, in which the Lyapunov functional is chosen as a weighted relative Fisher information functional. We derive a structure condition and formulate the Lyapunov constant explicitly. Given the positive Lyapunov constant, we prove the exponential convergence result for the probability density function towards its invariant distribution in the L1 norm. Several examples are presented: underdamped Langevin dynamics with variable diffusion

matrices, quantum SDEs in Lie groups (Heisenberg group, displacement group, and Martinet sub-Riemannian structure), three oscillator chain models with nearest-neighbor couplings, and underdamped mean field Langevin dynamics (weakly self-consistent VlasovFokkerPlanck equations).

Wuchen Li
University of South Carolina
wuchen@mailbox.sc.edu

MS61

From Langevin Based Optimizers to Diffusion Based Generative Models

We will review some recent advances in the field of Langevin based optimizers with emphasis on their application in training neural nets. We will highlight interesting links between this class of stochastic optimizers and diffusion-based (known also as scored-based) generative models.

Sotirios Sabanis
University of Edinburgh
Alan Turing Institute
s.sabanis@ed.ac.uk

MS61

High Order Normalizing Flow and Neural ODE for Mean Field Control Problems

We present a novel high-order normalizing flow approach to solve the mean field control problems. We derive a system of ordinary differential equations, which is a generalization of the neural ODE. With this system, we are able to compute the evolution of the not only the density but also the score function efficiently. Numerical demonstrations confirm the accuracy of our algorithm in capturing the dynamics of these elements.

Mo Zhou
University of California, Los Angeles
mozhou366@math.ucla.edu

MS62

Federated Learning with the Differential Privacy and More

In the burgeoning field of machine learning, federated learning (FL) has emerged as a pivotal paradigm for training models directly on distributed local datasets without centralizing them, thus mitigating privacy leakage. However, challenges remain in ensuring that the machine learning process itself does not leak sensitive information. This talk explores the integration of differential privacy within FL to address these challenges. We will discuss the principles of differential privacy and its implementation techniques that help protect participants local datasets during the FL process. Further, we will introduce recent advancements in this approach, highlighting their effect on the trade-off between model accuracy and privacy loss.

Rui Hu
University of Nevada, Reno
ruihu@unr.edu

MS62

Gradient-Based Methods for Explanation and Confidence Metrics in Distributed Learning

Gradient-based methods provide a practical computational approach for obtaining explanations of AI models. Algorithms based on path integrals have been shown to satisfy commonsense axiomatic properties similar to those expected from Shapley values in collaborative game theory. In this talk, we present a new gradient-based method for explaining AI decisions using integrated decision gradients an adaptive approach to computing path integrals that leads to robust explanation. In particular, the approach resolves the saturation problem of classical integrated gradients by employing an importance factor to focus on the area of the path integral where the model makes its decision. This approach naturally leads to an efficient attribution-based confidence metric for AI models. Finally, we demonstrate how these approaches naturally extend to distributed learning models. The talk will build upon our recent results published at NeurIPS19, IJCAI22, AAAI24, and DAC24.

Sumit K. Jha
Florida International University
sumit.jha@fiu.edu

MS62

Unveiling Threat Models in Gradient Inversion Attacks under Federated Learning

Federated Learning (FL) has emerged as a leading paradigm for decentralized, privacy preserving machine learning training. However, recent research on gradient inversion attacks (GIAs) have shown that gradient updates in FL can leak information on private training samples. While existing surveys on GIAs have focused on the honest-but-curious server threat model, there is a dearth of research categorizing attacks under the realistic and far more privacy-infringing cases of malicious servers and clients. Here we present a survey and novel taxonomy of GIAs that emphasize FL threat models, particularly that of malicious servers and clients. We first formally define GIAs and contrast conventional attacks with the malicious attacker. We then summarize existing honest-but-curious attack strategies, corresponding defenses, and evaluation metrics. Critically, we dive into attacks with malicious servers and clients to highlight how they break existing FL defenses, focusing specifically on reconstruction methods, target model architectures, target data, and evaluation metrics. Lastly, we discuss open problems and future research directions.

Olivera Kotevska
Oak Ridge National Laboratory
kotevskao@ornl.gov

Yichuan Shi
Massachusetts Institute of Technology
yshi23@mit.edu

Viktor Reshniak
Oak Ridge National Laboratory
reshniakv@ornl.gov

Abhishek Singh
Massachusetts Institute of Technology
abhi24@media.mit.edu

MS62

Privacy for the Activation Release in Collaborative

Training and Inference

There has been a recent advent of generative AI methods powered by transformers and large foundation models. This has been overshadowed by serious concerns about privacy-sensitivities of the raw datasets during training, fine-tuning and prompt tuning while maintaining computational efficiency in client-server settings. This necessitates new mechanisms of training with formal privacy guarantees for releasing intermediate activations (embeddings), by clients that hold the sensitive data to a computationally more powerful server. The privacy concern for NLP modalities, have arisen due to lack of privacy guarantees at the token level, word embedding level and sentence embedding level. Current differential privacy methods are applied upon extracting the pre-trained embeddings from a standard NLP architecture, while treating these embeddings as tabular data. As differential privacy is not immune to pre-processing, the privacy guarantees applied at this attack surface do not translate to privacy guarantees at the token, word or sentence level. Moreover, preventing model identifiability in this setting of inference serving systems where a server holds a model zoo of multiple models via latency and accuracy fingerprinting attacks is yet another important adjacent problem. This talk provides formal methods to cater to such relevant real-world problems in collaborative training and inference.

Praneeth Vepakomma
MIT
MBZUAI
vepakom@mit.edu

MS63

Kernel Methods Are Competitive for Operator Learning

We introduce a kernel-based framework for learning operators between Banach spaces. We show that even with simple kernels, our approach is competitive in terms of cost-accuracy trade-off and either matches or beats the performance of Neural Network methods on a majority of PDE-based benchmarks. Additionally, our framework offers several advantages inherited from kernel methods: simplicity, convergence guarantees, a priori error estimates, and Bayesian UQ. It is, therefore, a natural benchmark for operator learning problems.

Matthieu Darcy
Caltech
mdarcy@caltech.edu

MS63

Sample-Complexity Bounds for Operator Learning

Operator learning frameworks leverage neural networks and define a methodology for the data-driven approximation of operators. How much data is necessary to learn operators in such a purely data-driven manner? In this presentation, I will summarize recent work that provides partial answers to this question. Upper and lower bounds on the data-complexity of operator learning will be discussed.

Samuel Lanthaler
California Institute of Technology
slanth@caltech.edu

MS63

Any-Dimensional Equivariant Learning

Traditional supervised learning aims to learn an unknown mapping by fitting a function to a set of input-output pairs with a fixed dimension. The fitted function is then defined on inputs of the same dimension. However, in many settings, the unknown mapping takes inputs in any dimension; examples include graph parameters defined on graphs of any size and physics quantities defined on an arbitrary number of particles. We leverage a newly-discovered phenomenon in algebraic topology called representation stability to define equivariant neural networks that can be trained with data in a fixed dimension and then extended to accept inputs in any dimension. Our approach is black-box and user-friendly, requiring only the network architecture and the groups for equivariance, and can be combined with any training procedure.

Eitan Levin
Caltech
eitanl@caltech.edu

MS63

A Poincar Inequality and Consistency Results for Signal Sampling on Large Graphs

Large-scale graph machine learning is challenging as the complexity of learning models scales with the graph size. Subsampling the graph is a viable alternative, but sampling on graphs is nontrivial as graphs are non-Euclidean. Existing graph sampling techniques require not only computing the spectra of large matrices but also repeating these computations when the graph changes, e.g., grows. In this paper, we introduce a signal sampling theory for a type of graph limit—the graphon. We prove a Poincar inequality for graphon signals and show that complements of node subsets satisfying this inequality are unique sampling sets for Paley-Wiener spaces of graphon signals. Exploiting connections with spectral clustering and Gaussian elimination, we prove that such sampling sets are consistent in the sense that unique sampling sets on a convergent graph sequence converge to unique sampling sets on the graphon. We then propose a related graphon signal sampling algorithm for large graphs, and demonstrate its good empirical performance on graph machine learning tasks

Luana Ruiz
Department of Applied Mathematics and Statistics
Johns Hopkins University
lrubini1@jh.edu

MS64

Linear-Cost Vecchia Approximation of Multivariate Normal Probabilities

Multivariate normal (MVN) probabilities arise in myriad applications, but they are analytically intractable and need to be evaluated via Monte-Carlo-based numerical integration. For the state-of-the-art minimax exponential tilting (MET) method, we show that the complexity of each of its components can be greatly reduced through an integrand parameterization that utilizes the sparse inverse Cholesky factor produced by the Vecchia approximation, whose approximation error is often negligible relative to the Monte-Carlo error. Based on this idea, we derive al-

gorithms that can estimate MVN probabilities and sample from truncated MVN distributions in linear time (and that are easily parallelizable) at the same convergence or acceptance rate as MET, whose complexity is cubic in the dimension of the MVN probability. We showcase the advantages of our methods relative to existing approaches using several simulated examples. We also analyze a groundwater-contamination dataset with over twenty thousand censored measurements to demonstrate the scalability of our method for partially censored Gaussian-process models.

Jian Cao
University of Houston
jcao21@central.uh.edu

Matthias Katzfuss
University of WisconsinMadison
katzfuss@gmail.com

MS64

Enhanced Gaussian Process Surrogates for Optimization and Sampling by Pure Exploration

In this talk, we propose novel noise-free Bayesian optimization strategies that rely on a random exploration step to enhance the accuracy of Gaussian process surrogate models. The new algorithms retain the ease of implementation of the classical GP-UCB algorithm, but the additional random exploration step accelerates their convergence, nearly achieving the optimal convergence rate. Furthermore, to facilitate Bayesian inference with an intractable likelihood, we propose to utilize the optimization iterates as design points to build a Gaussian process surrogate model for the unnormalized log-posterior density. We provide bounds for the Hellinger distance between the true and the approximate posterior distributions in terms of the number of design points. The effectiveness of our algorithms is demonstrated in benchmark non-convex test functions for optimization, and in a black-box engineering design problem. We also showcase the effectiveness of our posterior approximation approach in Bayesian inference for parameters of dynamical systems.

Hwanwoo Kim
University of Chicago
hwanwoo.kim@duke.edu

MS64

Sampling from Gaussian Process Posteriors Using Stochastic Gradient Descent

The ability to deploy Gaussian-process-based decision-making systems such as Bayesian optimization at scale has traditionally been limited by computational costs arising from the need to solve large linear systems. The de-facto standard for solving linear systems at scale is via the conjugate gradient algorithm - in particular, stochastic gradient descent is known to converge near-arbitrarily-slowly on quadratic objectives that correspond to Gaussian process models linear systems. In spite of this, we show that it produces solutions which have low test error, and quantify uncertainty in a manner that mirrors the true posterior. We develop a spectral characterization of the error caused by finite-time non-convergence, which we prove is small both near the data, and sufficiently far from the data. Stochastic gradient descent therefore only differs from the true posterior between these regions, demonstrating a form of implicit bias caused by benign non-convergence. We conclude by showing, empirically, that stochastic gradient descent achieves state-of-the-art performance on sufficiently large-scale regression tasks, and produces uncertainty estimates which match the performance of significantly more expensive baselines on large-scale Bayesian optimization.

Alexander Terenin
Cornell University
avt28@cornell.edu

MS64

Asymptotic Theory for Linear Functionals of Kernel Ridge Regression

An asymptotic theory is established for linear functionals of the predictive function given by kernel ridge regression, when the reproducing kernel Hilbert space is equivalent to a Sobolev space. The theory covers a wide variety of linear functionals, including point evaluations, evaluation of derivatives, L_2 inner products, etc. We establish the upper and lower bounds of the estimates and their asymptotic normality. It is shown that $\lambda \sim n^{-1}$ is the universal optimal order of magnitude for the smoothing parameter to balance the variance and the worst-case bias. The theory also implies that the optimal L_∞ error of kernel ridge regression can be attained under the optimal smoothing parameter $\lambda \sim n^{-1} \log n$. These optimal rates for the smoothing parameter differ from the known optimal rate $\lambda \sim n^{-\frac{2m}{2m+d}}$ that minimizes the L_2 error of the kernel ridge regression.

Rui Tuo
Texas A&M University
ruituo@tamu.edu

Lu Zou
Hong Kong University of Science and Technology (Guangzhou)
lzou@connect.ust.hk

MS65

Structure-exploiting Neural Operators for Bayesian Inverse Problems and Optimal Control/Design Under Uncertainty

Deep neural networks (DNNs) have emerged as leading contenders for overcoming the challenges of constructing infinite dimensional surrogate models, which are known as neural operators. Black box application of DNNs for problems with infinite dimensional parameter fields may lead to inaccurate surrogates when training data are limited due to the expense of the model. Instead, by constructing a network architecture that captures the geometry of the map—in particular its smoothness, anisotropy, intrinsic low-dimensionality, and sensitivity—one can construct a dimension-independent reduced basis neural operator that is both accurate and optimization-aware using limited training data. We employ this reduced basis neural operator to make tractable the solution of PDE-constrained Bayesian inverse problems and optimal control & design under uncertainty, with application to problems governed by wave propagation, hyperelasticity, and viscous flows.

Lianghao Cao
California Institute of Technology
lianghao@caltech.edu

Peng Chen
Georgia Institute of Technology

pchen402@gatech.edu

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

Dingcheng Luo
Oden Institute for Computational Engineering and Sciences,
The University of Texas at Austin
dc.luo@utexas.edu

Thomas O'Leary-Roseberry
The University of Texas at Austin
Oden Institute for Computational and Engineering Sciences
tom@oden.utexas.edu

Umberto Villa
Oden Institute for Computational Engineering & Sciences
uvilla@oden.utexas.edu

MS65

Enabling Simulation and Optimization of Complex Systems Via Scientific Machine Learning

Complex systems are hard to simulation and even more difficult to optimize. In this talk, I will attempt an overview of data-driven scientific machine learning for the optimization of complex systems. In particular, I will showcase how surrogate models accelerate the evaluation of properties of solutions to partial differential equations. I will present a precise definition of the computational benefit of surrogate models and example surrogate models. We will then show how surrogate models can be combined to solve a challenging multiscale problem in optics. We will show that, through a synergistic combination of data-driven methods and direct numerical simulations, surrogate-based models present a data-efficient and physics-enhanced approach to simulating and optimizing complex systems. This approach has the benefit of being interpretable.

Raphaël Pestourie
MIT Mathematics
rpestourie3@gatech.edu

MS65

Multifidelity Linear Regression for Scientific Machine Learning from Scarce Data

Machine learning (ML) methods have garnered significant interest as potential methods for learning surrogate models for complex engineering systems for which traditional simulation is expensive. However, in many scientific and engineering settings, training data are scarce due to the cost of generating data from traditional high-fidelity simulations. ML models trained on scarce data have high variance and are sensitive to vagaries of the training data set. We propose a new multifidelity training approach for scientific machine learning that exploits the scientific context where data of varying fidelities and costs are available; for example high-fidelity data may be generated by an expensive fully resolved physics simulation whereas lower-fidelity data may arise from a cheaper model based on simplifying assumptions. We use the multifidelity data to define new multifidelity Monte Carlo estimators for the unknown parameters of linear regression models, and provide theoretical analyses that guarantee accuracy and improved robustness to small training budgets. Numerical results show that multifidelity learned models achieve order-of-magnitude lower expected error than standard training approaches when high-fidelity data are scarce.

Elizabeth Qian
Georgia Institute of Technology
eqian@gatech.edu

MS65

Fast Maxwell Solvers on Approximate Geometries for Interpretable Scientific Machine Learning and Nanophotonic Optimization

We present a fast solver for Maxwell's equations on approximate geometries that can leverage machine learning to explore the design space of high-dimensional shape optimization problems common in nanophotonics. Our solver qualitatively captures the physics of wave scattering in meta-materials using a model of layered, periodic surface impedances, and we train a neural network to generate these structures using the Physics-Enhanced Deep Surrogates (PEDS) architecture. The amount of data required to train a surrogate is reduced by the physical inductive bias embedded in our solver, yet we are able to calculate the transmission of light through complex, 3D devices much more efficiently than full-wave simulations due to the simple geometry. Furthermore, this architecture is inherently interpretable because the neural network generates structures that we can characterize using standard optical measurement techniques, such as ellipsometry and spectroscopy. We anticipate unlocking applications of photonic nanostructures in 3D by applying this surrogate model to large-scale inverse design problems.

Lorenzo X. Van Muñoz
Massachusetts Institute of Technology
Department of Physics
lxvm@mit.edu

Raphaël Pestourie
MIT Mathematics
rpestourie3@gatech.edu

Steven Johnson
MIT
stevenj@math.mit.edu

MS66

Solving Linear Systems Faster with Random Sampling and Sketching

Solving systems of linear equations has numerous applications across many areas, from data science and machine learning, to scientific computing, engineering and more. When these systems are too large to solve directly, iterative refinement methods have proven to be a powerful alternative. In this talk, I will present new algorithms and convergence guarantees for solving linear systems via sketch-and-project, a framework which unifies many known iterative methods that use randomized sampling and sketching, including randomized Kaczmarz, coordinate descent, and others. Our new results uncover a connection between stochastic iterative solvers and sketching-based randomized preconditioning algorithms: Whenever the spectral structure of a linear system is amenable to constructing a strong preconditioner via low-rank approximation, then one can construct a stochastic solver based on sketch-and-project that will implicitly take advantage of this spectral struc-

ture. In particular, I will show how this leads to solving an $n \times n$ linear system with at most k large (outlying) singular values in $\tilde{\mathcal{O}}\left(n^2+nk^2\right)$ arithmetic operations, which is faster than the $\tilde{\mathcal{O}}\left(n^2k\right)$ cost of constructing a good preconditioner for a deterministic iterative solver such as conjugate gradient.

Michal Derezinski
University of Michigan
derezin@umich.edu

MS66

Data-Driven Subgroup Identification for Linear Regression

Medical studies frequently require extracting the relationship between each covariate and a specified outcome. A commonly used approach is to fit a parametric model (e.g., linear regression) to a dataset, after which the model parameters can be used to draw qualitative conclusions about the data. However, it is common that the covariates may not have a uniform effect over the whole population and thus a unified simple model can miss the heterogeneous signal. For example, a linear model may be able to explain a subset of the data but fail on the rest due to the nonlinearity and heterogeneity in the data. In this talk, I will discuss methods for finding subpopulations of the data in which there is a simple, uniform (e.g., linear) relationship between the label and the covariates. Our algorithm, DDGroup (data-drive group discovery) outputs an interpretable region in the feature space in which a linear model is expected to hold. It is simple to implement and computationally tractable for use. Our theoretical results show that given a large enough sample, DDGroup recovers a region where a single linear model with low variance is well-specified (if one exists), and experiments on real-world medical datasets confirm that it can discover regions where a local linear model has improved performance. Our experiments also show that DDGroup can uncover subgroups with qualitatively different relationships which are missed by simply applying parametric approaches to the whole dataset.

Zachary Izzo
NEC Labs North America
zach@nec-labs.com

MS66

Randomized Mask Generation and Selection: From Unstructured to Structured Model Pruning

Recent progress in artificial general intelligence has led to large language models (LLMs) with billions of parameters. This scale necessitates the removal of unnecessary neurons or weights through model pruning. Traditional pruning methods typically focus on the magnitude of weights in a deterministic manner. However, using weight magnitude is a local metric without considering how it affects the model globally, and deterministic pruning can introduce errors that accumulate across layers. Conversely, randomized pruning can help even out these errors across different layers. In this talk, we introduce two inference-aware pruning criteria derived from the optimization perspective of output approximation, which surpass traditional training-aware metrics such as gradient and Hessian. Moreover, we introduce a two-step reconstruction technique to mitigate pruning errors without model retraining. Our experimental results showcase the superior performance of this approach across various datasets and models, markedly reducing both computational costs and hardware requirements.

Qi Lei
New York University
ql518@nyu.edu

MS67

Revisiting Hard Negative Sampling for Contrastive Learning: Optimal Representation Geometry and Neural Vs Dimensional Collapse

For a widely-studied data model and general loss and sample-hardening functions we prove that the Supervised Contrastive Learning (SCL), Hard-SCL (HSCL), and Unsupervised Contrastive Learning (UCL) risks are minimized by representations that exhibit Neural Collapse (NC), i.e., the class means form an Equiangular Tight Frame (ETF) and data from the same class are mapped to the same representation. We also prove that for any representation mapping, the HSCL and Hard-UCL (HUCL) risks are lower bounded by the corresponding SCL and UCL risks. Although the optimality of ETF is known for SCL, albeit only for InfoNCE loss, its optimality for HSCL and UCL under general loss and hardening functions is novel. Moreover, our proofs are much simpler, compact, and transparent. We empirically demonstrate, for the first time, that Adam optimization of HSCL and HUCL risks with random initialization and suitable hardness levels can indeed converge to the NC geometry if we incorporate unit-ball or unit-sphere feature normalization. Without incorporating hard negatives or feature normalization, however, the representations learned via Adam suffer from dimensional collapse (DC) and fail to attain the NC geometry.

Ruijie Jiang
Tufts University
ruijie.jiang@tufts.edu

Thuan Nguyen
Worcester Polytechnic Institute
tnguyen8@wpi.edu

Shuchin Aeron
Tufts University
shuchin@ece.tufts.edu

Prakash Ishwar
Boston University
pi@bu.edu

MS67

Recovering Simultaneously Structured Data Via Non-Convex Iteratively Reweighted Least Squares

We propose a new algorithm for the problem of recovering data that adheres to multiple, heterogeneous low-dimensional structures from linear observations. Focusing on data matrices that are simultaneously row-sparse and low-rank, we propose and analyze an iteratively reweighted least squares (IRLS) algorithm that is able to leverage both structures. In particular, it optimizes a combination of non-convex surrogates for row-sparsity and rank, a balancing of which is built into the algorithm. We prove locally quadratic convergence of the iterates to a simultaneously structured data matrix in a regime of minimal sample complexity (up to constants and a logarithmic factor), which is known to be impossible for a combination of

convex surrogates. In experiments, we show that the IRLS method exhibits favorable empirical convergence, identifying simultaneously row-sparse and low-rank matrices from fewer measurements than state-of-the-art methods.

Christian Kümmerle
Johns Hopkins University
kuemmerle@uncc.edu

Johannes Maly
University of Munich
maly@math.lmu.de

Christian Kümmerle
Johns Hopkins University
kuemmerle@uncc.edu

MS67

Randomized Iterative Solvers for Linear Systems with Fast Spectral Decay

Thesketch-and-projectis a unifying framework for many known randomized iterative methodsfor solving linear systems, such as randomized Kaczmarz and coordinate descent algorithms, their block variants as well as the extensions to non-linear optimization problems.In the linear case, convergence rates of the sketch-and-project methods are typically measured in terms of the condition number of the system, that, depending on the structure of the data, can be very small. However, a strong side of employing randomized iterative methods like the sketch-and-project, or randomized Kaczmarz, is in their adaptivity. In this talk, I will discuss how and why certain variants of these methods become efficient on the systems with fast, or partially known spectral decay. This phenomenon is rooted into more delicate ways to quantify the convergence of the sketch-and-project type methods via the so-called spectral tail condition number or projected condition number. Based on the works joint with J. Lok, M. Derezinski, D. Needell and D. LJeune

Elizaveta Rebrova
Princeton University
elre@princeton.edu

MS67

Benign and Tempered Overfitting for Least Squares Error in Variables Regression with Low Rank Data

Despite the importance of denoising in modern machine learning and ample empirical work on supervised denoising, its theoretical understanding is still relatively scarce. We study supervised denoising and noisy-input regression under distribution shift. We add three considerations to increase the applicability of our theoretical insights to real-life data and modern machine learning. First, while most past theoretical work assumes that the data covariance matrix is full-rank and well-conditioned, empirical studies have shown that real-life data is approximately low-rank. Thus, we assume that our data matrices are low-rank. Second, we drop independence assumptions on our data. Third, the rise in computational power and dimensionality of data have made it important to study non-classical regimes of learning. Thus, we work in the non-classical proportional regime, where data dimension d and number of samples N grow as $d/N = c + o(1)$. For this setting, we derive data-dependent, instance specific expressions for the test error for both denoising and noisy-input regression, and study when overfitting the noise is benign, tempered or catastrophic. We show that the test error exhibits double descent under general distribution shift, providing insights for data augmentation and the role of noise as an implicit regularizer.

Rishi Sonthalia
University of California Los Angeles
rsonthal@math.ucla.edu

Chinmaya Kausik
University of Michigan
ckausik@umich.edu

Kashvi Srivastava
University of Michigan Ann Arbor
kashvi@umich.edu

MS68

Physics Extraction Pods: Navigating the Landscape Between Learning and Knowing

Physics Extraction Pods (PEP) are sensors places within numerical or laboratory experiments with the ability to measure their surrounding over finite time intervals and to process the collected data. The sensors navigate in the dataset, modifying both their sensing and analysis modalities with the aim of substantial learning. PEP is equipped with sufficient algorithmic autonomy to achieve its goal. I will describe the main features of these PEP and demonstrate their application to problems of importance in science and engineering. I will explain the significance and distinction between sensing, learning, substantial learning.

Zheming Gou, Vinay Kumar, Xiaoshu Zeng, Philippe Hawi
University of Southern California
zgou@usc.edu, vinaykum@usc.edu, xiaoshuz@usc.edu, hawi@usc.edu

Roger Ghanem
University of Southern California
Civil and Environmental Engineering
ghanem@usc.edu

MS68

Towards a Notion of Model Correctness for Deep Learning

The remarkable success of deep learning (DL) models is attended by troubling features, including gaps in validity, resiliency, and explainability/interpretability that create obstacles to employing DL methods as parts of scientific models capable of being verified and validated. These gaps can be traced to the heuristic nature of DL modeling: no mathematical theory exists that can provide predictions of DL outputs, and without such predictions, there is no notion of *model correctness* for DL, and no opportunity for true verification, validation, and uncertainty quantification. I will describe a program for creating such mathematical theory based on a program that abstracts the models themselves as data distribution estimators coupled to Bayes-optimal decision machines, wherein the two are concomitantly trained. Such an abstraction encourages a statistical outlook that emphasizes the distributional nature of the data, the structure of the decision space, and the information flow between them. I will discuss the opportunities to create strong, validatable predictions of DL

models, to bound errors, to eliminate hallucinations, and to make model behavior effectively transparent rather than opaque.

Carlo Graziani
Argonne National Laboratory
cgraziani@anl.gov

MS68

Towards Rigorous Learning of Energy-based Models

In this talk, we discuss the perspectives for developing a rigorous learning framework for energy-based models (EBMs), including those based on deep neural network architectures. We make a case that development of a rigorous approach to EBMs can be achieved by learning from existing AI frameworks such as undirected graphical models and Markov random fields. These frameworks are widely used in science to represent joint probability distributions with an underlying conditional dependence structure. The inverse problem of learning a graphical model given independent samples from its joint distribution can be solved with near-optimal sample complexity using a convex optimization method known as Interaction Screening. But the computational cost of Interaction Screening becomes prohibitive when the energy function of the true graphical model has higher order terms. We introduce NeurISE, a neural-net based algorithm for EBM learning, to tackle this limitation while retaining some of the reconstruction guarantees. We use neural nets as function approximators in an Interaction Screening objective function. The optimization of this objective then produces a neural-net representation for the EBM conditionals. We also show how NeurISE can be used to learn the underlying structure of the true model with some simple modifications to its training procedure.

Abhijith Jayakumar, Andrey Lokhov, Sidhant Misra
Los Alamos National Laboratory
abhijithj@lanl.gov, lokhov@lanl.gov, sidhant@lanl.gov

Marc Vuffray
Theoretical Division and CNLS
Los Alamos National Laboratory
vuffray@lanl.gov

MS68

Learning from Data Via Measure Transport

Optimal measure transport, broadly construed, deals with the problem of minimizing the cost of transporting, via a transport map, a (probability) measure defined on a reference space to another defined on a target space. In the context of learning, methods of measure transport offer many advantages, since they allow for a unified framework for the analysis of data distributed according to a wide class of probability measures. This is one reason why measure transport techniques are being actively developed, including within deep learning (e.g., normalizing flows, diffusion maps). We will discuss our work concerning the use of measure transport for density estimation, focusing on cases when sample data for model training is limited in amount. More precisely, given a fixed, easy to sample probability measure on the reference space (e.g., Gaussian), a limited amount of samples drawn from an unknown target probability measure, and restricting the sought after transport maps to the class of triangular maps, we devise a randomized technique to statistically learn properties of the transport maps and a posteriori infer from them properties of the unknown target probability measure. This is done by learning series of triangular transport maps represented by finite dimensional approximations in an appropriate function space dictated by the choice of the reference space. Extensions of these techniques for the purpose of model calibration in the target space will be mentioned as well.

Vanessa Lopez-Marrero
Brookhaven National Laboratory
vlopezmar@bnl.gov

MS69

Sparse Bayesian Learning Techniques for Big Data Inverse Problem

Sparse Bayesian learning (SBL) is an advanced statistical modeling technique for inverse problems that builds upon traditional Bayesian methods by integrating hierarchical structures within prior distributions. This approach allows for extracting intricate relationships between parameters at various levels, fostering information sharing throughout the model. It is particularly effective when dealing with limited, noisy, or indirect data, yielding more accurate and robust inferences. Consequently, SBL has proven successful in diverse fields, such as machine learning, signal processing, and remote sensing. In this talk, we focus on the challenges of applying SBL to big data inverse problems, such as the need for prior conditioning and iterative solvers for large linear systems. In particular, we discuss recent techniques based on oblique projection approaches.

Jan Glaubitz
Massachusetts Institute of Technology
Glaubitz@mit.edu

Jonathan Lindbloom
Dartmouth College
jonathan.t.lindbloom.gr@dartmouth.edu

Anne Gelb
Dartmouth College
Department of Mathematics
annegelb@math.dartmouth.edu

Youssef M. Marzouk
Massachusetts Institute of Technology
ymarz@mit.edu

MS69

A Structurally Informed Data Assimilation Approach for Discontinuous State Variables

Ensemble-based Kalman filtering data assimilation is a scientific process that combines available observations with numerical simulations to obtain statistically accurate and reliable state representations in dynamical systems. However, it is well known that the commonly used Gaussian distribution assumption introduces biases for state variables that admit discontinuous profiles, which are prevalent in nonlinear partial differential equations. In this talk, we focus on the design of a new structurally informed prior that exploits statistical information from the simulated state variables. In particular, based on the second moment information of the state variable gradient, we construct a new weighting matrix for the numerical simulation contribution in the data assimilation objective function. This replaces the typical prior covariance matrix used for this purpose. We further adapt our weighting matrix to include

information in discontinuity regions via a clustering technique. Our numerical experiments demonstrate that this new approach yields more accurate estimates than those obtained using standard ensemble-based Kalman filtering on shallow water equations.

Tongtong Li
Dartmouth College
tongtong.li@umbc.edu

Yoonsang Lee, Anne Gelb
Dartmouth College
Department of Mathematics
yoonsang.lee@dartmouth.edu, annegelb@math.dartmouth.edu

MS70

Reduced Order Approximation of Kohn-Sham Density Functional Theory

Density Functional Theory (DFT) is a formulation of the electronics structure problem that offers a practical balance between computational complexity and accuracy. As the predictive tool of choice for materials science applications, DFT codes account for a large fraction of the utilization on many supercomputing systems. However, the high computational cost associated with the solution of the DFT equations poses a significant limitation for practical quantum molecular dynamics (MD) simulations, necessary to obtain meaningful inference for many realistic systems. To address this limitation, we investigate data-driven strategies which combine linear scaling algorithms for DFT and reduced order models (ROMs), constrained by the physics, to enable the exploration of intrinsic low dimensional subspace and manifold through data. Combining ROMs with DFT provides a natural way to incorporate physics through the underlying PDE with favorable implications for error control, yielding low computational cost, yet high fidelity solutions. In this talk, we will present the data-driven strategies and numerical results on prototype problems which exhibit the technical difficulties posed by Kohn-Sham Density Functional Theory.

Siu Wun Cheung
Lawrence Livermore National Laboratory
cheung26@llnl.gov

MS70

Data-Driven Methods for the Allen-Cahn Equation

The Allen-Cahn equation describes the process of phase separation and transition in phase field modeling of multi-component physical systems. High-fidelity numerical simulations of the Allen-Cahn equation can be computationally expensive due to its stiffness, compounded by the necessity for multiple simulations in parametric Allen-Cahn scenarios within a multi-query setting. In this presentation, we will introduce two approaches to accelerate the numerical simulations of the Allen-Cahn equation. One approach is based on model order reduction, wherein we develop a structure-preserving, gradient-preserving operator inference technique for learning the reduced operators. The other approach is deep learning-based, involving the design of specialized convolutional neural network models to learn fully-discrete operators. We will present numerical experiments demonstrating the efficiency of the proposed methods.

Yuwei Geng
University of South Carolina
ygeng@email.sc.edu

Lili Ju
University of South Carolina
Department of Mathematics
ju@math.sc.edu

Zhu Wang
Department of Mathematics
University of South Carolina
wangzhu@math.sc.edu

MS70

Structure-Preserving Moment Closure for Kinetic Equations

In this talk, we present our work on structure-preserving machine learning (ML) moment closure models for kinetic equations. Most of the existing ML closure models are not able to guarantee the stability, which directly causes blow up in the long-time simulations. In our work, with carefully designed neural network architectures, the ML closure model can guarantee the provable stability (or hyperbolicity). Moreover, other mathematical properties, such as physical characteristic speeds, are also discussed. Extensive benchmark tests show the good accuracy, long-time stability, and good generalizability of our ML closure model.

Juntao Huang
Texas Tech Univeresity
juntao.huang@ttu.edu

MS70

Data-Assisted Algorithms for Inverse Random Source Problems

Inverse source scattering problems are essential in various fields, including antenna synthesis, medical imaging, and earthquake monitoring. In many applications, it is necessary to consider uncertainties in the model, and such problems are known as stochastic inverse problems. Traditional methods require a large number of realizations and information on medium coefficients to achieve accurate reconstruction for inverse random source problems. To address this issue, we propose a data-assisted approach that uses boundary measurement data to reconstruct the statistical properties of the random source with fewer realizations. We compare the performance of different data-driven algorithms under this framework to enhance the initial approximation obtained from integral equations. Our numerical experiments demonstrate that the data-assisted approach achieves better reconstruction with only 1/10 of the realizations required by traditional methods.Among the various Image-to-Image translation algorithms that we tested, the pix2pix method outperforms others in reconstructing well-separated inclusions with accurate positions. Our proposed approach results in stable reconstruction with respect to the observation data noise.

Ying Liang
Duke University
ying.liang@duke.edu

Peijun Li
Purdue University
lipeijun@purdue.edu

Yuliang Wang
Beijing Normal University
BNU-HKBU United International College
yuliangwang@uic.edu.cn

MS71

Bayesian Autoencoders for Data-Driven Discovery of Coordinates, Governing Equations and Fundamental Constants

Recent progress in autoencoder-based sparse identification of nonlinear dynamics (SINDy) under ℓ_1 constraints allows joint discoveries of governing equations and latent coordinate systems from spatio-temporal data, including simulated video frames. However, it is challenging for ℓ_1-based sparse inference to perform correct identification for real data due to the noisy measurements and limited sample sizes. To address the data-driven discovery of physics in the low-data and high-noise regimes, we propose Bayesian SINDy autoencoders, which incorporate a hierarchical Bayesian Spike-and-slab Gaussian Lasso prior. Bayesian SINDy autoencoder enables the joint discovery of governing equations and coordinate systems with uncertainty estimate. To resolve the challenging computational tractability of the Bayesian hierarchical setting, we adapt an adaptive empirical Bayesian method with Stochastic Gradient Langevin Dynamics (SGLD) which gives a computationally tractable way of Bayesian posterior sampling within our framework. Bayesian SINDy autoencoder achieves better physics discovery with lower data and fewer training epochs, along with valid uncertainty quantification suggested by the experimental studies. The Bayesian SINDy autoencoder can be applied to real video data, with accurate physics discovery which correctly identifies the governing equation and provides a close estimate for standard physics constants like gravity g, for example, in videos of a pendulum.

Mars L. Gao
Universty of Washington
Paul G. Allen School of Computer Science & Engineering
marsgao@uw.edu

J. Nathan Kutz
University of Washington, Seattle
Dept of Applied Mathematics
kutz@uw.edu

MS71

Numerical Strategies for Model Selection of Differential Algebraic Equations

Differential Algebraic Systems (DAEs) form a broad category of differential equations wherein ordinary differential equations (ODEs) are combined with algebraic equations, resulting in a highly coupled system. DAEs typically arise due to the separation of time scales, conservation laws, or algebraic constraints in the dynamical systems. Sparse optimization represents a recently expanding set of techniques used to discover the model of a dynamical system from a model library and data. Despite the extensive study of numerical solvers for DAEs in the literature and their wide range of applications, model discovery algorithms for such systems remain a less explored area of research. Previous approaches focused on using quasi-steady state approximations to reduce the system of DAEs into an implicit system of ODEs, and then trying to discover the reduced system. However, the reduction often introduces numerical instability, and existing algorithms based on Sparse Inference of Nonlinear Dynamics (SINDy) have highly correlated library terms due to algebraic relationships between candidate features in the library. We propose a novel method called Sparse Optimization for Differential Algebraic systems (SODAs), that can identify the full system of DAEs without any reduction. Additionally, we demonstrate that this method can handle higher levels of noise in the data and work with simpler candidate library functions compared to the existing methods to discover implicit ODE systems.

Manu Jayadharan, Finn Hagerty
Northwestern University
manu.jayadharan@northwestern.edu,
finnhagerty2025@u.northwestern.edu

Arthur N. Montanari
University of Luxemburg
montanariarthur@gmail.com

Christina Catlett
Northwestern University
ccatlett@u.northwestern.edu

Niall M. Mangan
Engineering Sciences and Applied Mathematics
Northwestern University
niall.mangan@northwestern.edu

MS71

Estimating the Parameters of Multicompartment Hodgkin-Huxley Models from Extracellular Voltage Recordings

Biophysical models such as the Hodgkin-Huxley model characterize electrical signal propagation across compartments of a neuron. To estimate the parameters of these models, existing methods typically rely on measurements of voltage from a single compartment (the soma). Recent advances in neural recording technology with high-density Neuropixel Ultra probes enable dense sampling of extracellular voltage from many sites surrounding a neuron, enabling indirect measurement of many compartments of a cell. Here, we propose to use these extracellular recordings to fit Hodgkin-Huxley models. We extend biophysical models to account for extracellular measurements, and we use Sequential Monte Carlo (SMC) and Extended Kalman Smoothing (EKS) to infer membrane voltage from observed extracellular action potentials. To account for the fact that multiple parameters might produce a similar extracellular signal, we propose using MCMC combined with SMC or EKS to approximate a posterior distribution over voltage traces and parameters. Our JAX implementation enables parallelization and hardware acceleration of these complex models. We demonstrate the performance of these approaches for multi-compartment neurons using real and simulated data.

Ian Tanoh, Dieterich Lawson, Scott Linderman
Stanford University
ictanoh@stanford.edu, jdlawson@stanford.edu,
scott.linderman@stanford.edu

MS71

Hierarchical Learning of Partially Observed Interrelated Dynamical Systems

We present a novel hierarchical Bayesian inference toolbox that leverages differentiable data-assimilation to learn a

set of interconnected, partially observed stochastic dynamical systems. Contrary to the common assumption that each observed data trajectory represents an independent realization of a different (unknown) dynamical system, we study the case where each observed trajectory is unique, yet related to a set of observed, akin trajectories. This set-up underpins many real-world time-series data (e.g., in biomedicine and engineering), as it accommodates the idiosyncrasies of each individual time-series, yet incorporates the similarities across a group of interrelated time-series. Namely, each dynamical system trajectory is idiosyncratic (the individual), yet shares commonalities with other observed trajectories (the population). We merge hierarchical Bayesian modeling with differentiable data-assimilation to efficiently tackle the challenges of filtering, smoothing, predicting, and system identification in these scenarios. We show how the proposed modeling framework and the presented toolbox enables efficient and accurate Bayesian inference of interrelated, linear and/or nonlinear continuous-discrete, partially observed dynamical systems.

Iñigo Urteaga
Basque Center for Applied Math
Bilbao, Spain
iurteaga@bcamath.org

Matthew Levine
Broad Institute
mattlevine22@gmail.com

MS72

Why Does the Two-Timescale Q-Learning Converge to Different Mean Field Solutions? A Unified Convergence Analysis

This talk will revisit the unified two-timescale Q-learning algorithm as initially introduced by Angiuli et al. This algorithm demonstrates efficacy in solving mean field game (MFG) and mean field control (MFC) problems, simply by tuning the ratio of two learning rates for mean field distribution and the Q-functions respectively. In this talk, we will provide a comprehensive theoretical explanation of the algorithms bifurcated numerical outcomes under fixed learning rates. We achieve this by establishing a diagram that correlates continuous-time mean field problems to their discrete-time Q-function counterparts, forming the basis of the algorithm. Our key contribution lies in the construction of a Lyapunov function integrating both mean field distribution and Q-function iterates. This Lyapunov function facilitates a unified convergence of the algorithm across the entire spectrum of learning rates, thus providing a cohesive framework for analysis.

Jing An
Department of Mathematics
Duke University
jing.an@duke.edu

Jianfeng Lu
Mathematics Department
Duke University
jianfeng@math.duke.edu

Yue Wu
Department of Mathematics
The Hong Kong University of Science and Technology
ywudb@connect.ust.hk

Yang Xiang
Hong Kong University of Science and Technology
maxiang@ust.hk

MS72

Kernel Methods for Dynamic Generative Modeling: Theory and Algorithms

This talk will present a general framework for the transport of probability measures towards minimum divergence generative modeling and sampling using ordinary differential equations (ODEs) and Reproducing Kernel Hilbert Spaces (RKHSs), inspired by ideas from diffeomorphic matching and image registration. A theoretical analysis of the proposed method is presented, giving a priori error bounds in terms of the complexity of the model, the number of samples in the training set, and model misspecification. An extensive suite of numerical experiments further highlights the properties, strengths, and weaknesses of the method and extends its applicability to other tasks, such as conditional simulation and inference.

Biraj Pandey
University of Washington in Seattle
bpandey@uw.edu

Bamdad Hosseini
University of Washington
bamdadh@uw.edu

Pau Batlle
Caltech
pbatllef@caltech.edu

Houman Owhadi
California Institute of Technology
owhadi@caltech.edu

MS72

Statistical Spatially Inhomogeneous Diffusion Inference

Inferring a diffusion equation from discretely observed measurements is a statistical challenge of significant importance in a variety of fields, from single-molecule tracking in biophysical systems to modeling financial instruments. Assuming that the underlying dynamical process obeys a d-dimensional stochastic differential equation of the form $dx_t = b(x_t)dt + \Sigma(x_t)dw_t$, we propose neural network-based estimators of both the drift b and the spatially-inhomogeneous diffusion tensor $D = \Sigma\Sigma^T/2$ and provide statistical convergence guarantees when b and D are s-Hölder continuous. Notably, our bound aligns with the minimax optimal rate $N^{-\frac{2s}{2s+d}}$ for nonparametric function estimation even in the presence of correlation within observational data, which necessitates careful handling when establishing fast-rate generalization bounds. Our theoretical results are bolstered by numerical experiments demonstrating accurate inference of spatially-inhomogeneous diffusion tensors.

Yinuo Ren
Stanford University
yinuoren@stanford.edu

Yiping Lu
Courant Institute, New York University
Northwestern University
yiping.lu@northwestern.edu

Grant Rotskoff
Stanford University
rotskoff@stanford.edu

Lexing Ying
Stanford University
Department of Mathematics
lexing@stanford.edu

MS72

Statistical Optimality of ODE-Based Generative Modelling

Ordinary differential equations (ODEs), via their induced flow maps, provide a powerful framework to parameterize invertible transformations for the purpose of representing complex probability distributions. While such models have achieved enormous success in machine learning, particularly for generative modeling and density estimation, little is known about their statistical properties. In this talk, we will present our recent series of work, which establish the first minimax-optimal nonparametric statistical convergence rate for ODE-based generative modeling trained through maximum likelihood estimation. We begin by proving a convergence theorem applicable to an arbitrary class of velocity field satisfying certain simple boundary constraints. This general result captures the trade-off between approximation error ('bias') and the statistical complexity of the ODE model ('variance'). The statistical complexity is then quantified via the metric entropy of the velocity field class. We further apply this general framework to the setting of learning densities with prescribed smoothness and establish minimax-optimal convergence rates for the class of sparse neural networks. In terms of practical impact, our work reveals that neural ODE is a statistically optimal estimator for generative modeling.

Haoxuan Chen
Institute for Computational and Mathematical Engineering
Stanford University
haoxuanc@stanford.edu

Zhi Ren
Massachusetts Institute of Technology
zren@mit.edu

Sven Wang
Department of Mathematics
Humboldt University Berlin
sven.wang@hu-berlin.de

Youssef M. Marzouk
Massachusetts Institute of Technology
ymarz@mit.edu

Jose H. Blanchet
Stanford University
jose.blanchet@stanford.edu

MS73

Data-Driven Aerodynamic Shape Design with Distributionally Robust Optimization Approaches

We formulate and solve data-driven aerodynamic shape design problems with distributionally robust optimization (DRO) approaches. DRO aims to minimize the worst-case expected performance in a set of distributions that is informed by observed data with uncertainties. Building on the findings of the work [J.-y. Gotoh, M. J. Kim, A. E. Lim, Robust empirical optimization is almost the same as meanvariance optimization], we study the connections between a class of DRO and robust design optimization, which is classically based on the mean-variance (standard deviation) optimization formulation pioneered by Taguchi. Our results provide a new perspective to the understanding and formulation of robust design optimization problems. It enables data-driven and statistically principled approaches to quantify the trade-offs between robustness and performance, in contrast to the classical robust design formulation that captures uncertainty only qualitatively.

Long Chen
RPTU Kaiserslautern, Germany
long.chen@scicomp.uni-kl.de

Jan Rottmayer
RPTU Kaiserslautern
jan.rottmayer@scicomp.uni-kl.de

Lisa Kusch
TU Eindhoven
l.kusch@tue.nl

Nicolas R. Gauger
TU Kaiserslautern
nicolas.gauger@scicomp.uni-kl.de

Yinyu Ye
Stanford University
yinyu-ye@stanford.edu

MS73

Shape Optimization Via Pinns and Stochastic Interpretation of PDE

We introduce a novel mesh-free and direct method for computing the shape derivative in PDE-constrained shape optimization problems. Our approach is based on a probabilistic representation of the shape derivative and is applicable for second-order elliptic PDEs with Dirichlet boundary conditions and a general class of target functions. The proba- bilistic representation derives from a boundary sensitivity result for diffusion processes due to Constantini, Gobet and El Karoui. We provide a Taylor test to verify the accuracy of our methodology.

Maximilian Wuerschmidt
Trier University, Germany
wuerschmidt@uni-trier.de

Volker H. Schulz
University of Trier
Department of Mathematics
Volker.Schulz@uni-trier.de

Frank Seifried
University of Trier
seifried@uni-trier.de

MS73

AONN-2: An Adjoint-oriented Neural Network Method for PDE-constrained Shape Optimization

Shape optimization has been playing an important role in a large variety of engineering applications. Existing shape

optimization methods are generally mesh-dependent and therefore encounter challenges due to mesh deformation. To overcome this limitation, we present a new adjoint-oriented neural network method, AONN-2, for PDE-constrained shape optimization problems. This method extends the capabilities of the original AONN method, which is developed for efficiently solving parametric optimal control problems. AONN-2 inherits the direct-adjoint looping (DAL) framework for computing the extremum of an objective functional and the neural network methods for solving complicated PDEs from AONN. Furthermore, AONN-2 expands the application scope to shape optimization by taking advantage of the shape derivatives to optimize the shape represented by discrete boundary points. AONN-2 is a fully mesh-free shape optimization approach, naturally sidestepping issues related to mesh deformation, with no needs for maintaining mesh quality and additional mesh corrections. A series of experimental results are presented, highlighting the flexibility, robustness, and accuracy of AONN-2.

Xili Wang
Peking University, Beijing, China
xiliwang@stu.pku.edu.cn

Pengfei Yin, Bo Zhang, Chao Yang
Peking University
Beijing, China
pengfeiyin@pku.edu.cn, zhang_bo@pku.edu.cn, yang_chao@pku.edu.cn

MS75

Neighbor Embeddings with Contrastive Learning and Pykeops

Neighbor embedding methods like t-SNE and UMAP are popular tools for dimensionality reduction and visualization. These algorithms are typically perceived by practitioners as very different, with unrelated loss functions, optimization strategies, and mathematical underpinnings. Here we show that in fact they are conceptually very similar and fall on a single attraction-repulsion spectrum, obtained by varying attraction strength between the neighboring points. Different embeddings on the spectrum can be obtained by varying exaggeration strength in t-SNE, or alternatively by varying the normalization constant in a contrastive loss called negative sampling. We show that UMAP is effectively negative sampling applied to the t-SNE loss function, and explain the difference between negative sampling and noise-contrastive estimation, which has been used to optimize t-SNE under the name NCVis.
The spectrum is characterized by the trade-off between preservation of discrete cluster structure and preservation of continuous manifold structure. Increasing neighbor attraction moves the embedding from t-SNE to UMAP, bringing out continuous structures (e.g. global trajectories) at the expense of losing local details (e.g. fine cluster structure).
As an alternative to contrastive losses or approximation schemes, like Barnes-Hut, neighbor-embedding methods can also be quickly optimized via the GPU acceleration provided by the pykeops package.

Sebastian Damrich, Jan Niklas Böhm
Hertie Institute for AI in Brain Health
University of Tuebingen
sebastian.damrich@uni-tuebingen.de,
jan-niklas.boehm@uni-tuebingen.de

Philipp Berens
Hertie Institute for AI in Brain Health University of Tuebin
philipp.berens@uni-tuebingen.de

Fred A. Hamprecht
IWR, Heidelberg University
fred.hamprech@iwr.uni-heidelberg.de

Dmitry Kobak
Hertie Institute for AI in Brain Health University of Tuebin
dmitry.kobak@uni-tuebingen.de

MS75

Euclidean Mirrors and Dynamics in Network Time Series

Analyzing changes in network evolution is central to statistical network inference, as underscored by recent challenges of predicting and distinguishing pandemic-induced transformations in organizational and communication networks. We consider a joint network model in which each node has an associated time-varying low-dimensional latent vector of feature data, and connection probabilities are functions of these vectors. Under mild assumptions, the time-varying evolution of the latent vectors exhibits low-dimensional manifold structure under a suitable notion of distance. This distance can be approximated by a measure of separation between the observed networks themselves, and there exist Euclidean representations for underlying network structure, as characterized by this distance, at any given time. These Euclidean representations, called Euclidean mirrors, permit the visualization of network evolution and transform network inference questions such as change-point and anomaly detection into a classical setting. We illustrate our methodology with real and synthetic data, and identify change points corresponding to massive shifts in pandemic policies in a communication network of a large organization.

Zachary Lubberts
University of Virginia
zlubberts@virginia.edu

Avanti Athreya
Johns Hopkins University
Department of Applied Mathematics and Statistics
dathrey1@jhu.edu

Youngser Park, Carey Priebe
Johns Hopkins University
youngser@jhu.edu, cep@jhu.edu

MS75

Learning Low-Dimensional Nonlinear Structures from High-Dimensional Noisy Data: An Integral Operator Approach

We propose a kernel-spectral embedding algorithm for learning low-dimensional nonlinear structures from noisy and high-dimensional observations, where the data sets are assumed to be sampled from a nonlinear manifold model and corrupted by high-dimensional noise. The algorithm employs an adaptive bandwidth selection procedure which does not rely on prior knowledge of the underlying manifold. The obtained low-dimensional embeddings can be further utilized for downstream purposes such as data visualization, clustering and prediction. Our method is the-

oretically justified and practically interpretable. Specifically, for a general class of kernel functions, we establish the convergence of the final embeddings to their noiseless counterparts when the dimension grows polynomially with the size, and characterize the effect of the signal-to-noise ratio on the rate of convergence and phase transition. We also prove the convergence of the embeddings to the eigenfunctions of an integral operator defined by the kernel map of some reproducing kernel Hilbert space capturing the underlying nonlinear structures. Our results hold even when the dimension of the manifold grows with the sample size. Numerical simulations and analysis of real data sets show the superior empirical performance of the proposed method, compared to many existing methods, on learning various nonlinear manifolds in diverse applications.

Rong Ma
Harvard University
rongma@hsph.harvard.edu

MS75

Contrastive Independent Component Analysis

Visualizing data and finding patterns in data are ubiquitous problems in the sciences. Increasingly, applications seek signal and structure in a contrastive setting: a foreground dataset relative to a background dataset. For this purpose, we propose contrastive independent component analysis (cICA). This generalizes independent component analysis to independent latent variables across a foreground and background. We propose a hierarchical tensor decomposition algorithm for cICA. We study the identifiability of cICA and demonstrate its performance visualizing data and finding patterns in data, using synthetic and real-world datasets, comparing the approach to existing contrastive methods.

Kexin Wang
Harvard University
kexin_wang@g.harvard.edu

Aida Maraj
Max Planck Institute for Mathematics in the Sciences
maraja@umich.edu

Anna Seigal
Harvard University, U.S.
aseigal@seas.harvard.edu

MS76

Towards Understanding the Dynamics of Gaussian-Stein Variational Gradient Descent

Stein Variational Gradient Descent (SVGD) is a nonparametric particle-based deterministic sampling algorithm. Despite its wide usage, understanding the theoretical properties of SVGD has remained a challenging problem. For sampling from a Gaussian target, the SVGD dynamics with a bilinear kernel will remain Gaussian as long as the initializer is Gaussian. Inspired by this fact, we undertake a detailed theoretical study of the Gaussian-SVGD, or equivalently, Gaussian variational inference (GVI) with SVGD. We present a complete picture by considering both the mean-field PDE and discrete particle systems. When the target is strongly log-concave, the mean-field Gaussian-SVGD dynamics is proven to converge linearly to the Gaussian distribution closest to the target in KL divergence. In the finite-particle setting, there is both uniform in time convergence to the mean-field limit and linear convergence in time to the equilibrium if the target is Gaussian. In the general case, we propose a density-based and a particle-based implementation of the Gaussian-SVGD, and show that several recent algorithms for GVI, proposed from different perspectives, emerge as special cases of our unified framework. Interestingly, one of the new particle-based instance from this framework empirically outperforms existing approaches. Our results make concrete contributions towards obtaining a deeper understanding of both SVGD and GVI.

Krishnakumar Balasubramanian
University of California, Davis, U.S.
kbala@ucdavis.edu

MS76

On Independent Samples Along the Langevin Diffusion and Algorithm

Markov chains are widely used to sample from probability distributions, and their mixing time tells us how long we need to wait before the law of the iterates is close to the stationary distribution. In this talk, we discuss the "independence time" of a Markov chain, which tells us how long we need to wait before the iterates have small mutual information with the initial random variable. We study the independence time for the Langevin diffusion along with its discrete time implementations, and show that for strongly-log concave targets, the mutual information goes to 0 exponentially fast. These convergence rates are tight and our results are proven using strong data processing inequalities and regularity properties of these Markov chains. Based on joint work with Jiaming Liang and Andre Wibisono.

Siddharth Mitra
Yale University
siddharth.mitra@yale.edu

MS76

Improved Dimension Dependence of a Proximal Algorithm for Sampling

In this proposal, I will introduce a novel sampling algorithm based on the proximal sampler framework introduced by Lee et al. (2021). This new approximate proximal sampler achieves superior complexity bounds across almost all the settings, such as strong log-concavity, log-concavity, functional inequalities, as well as those involving semi-smooth or composite potentials. The core of our algorithm is an inexact realization of the restricted Gaussian oracle (RGO) based on approximate rejection sampling. The state-of-the-art complexity bounds are largely established by a new concentration inequality for semi-smooth functions over Gaussian distributions. For strongly log-concave distributions, our method has complexity bound $\tilde{\mathcal{O}}(\kappa d^{1/2})$ without warm start, better than the minimax bound for MALA.

Bo Yuan
Georgia Institute of Technology
byuan48@gatech.edu

MS76

Sampling from Interactive Energy Minimizers

The problem of sampling from distributions which minimize certain energy functionals is fundamental to sampling and probabilistic inference. This talks treats the case where

the energy functional contains an "interaction". Such functionals arise in many different statistical applications, from variational inference to deep learning. Our main insight is to decouple the two key aspects of this problem: (1) approximation of the mean-field SDE via a finite-particle system, via uniform-in-time propagation of chaos, and (2) sampling from the finite-particle stationary distribution, via standard log-concave samplers. Our approach is conceptually simpler and its flexibility allows for incorporating the state-of-the-art for both algorithms and theory.

Matthew Zhang
University of Toronto
matthew.zhang@mail.utoronto.ca

Mufan Li
Princeton University
mufan.li@princeton.edu

Sinho Chewi
Institute for Advanced Study
schewi@ias.edu

Murat Erdogdu
University of Toronto
erdogdu@cs.toronto.edu

Yunbum Kook
KAIST
Discrete Mathematics Group, Institute for Basic Science
yb.kook@gatech.edu

MS77

On Representing (Mixed-Integer) Linear Programs by Graph Neural Networks

While Mixed-integer linear programming (MILP) is NP-hard in general, practical MILP has received roughly 100-fold speedup in the past twenty years. Still, many classes of MILPs quickly become unsolvable as their sizes increase, motivating researchers to seek new acceleration techniques for MILPs. With deep learning, they have obtained strong empirical results, and many results were obtained by applying graph neural networks (GNNs) to making decisions in various stages of MILP solution processes. We study the theoretical foundation and discover a fundamental limitation: there exist feasible and infeasible MILPs that all GNNs will, however, treat equally, indicating GNN's lacking power to express general MILPs. Then we show that linear programs (LPs) without integer constraints do not suffer from this limitation and that, by restricting the MILPs to unfoldable ones or by adding random features, there exist GNNs that can reliably predict MILP feasibility, optimal objective values, and optimal solutions up to prescribed precision. We also show that second-order GNNs can represent branching strategies for MILPs and conduct small-scale numerical experiments to validate our theoretical findings.

Ziang Chen
MIT
ziang@mit.edu

MS77

Reconstructing Graph Diffusion History from a Single Snapshot

Diffusion on graphs is ubiquitous with numerous high-impact applications, ranging from the study of residential segregation in socioeconomics and activation cascading in neuroscience, to the modeling of disease contagion in epidemiology and malware spreading in cybersecurity. In these applications, complete diffusion histories play an essential role in terms of identifying dynamical patterns, reflecting on precaution actions, and forecasting intervention effects. Despite their importance, complete diffusion histories are rarely available and are highly challenging to reconstruct due to ill-posedness, explosive search space, and scarcity of training data. To date, only very few methods exist for diffusion history reconstruction. They are exclusively based on the maximum likelihood estimation (MLE) formulation and require knowing exact values of diffusion parameters. In this work, we study an even harder problem, namely reconstructing Diffusion history from A single SnapsHot (DASH), where we seek to reconstruct the history from only the final snapshot without knowing true diffusion parameters. We start with theoretical analyses that reveal a fundamental limitation of the MLE formulation. We prove: (a) the estimation error of diffusion parameters is unavoidable due to the NP-hardness of diffusion parameter estimation, and (b) the MLE formulation is sensitive to the estimation error of diffusion parameters. To overcome the inherent limitation of the MLE formulation, we propose a novel barycenter formulation: finding the barycenter of the posterior distribution of histories, which is provably stable against the estimation error of diffusion parameters. We further develop an effective solver named DIffusion hiTting Times with Optimal proposal (DITTO) by reducing the problem to estimating posterior expected hitting times via the MetropolisHastings Markov chain Monte Carlo method (MH MCMC) and employing an unsupervised graph neural network to learn an optimal proposal to accelerate the convergence of MH MCMC. We conduct extensive experiments to demonstrate the efficacy of the proposed method.

Ruizhong Qui
University of Illinois at Urbana-Champaign
rq5@illinois.edu

MS77

Regularizing Inverse Problems on Graphs with Score-based Priors

Consider the problem of inferring the topology of a graph from some observation model while having access to other graphs from the same domain. For example, we might want to infer a brain graph of a specific patient while having access to many other human brain graphs. This raises the question: How can we use known graphs (of potentially different sizes) to regularize the inference of the new graph? In this context, we set up a Bayesian framework to solve inverse problems on graphs by sampling from the posterior distribution over graphs, given the likelihood of observations and some prior over the graph structure. Our main novelty lies in using: i) Annealed Langevin diffusion as an MCMC sampler in the space of graphs, and ii) a graph-neural network to learn the score of the prior from examples of graphs. We analyze the use of score-based priors on several inverse problems on graphs, including network topology inference from diffused graph signals, network state-space estimation, and Gaussian graphical models.

Santiago Segarra
Rice University

segarra@rice.edu

MS77

Learning on Graphs: What is Next?

Graphs provide a universal representation of data with numerous types. In the last decades, we have witnessed techniques to extract knowledge from graphs for various real-world applications. In this talk, I will first talk about how these techniques evolved, what principles they shared and what unique advances they introduced. Now we are in the era of foundation models. Then I will discuss what progress we have made and what next steps we will potentially take towards graph foundation models.

Jiliang Tang
Michigan State University
tangjili@msu.edu

MS78

Searching for Unicorns

In this talk, we explore the transformative potential of off-the-shelf reinforcement learning (RL) algorithms in accelerating solutions to complex, research-level mathematical challenges. We begin by illustrating how these algorithms have achieved a 10X improvement in areas where previous advances of the same magnitude required many decades. A comparative analysis of different network architectures is presented to highlight their performance in this context. We then delve into the application of RL algorithms to exceptionally demanding tasks, such as those posed by the Millennium Prize problems and the smooth Poincar conjecture in four dimensions. Drawing on our experiences, we discuss the prerequisites for developing new RL algorithms and architectures that are tailored to these high-level challenges.

Sergei Gukov
California Institute of Technology
sergei.gukov@gmail.com

MS78

The Effect of Smooth Parametrizations on Nonconvex Optimization Landscapes

Given a constrained optimization problem, we often tackle it by choosing a parameterization of the constraint set and then optimizing over the parameters (which may be unconstrained). For example, if we need to optimize a real-valued cost over bounded-rank matrices, then we can parameterize the domain using a low-rank factorization (e.g., the SVD) and then optimize over the factors. Alternatively, in deep learning when optimizing over the function space represented by a neural network, we have parameterized the space by the weights and biases in the network. In such situations, a natural question is: does the choice of parameterization affect the nonconvex optimization landscape? And: are some parameterizations better than others? In this talk, I'll present a geometric framework to formalize these questions and analysis tools to help answer them. The theory will be applied to several examples, including the aforementioned ones as well as optimization problems in tensor decomposition and semidefinite programming. Joint work with Eitan Levin (Caltech) and Nicolas Boumal (EPFL).

Joe Kileel
University of Texas at Austin, U.S.
jkileel@math.utexas.edu

MS78

Linear Causal Disentanglement Via Higher-order Cumulants

Linear causal disentanglement is a recent method in causal representation learning to describe a collection of observed variables via latent variables with causal dependencies between them. We study its identifiability, assuming access to data under multiple contexts, each given by an intervention on a latent variable. We show that one perfect intervention on each latent variable is sufficient and in the worst case necessary to recover parameters under perfect interventions, generalizing previous work to allow more latent than observed variables. We give a constructive proof that computes parameters via a coupled tensor decomposition. For soft interventions, we find the latent graphs consistent with observed data, via the study of a system of polynomial equations. Based on https://arxiv.org/abs/2407.04605, joint work with Paula Leyes Carreno and Chiara Meroni.

Paula Leyes Carreno
Apple
pleyescarreno@college.harvard.edu

Chiara Meroni
ETH
chiara.meroni@eth-its.ethz.ch

Anna Seigal
Harvard University, U.S.
aseigal@seas.harvard.edu

MS79

Statistical Modeling of Topological Features in Medical Imaging: Enhancing Prognostic Precision and Interpretation

Tumor shape significantly influences growth and metastasis. We introduce a topological feature obtained by persistent homology to characterize tumor progression in brain tumor patients using radiology images, focusing on its influence on time-to-event data. These topological features, invariant to scale-preserving transformations, capture diverse tumor shape patterns. We introduce a functional spatial Cox proportional hazards model that represents these topological features in a functional space, utilizing them as functional predictors alongside their spatial locations. This model allows for interpretable analysis of the relationship between topological shape features and survival risks.

Yuhyeong Jang
Southern Methodist University
yuhyeongj@smu.edu¿

Tu Dan
The University of Texas Southwestern Medical Center
tu.dan@utsouthwestern.edu

Chul Moon
Southern Methodist University
Dallas, TX, USA
chulm@smu.edu

MS79

Usage of Topological Clustering Provides New In-

sights into Genomic Data Analysis

Mapper, a topological clustering algorithm, is frequently used as an exploratory tool to build a graphical representation of data. This representation can help us better understand the intrinsic shape of high-dimensional genomic data and retain information that may be lost using standard clustering algorithms. Although Mapper shows promise in analyzing high-dimensional data, tools to statistically analyze Mapper graphical structures are limited in the existing literature. In this talk, we will present two novel pipelines integrating Mapper with existing methods for molecular analysis and graph theory that provide empirical settings for statistical inferences. Firstly, we discuss a framework to process and analyze RNA-seq data combining Mapper, differential gene expression, and spectral shape analysis. To be precise, a scoring method using heat kernel signatures of Mapper networks. Secondly, we describe the utility of Mapper for detecting trait-associated genes from brain gene expression data. In particular, the graphical structures obtained by Mapper are further analyzed using Dijkstras algorithm to identify the best group of subjects for sparse modeling. Sparse modeling is then implemented to detect genes that are significant predictors of the trait of interest. The results show that the Mapper graphs have utilities beyond exploratory and can provide new insights into genomics data analysis.

Farzana Nasrin
University of Hawaii
fnasrin@hawaii.edu

Motoko Iwashita
University of
motokoi@hawaii.edu

Erik Morataya
University of Missouri
Columbia, MO, USA
eah4d@missouri.edu

Kathleen Storey
Lafayette College
storeyk@lafayette.edu

Masato Yoshizawa
University of Hawaii
yoshizaw@hawaii.edu

MS79

Manifold-Aware Model Reduction in Recurrent Neural Networks

A novel sparsification technique relying on manifold-aware regularization is proposed to reduce computational complexity during model learning in recurrent neural networks, while not sacrificing performance. By incorporating topological information via moduli regularization in the learning cost used to train the network, the proposed method is able to identify critical network weights while eliminating a significant portion of nuisance parameters. Moduli regularization allows to induce a geometric relationship between neurons in the hidden state of a recurrent network that further enables to identify important hidden state weights via an intelligently scaled norm-one penalization term. This effective norm-one based regularization allows for effective optimization of the learning cost in a computationally efficient manner. The effectiveness of our topology-aware learning scheme is verified on RNNs for navigation, natural language processing, and the adding problem with effective model sparsification of up to 90% of learning weights with negligible loss of accuracy.

Ioannis Schizas
Army Research Lab
Aberdeen Proving Ground, MD, USA
ioannis.d.schizas.civ@army.mil

MS79

Social and Environmental Drivers of Opioid Risk: A Topological Study

It is clear that the number of opioid prescriptions which inundated the United States over the past few decades has fueled the opioid epidemic. Less clear is the role that social and environmental factors have played in allowing the epidemic to spread. These factors are confounded by wide regional disparities in opioid overdoses, suggesting that the primary drivers of overdose risk at a national level may extend beyond a single set of societal and environmental factors and are a combination of regional factors. In this study, we study the role of social and environmental factors in the opioid overdose landscape through the lens of topology. We model the relationship between opioid overdoses and a subset of social and environmental factors as measured by the social vulnerability index (SVI) in regions exhibiting a diverse collection of theses factors. Considering that the geographical distance between counties may not reflect their social vulnerability distance, we construct topological adjacencies in the SVI space, which are used for a predictive model of opioid overdose. We present results in geographic regions exhibiting diverse SVI components from the years 2014-2020 displaying correlations between SVI components and opioid overdose risk. These results set the stage for future work in moving from correlation to causation in identifying primary social and environmental factors influencing the opioid epidemic in different geographic regions.

Adam Spannaus
Oak Ridge National Laboratory
Oak Ridge, TN, USA
spannausat@ornl.gov

Hashan Fernando, Anuj J. Kapadia
Oak Ridge National Laboratory
hfernand@vols.utk.edu, kapadiaaj@ornl.gov

Jodie Trafton
Veterans Health Administration
jodie.trafton@va.gov

MS80

Moment Constraints in Phase Recovery for Multireference Alignment

Multireference alignment (MRA) refers to the problem of recovering a signal from noisy samples subject to random circular shifts. Expectation maximization (EM) and variational approaches use statistical modeling to achieve high accuracy at the cost of solving computationally expensive optimization problems. The method of moments, instead, achieves fast reconstructions by utilizing the power spectrum and bispectrum to determine the signal up to shift. Our approach combines the two philosophies by viewing the power spectrum as a manifold on which to constrain the signal. We then maximize the data likelihood function on this manifold with a gradient-based approach to esti-

mate the true signal. Algorithmically, our method involves iterating between template alignment and projections onto the manifold. The method offers increased efficiency compared to EM and demonstrates improved accuracy over the method of moments applied to the bispectrum.

Vahid Shahverdi
KTH Royal Institute of Technology
vahidsha@kth.se

Emanuel Ström
KTH (Royal Institute of Technology, Stockholm)
emastr@kth.se

MS81

A Large-Scale Cohort Study Leverages a Machine Learning Model of Cell-Free DNA for Early Breast Cancer Detection

Cell-free DNA (cfDNA) from liquid biopsy offers a new approach with potential benefits for early breast cancer (BC) detection and screening in a non-invasive diagnostic. However, identifying breast cancer from cfDNA remains a major challenge due to the low tumor fraction. To overcome the difficulties, we developed a nanopore-based sequencing technology to capture the range of 1 to 4 million CpG sites from 1 ml of plasma for each sample. Moreover, we extended the sequencing to a large case-control cohort with a total of 1080 samples, consisting of 440 untreated breast cancer patients and 640 healthy controls. The feature engineering procedure was used to build the methylation-based scores and fragmentations to overcome the high dimensionality and sparsity. Then, we developed an ensemble classification model utilizing a leave-one-out cross-validation (LOOCV) for early BC detection leveraging the large cohort. Two external independent cohorts validated that our model achieved excellent performance in early breast cancer detection. The machine-learning approach holds the potential to improve cancer diagnosis and treatment outcomes in the real world.

Xiangqi Bai
Stanford University
School of Medicine
xiangqi@stanford.edu

Li C. Xia
South China University of Technology
lcx.scut@outlook.com

MS81

Uncertainty Estimation and Selective Generation in Language Models

Large language models (LLMs) have demonstrated remarkable capabilities in comprehending and generating human language text. However, LLMs are found to overconfidently generate wrong answers for questions they do not know (also known as hallucination). Having a reliable estimation for models' uncertainty on their generated content could help to determine when to abstain or to selectively generate, ensuring safe deployment of language models. While likelihood-based metrics such as perplexity are widely employed, recent research has demonstrated the limitations of using sequence-level probability estimates given by LLMs as reliable indicators of generation quality. Therefore, in this talk, we present a few newly developed methods to improve uncertainty estimation for large language models, including self-evaluation, self-consistency, and measuring distance to training domain. We conclude the talk with discussion on limitations of existing methods and the future directions of this problem.

Jie Ren
Google Inc.
Google Brain
jjren@google.com

MS81

Learning Complex Cellular Dynamics from Time-Series Single-Cell Data

Time-series single-cell RNA sequencing (scRNA-seq) data are emerging and provide unprecedented opportunities to learn dynamic processes of cellular systems. However, dynamic inference based on time-series scRNA-seq data is challenging due to the destructive nature of single-cell sequencing, and it remains a challenge to link the scRNA-seq snapshots sampled at different time points. This requires the development of mathematical models and machine learning methods capable of reconstructing cellular dynamics and uncovering the non-linear cell-cell interactions. Optimal transport is a powerful tool for analyzing complex data because it learns an optimal, cost-effective mapping between data distributions. In this talk, I will report our recent work on developing a mean-field modeling and learning framework for cell population dynamics inference from time-series scRNA-seq data. This framework solves the mean field equations via optimal transport coupled with neural networks. It not only reconstructs the cellular dynamic processes, but also learns the nonlinear cell-cell interactions.

Lin Wan
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
lwan@amss.ac.cn

MS81

Improving Cell Type Deconvolution for Spatial Transcriptomics by Leveraging Temporal Information

Spatial transcriptomics (ST) data allows for the study of tissue organization and facilitates the inference of cell-cell interactions. However, the widely used spot-based ST data measures multiple cells at a spot, which makes it challenging to identify the locations of cell types. Cell type deconvolution methods have been developed to estimate the cell type compositions in each spot, and such results are critical for further analysis of the spatial data. Since tissues are formed over a temporal process, we leverage temporal information of cells in order to achieve higher accuracy of cell type deconvolution. In our method TemSOMap, we infer the clonal information of cells and with SpaDecoder we perform deconvolution for temporal spatial samples or slices from a 3D tissue.

Xiuwei Zhang
Georgia Institute of Technology
School of Computational Science and Engineering
xiuwei.zhang@gatech.edu

MS82

Autm Flow: Atomic Unrestricted Time Machine As a Monotonic Normalizing Flow

Normalizing flows constitute an important class of gen-

erative models. It tries to provide a generally nonlinear bijective mapping between the base distribution and the target distribution. How to design an architecture that is computationally efficient and expressive is a central topic. We present a novel normalizing flow called AUTM that allows triangular Jacobian, tractable inverse calculation, and provable universality. We compare AUTM to popular flow architectures such as RealNVP, GLOW, NSF, FFJORD, UMNN, BNAF, etc. to demonstrate the efficiency and expressive power over machine learning and large scale image data sets.

Difeng Cai
Southern Methodist University
ddcai@smu.edu

MS82

Topological Machine Learning for Drug Discovery

In computer-aided drug discovery (CADD), virtual screening (VS) is used for identifying the drug candidates that are most likely to bind to a molecular target in a large library of compounds. In this project, we developed a novel method using multiparameter persistence (MP) homology that produces topological fingerprints of the compounds as multidimensional vectors. Our primary contribution is framing the VS process as a new topology-based graph ranking problem by partitioning a compound into chemical substructures informed by the periodic properties of its atoms and extracting their persistent homology features at multiple resolution levels. We demonstrate that our models, enhanced by the MP signatures, outperform state-of-the-art methods on benchmark datasets by a wide and highly statistically significant margin. Associated Poster: Topological Machine Learning for Drug Discovery

Baris Coskunuzer
University of Texas at Dallas
coskunuz@utdallas.edu

MS82

Data-Driven Solutions for Neuroscience Challenges: Integrating Theory, Computation, and Practice

In this talk, we will explore three recent challenges that required data-driven solutions. First, we will develop machine learning models to forecast subjects' attentiveness based on ocular measurements, with a focus on the impact of goals, feedback, and rewards on sustained attention. Next, we will introduce an automated parameter estimation method for a three-population neuronal network, capable of producing cyclical behavior modeled from human sleep hypnograms. Finally, we will reconstruct Local Field Potential time series from anesthetized and awake rats, both before and during CO2 euthanasia, using the Eigensystem Realization Algorithm to identify an underlying linear dynamical system responsible for generating the observed data.

Pedro Maia
University of Texas at Arlington
pedro.maia@uta.edu

MS82

Computational Nonlinear Filtering: A Deep Learning Approach

Nonlinear filtering is a fundamental problem in signal processing, information theory, communication, control and optimization, and systems theory. In the 1960s, celebrated results on nonlinear filtering were obtained. Nevertheless, the computational issues for nonlinear filtering remained to be a long-standing (60-year-old) and challenging problem. In this talk, in lieu of treating the stochastic partial differential equations for obtaining the conditional distribution or conditional measure, we construct finite-dimensional approximations using deep neural networks for the optimal weights. Two recursions are used in the algorithm. One of them is the approximation of the optimal weight and the other is for approximating the optimal learning rate. [This is a joint work with Qing Zhang (University of Georgia), and Hongjiang Qian (University of Cornell).]

Hongjiang Qian
Cornell University
hongjiang.qian@uconn.edu

George Yin
University of Connecticut, Storrs
gyin@uconn.edu

Qing Zhang
University of Georgia
Department of Mathematics
qz@uga.edu

MS83

A Superlinearly Convergent First-order Method for Nonsmooth Optimization

Nonsmooth optimization problems appear throughout machine learning and signal processing. However, standard first-order methods for nonsmooth optimization can be slow for "poorly conditioned" problems. In this talk, I will present a locally accelerated first-order method that is less sensitive to conditioning and achieves superlinear (i.e., double-exponential) convergence near solutions for a broad family of problems. The algorithm is inspired by Newton's method for solving nonlinear equations.

Vasilis Charisopoulos
University of Chicago
vchariso@uchicago.edu

Damek Davis
Cornell University
ORIE
dsd95@cornell.edu

MS83

A New Hybrid Convex-Nonconvex Method for Large-Scale SDPs

This paper introduces HALLaR, a new first-order method for solving large-scale semidefinite programs (SDPs) with bounded domain. HALLaR is an inexact augmented Lagrangian (AL) method where the AL subproblems are solved by a novel hybrid low-rank (HLR) method. The recipe behind HLR is based on two key ingredients: 1) an adaptive inexact proximal point method with inner acceleration; 2) Frank-Wolfe steps to escape from spurious local stationary points. In contrast to the low-rank method of Burer and Monteiro, HALLaR finds a near optimal solution of SDP instances satisfying strong duality. Computational results comparing HALLaR to state-of-the-art solvers on several large SDP instances arising from maximum stable set, phase retrieval, and matrix completion, show that the

former finds highly accurate solutions in substantially less CPU time than the latter ones. For example, in less than 20 minutes, HALLaR can solve a maximum stable set SDP instance with dimension pair $(n, m) \sim (10^6, 10^7)$ within 10-5 relative precision.

Diego Cifuentes
Georgia Institute of Technology, U.S.
diego.cifuentes@isye.gatech.edu

MS83

The radius of statistical efficiency

Classical results in asymptotic statistics show that the Fisher information matrix controls the difficulty of estimating a statistical model from observed data. In this work, we introduce a companion measure of robustness of an estimation problem: the radius of statistical efficiency (RSE) is the size of the smallest perturbation to the problem data that renders the Fisher information matrix singular. This quantity arises naturally as a statistical analogue of the "distance to ill-posedness", commonly used in numerical analysis and optimization. We compute RSE up to numerical constants for a variety of test bed problems, including principal component analysis, generalized linear models, phase retrieval, blind deconvolution, and matrix completion. In all cases, the RSE quantifies the compatibility between the covariance of the population data and the underlying model parameter.

Mateo Diaz
Johns Hopkins University
mateodd@jhu.edu

MS83

: Statistical trajectory predictions for complex optimization algorithms with random data

Iterative algorithms are the workhorses of modern statistical signal processing and data science. While the choice of an algorithm and its hyperparameters determines both the speed and fidelity of the learning pipeline, it is common for this choice to be made heuristically, either by expensive trial-and-error or by comparing upper bounds on convergence rates of various candidate algorithms. Motivated by these issues, I will present a toolbox for deriving state evolutions for a wide variety of algorithms with random data. These are non-asymptotic and provably near-exact predictions of the statistical behavior of the algorithm, which apply even when the underlying optimization problem is nonconvex or the algorithm is randomly initialized. We will showcase these predictions on deterministic and stochastic variants of complex algorithms employed in some canonical models in statistical machine learning.

Ashwin Pananjady
Georgia Institute of Technology
ashwinpm@gatech.edu

MS84

Hybrid s-step SGD and Federated SGD for Large Scale Optimization

We develop a hybrid, 2D parallel distributed-memory SGD algorithm by combining ideas from s-step methods and federated learning. The proposed algorithm attains a continuous tradeoff between accuracy and performance, while scaling beyond what is currently enabled by s-step SGD and federated SGD alone. We show empirical results that highlight better convergence and better performance when combining the s-step and federated SGD approaches into a 2D parallel algorithm. We also show theoretical parallel cost analyses which highlight the benefits of the hybrid, 2D SGD algorithm.

Aditya Devarakonda
Wake Forest University
devaraa@wfu.edu

MS84

FedOSAA: Accelerating Federated Learning with One-Step Anderson Acceleration

Federated Learning is a decentralized machine learning approach that enables multiple local clients and a central servers to collaboratively learn a model while keeping their data locally. In this work, we introduce FedOSAA, an innovative optimization algorithm for federated learning that combines the simplicity of first-order methods with the accelerated convergence performance typically associated with second-order methods. During the local training, FedOSAA takes a few local gradient descent steps, followed by one Anderson acceleration step which accelerates the convergence while avoiding the common expense of Newton-based methods, i.e., the construction and inversion of the Hessian matrix. We establish a linear convergence rate to the global minimizer for FedOSAA on strongly convex losses. We compare FedOSAA with other state-of-the-art federated learning methods such as FedSVRG and GIANT on logistic regression problems. Numerical evidence demonstrates the superior performance of our algorithm in terms of communication and computation efficiency, which are bottlenecks in federated learning.

Xue Feng
University of California, Davis, U.S.
xffeng@ucdavis.edu

Paul Laiu
Oak Ridge National Laboratory
laiump@ornl.gov

Thomas Strohmer
University of California,Davis
Department of Mathematics
strohmer@math.ucdavis.edu

MS84

Federated Learning from Heterogenous Private Data Silos

Biomedical data, such as electronic medical records, is typically siloed within institutions and contains private information that limits sharing. Federated learning approaches incorporating privacy-preserving algorithms, such as differential privacy, provide an avenue to enabling the development of deep learning models from biomedical data. However, incorporating privacy mechanisms poses challenges to maintaining model accuracy, particularly in the case of underrepresented classes in heterogeneous data silos. We consider the effectiveness of centralized and distributed synthetic data generation approaches to addressing these challenges, using both naive and generative AI methods. We evaluate these approaches in the context of developing models with privacy-preserving federated learning on cancer surveillance datasets from the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results

(SEER) program.

John Gounley, Patrycja Krawczuk, Christopher Stanley, Heidi Hanson
Oak Ridge National Laboratory
gounleyjp@ornl.gov, krawczukpn@ornl.gov, stanleycb@ornl.gov, hansonha@ornl.gov

MS84

Dynamical Low-Rank Training Strategies for Federated Learning

Computational cost on client machines and communication latency are central challenges for federated learning, where the cost of communicating the whole weight matrices and training the entire model on resource-constraint edge devices is prohibitively expensive. This work introduces Federated dynamical low-rank training (FeDLRT), a technique for improving the computing and communication efficiency of federated learning in training neural networks across decentralized devices. Drawing from dynamical low-rank approximation, FeDLRT enables provably robust optimization using only a low-rank factorization on the clients and server of a federated learning setup. FeDLRT offers several key benefits. It focuses on automatic rank adaptation of the low-rank factors of the weight matrices, a feature that significantly reduces communication and computational costs while maintaining model performance. Additionally, it incorporates variance reduction techniques and provides theoretical proof of global loss descent in the federated learning setting. Numerical experiments further demonstrate the effectiveness of FeDLRT in enhancing the efficiency and scalability of federated learning for neural networks.

Steffen Schotthoefer
Oak Ridge National Laboratory
schotthofers@ornl.gov

MS85

Principled Structures in Deep Learning-Based Autoregressive Models of Dynamical Systems

We present a general theoretical framework, based on linear stability analysis and random matrix theory to analyze deep neural networks (DNNs) when applied to nonlinear PDEs, for example, in terms of stability and convergence. Data-driven solvers, which are fast and cheap once trained, have a broad range of applications, e.g., to perform probabilistic weather forecasting or efficient long-term emulation of the climate system. Many recent studies have shown promising results, for short-term forecasts, on canonical PDEs, and even for real-world data, such as atmospheric observations. However, there are also challenges such as long-term instabilities (drifts and blow-ups) and lack of convergence. These challenges have been difficult to address due to the lack of a rigorous (i.e., system- and architecture-agnostic) framework for analyzing DNNs when applied to PDEs. Here, we present such a system- and architecture-agnostic framework that combines numerical analysis, deep learning theory, and ransom matrix theory. The framework is based on linear stability analysis of DNNs wherein we demonstrate the predictive capabilities of the linear operator to the general behavior of the models in short- and long term using ideas from random matrix theory. We show that this framework can guide the development of stable, convergent neural PDE-solvers for any system and architecture, which has wide-ranging applications on nonlinear geophysics, e.g., to build data-driven weather/climate models.

Ashesh Chattopadhyay
UC Santa Cruz
aschatto@ucsc.edu

MS85

Paired Autoencoder for Inverse Problems and Regularization

We consider a decoupled approach for surrogate modeling, where unsupervised learning approaches are used to efficiently represent the input and target spaces separately, and a supervised learning approach is used to represent the mapping from one latent space to another. We demonstrate that our likelihood-free approach can outperform other approaches in scenarios where forward and adjoint operations are computationally prohibitive. We investigate how the approach can be used for defining regularization or prior knowledge, and/or as a surrogate model for inversion for various large-scale applications.

Matthias Chung
Emory University, U.S.
matthias.chung@emory.edu

Eldad Haber
Department of Mathematics
The University of British Columbia
haber@eoas.ubc.ca

Julianne Chung
Department of Mathematics
Emory University
julianne.mei-lynn.chung@emory.edu

Emma Hart
Emory University
ehart5@emory.edu

MS85

Neural Differentiable Modeling with Diffusion-Based Super-Resolution for Predicting Two-Dimensional Turbulence

Simulating spatiotemporal turbulence remains a significant challenge due to its intricate multiscale nature and prohibitive computational demands. Traditional approaches attempt to represent small-scale features in an unresolved manner, which, however, often sacrifice accuracy and lose high-frequency/wavenumber information. In this paper, we introduce an innovative neural differentiable modeling framework designed to enhance the predictability and efficiency of spatiotemporal turbulence simulations. Our approach features hybrid differentiable modeling techniques that seamlessly integrate deep neural networks with numerical PDE solvers within a differentiable programming framework, synergizing deep learning techniques with physics-based CFD modeling. Specifically, a hybrid differentiable neural solver is constructed on a coarser grid to capture large-scale turbulent phenomena, followed by the application of a Bayesian conditional diffusion model that generates the small-scale turbulence conditioned on large-scale flow predictions. The performance is evaluated through comparative analysis against conventional large eddy simulation techniques with physics-based subgrid-scale closures and purely data-driven deep neural solvers. The findings underscore the neural differentiable modeling framework's potential to significantly enhance the accuracy

and computational efficiency of turbulence simulations.

Xiantao Fan, Deepak Akhare
University of Notre Dame
fanxiantao1@gmail.com, dakhare@nd.edu

Jian-Xun Wang
Dept. of Aerospace and Mechanical Engineering
University of Notre Dame
jwang33@nd.edu

MS85

SciML for Bayesian Inference in Optimization Problems Governed by Nonlinear Dynamical Systems

We have developed a data-driven computational framework for solving optimization problems governed by nonlinear dynamical systems. Our approach involves neural networks to learn the reconstruction map from noisy observations to model and noise parameters. We are able to estimate the covariance matrix of the posterior distribution, which allows us to expose uncertainties that arise in our problem formulation. We have tested our approach under different noise conditions and input-output pairs. We report results for various metrics to evaluate the performance of the proposed methodology. We have also investigated how changes in hyperparameters and the network architecture affect the methodology's performance. Lastly, we report results on dedicated hardware to expose the benefits of training neural networks on modern high-performance computing platforms.

Andreas Mang
University of Houston
Department of Mathematics
amang@Central.UH.EDU

MS86

Cut Improvement and Clustering Using Compressed Sensing

Let $G = (V, E)$ be a (possibly weighted) graph. Typically, the clustering problem refers to decomposing V into disjoint subsets $C_1, \ldots, C_k$. such that each C_a has many internal edges and few external edges. However when $n := |V|$ and k are large, finding all k clusters can be undesirable and computationally wasteful. In this talk we study two refinements of the clustering problem:

1. Local Clustering: Given a small set of seed vertices, $\Gamma \subset C_a$,find or approximate C_a.
2. Cut Improvement: Given an approximation $\Omega \approx C_a$, refine Ω to an even better approximation of C_a.

In particular, we introduce a novel algorithm based on techniques from the signal processing field of compressive sensing which is able to solve both problems. We shall show that this algorithm, which we call ClusterPursuit, exhibits strong performance in theory and in practice.

Daniel McKenzie
Colorado School of Mines
dmckenzie@mines.edu

Lai Ming-Jun
University of Georgia
mjlai@uga.edu

MS86

Intrinsic Models in Wasserstein Space with Applications to Molecular Dynamics

We study the problems of efficient modeling and representation learning for probability distributions in Wasserstein space. We consider a general barycentric coding model in which data are represented as Wasserstein-2 (W2) barycenters of a set of fixed reference measures. Leveraging the geometry of W2-space, we develop a tractable optimization program to learn the barycentric coordinates and provide a consistent statistical procedure for learning these coordinates when the measures are accessed only by i.i.d. samples. Our consistency results and algorithms exploit entropic regularization of the optimal transport problem, and the statistical convergence of entropic optimal transport maps will be discussed. We also consider the problem of learning reference measures given observed data. Our regularized approach to dictionary learning in W2-space addresses core problems of ill-posedness and in practice learns interpretable dictionary elements and coefficients useful for downstream tasks. Applications of optimal transport to compression of molecular dynamics simulations will be considered

James Murphy
Tufts University
jm.murphy@tufts.edu

MS86

Graph-Based Clustering with and Without Labeled Nodes

Local clustering aims at extracting a local structure inside a graph without the necessity of knowing the entire graph structure. As the local structure is usually small in size compared to the entire graph, one can think of it as a compressive sensing problem where the indices of target cluster can be thought as a sparse solution to a linear system. In this talk, we apply this idea based on two pioneering works under the same framework and propose a new semi-supervised local clustering approach using only few labeled nodes. Our approach improves the existing works by making the initial cut to be the entire graph and hence overcomes a major limitation of the existing works, which is the low quality of initial cut. Extensive experimental results on various datasets demonstrate the effectiveness of our approach. Finally, we will discuss how the current framework for local clustering can be extended to the unsupervised setting.

Zhaiming Shen, Ming-Jun Lai
University of Georgia
zshen49@gatech.edu, mjlai@uga.edu

Sheng Li
University of Virginia
shengli@virginia.edu

MS86

Localization from Structured Distance Matrices Via Low-rank Matrix Recovery

We study the problem of determining the configuration of n points by using their distances to m nodes, referred to as anchor nodes. One sampling scheme is Nystrom sampling,

which assumes known distances between the anchors and between the anchors and the n points, while the distances among the n points are unknown. For this scheme, a simple adaptation of the Nystrom method could be used to estimate the configuration of the anchors and the n points. We propose a modified version of Nystrom sampling, where the distances from every node to one central node are known, but all other distances are incomplete. We show that this problem can be framed as the recovery of a low-rank submatrix of a Gram matrix. Using synthetic and real data, we demonstrate that the proposed approach can exactly recover configurations of points given sufficient distance samples. Associated Poster: Localization from structured distance matrices via low-rank matrix sensing

Abiy F. Tasissa
Tufts University
Abiy.Tasissa@tufts.edu

Samuel Lichtenberg
John Hopkins University
splichte@gmail.com

MS87

Frequentist Confidence Intervals Via Optimization: Resolving the Burrus Conjecture

We introduce an optimization-based framework to construct confidence intervals for functionals in constrained inverse problems, ensuring valid one-at-a-time frequentist coverage guarantees. Our method builds upon the now-called strict bounds intervals which offer ways to directly incorporate any side information about parameters during inference without introducing external biases. By tying these intervals to an inversion of a constrained likelihood ratio test, we translate interval coverage guarantees into type-I error control, and characterize the resulting interval via solutions of optimization problems. Along the way, we refute the Burrus conjecture. Our framework provides a novel approach to analyze the conjecture and construct a counterexample by employing a stochastic dominance argument, which we also use to disprove a general form of the conjecture. We illustrate our framework with several numerical examples and provide directions for extensions beyond the Rust-Burrus method for non-linear, non-Gaussian settings with general constraints.

Pau Batlle
Caltech
pbatllef@caltech.edu

MS87

Optimal Ridge Regularization for Out-of-Distribution Prediction

We study the behavior of optimal ridge regularization and optimal ridge risk for out-of-distribution prediction, where the test distribution deviates arbitrarily from the train distribution. We establish general conditions that determine the sign of the optimal regularization level under covariate and regression shifts. These conditions capture the alignment between the covariance and signal structures in the train and test data and reveal stark differences compared to the in-distribution setting. For example, a negative regularization level can be optimal under covariate shift or regression shift, even when the training features are isotropic or the design is underparameterized. Furthermore, we prove that the optimally tuned risk is monotonic in the data aspect ratio, even in the out-of-distribution setting and when optimizing over negative regularization levels. In general, our results do not make any modeling assumptions for the train or the test distributions, except for moment bounds, and allow for arbitrary shifts and the widest possible range of (negative) regularization levels.

Pratik Patil
Berkeley
pratikpatil@berkeley.edu

MS87

Efficient Constrained Estimation via Projected Mirror Descent

Constraints on parameter spaces promote various structures in statistical and machine learning tasks. However, they present methodological and computational challenges. These challenges only become more evident in non-Euclidean settings. To address these challenges, we advocate for the use of the Projected Mirror Descent algorithm. In addition to asymptotic consistency, we show how this algorithm can under certain conditions produce asymptotically efficient estimators. Moreover, we connect Projected Mirror Descent to other algorithms of interest and demonstrate our theoretical analysis through various applications, highlighting ways this algorithm can be extended to a broad class of problems.

Rick Presman
Duke University
rick.presman@duke.edu

MS87

Bayesian Constraint Relaxation: Distance and Divergence Penalization

We consider regularizing the squared distance to set-based constraints for several statistical tasks that can be cast as constrained optimization. These distance-to-set penalties are more flexible than many existing algebraic and regularization penalties, and often avoid drawbacks that arise from alternatives such as shrinkage. Moreover, these translate naturally to a flexible class of priors for incorporating relaxed constraints within a Bayesian framework. We show how they are amenable to gradient-based samplers, and derive a natural extension to (Bregman) divergence-to-set priors. We discuss several examples showcasing how this generalization improves performance by making use of information geometry of the data generating mechanism.

Jason Xu
Duke University
jason.q.xu@duke.edu

MS88

A Tensor Approach to Group Fairness in Classification Models

I introduce the problem of defining and measuring fairness at the level of comparing predictive outcomes across subpopulations. I overview the fairness-confusion tensor (FaCT) introduced in Kim, Chen and Talwalkar, ICML 2020, and show that the so-called impossibility theorems for fairness are a natural consequence of the solutions to linear programs defined on the non-negative orthant of possible confusion matrices, suitably augmented with group membership labels. Our FaCT formalism reveals how various fairness definitions can be classified by the codimension of the solution space and imposed geometric constraints on

the solution manifold. I also discuss practical issues of measuring fairness in practice, such as the partially unobserved nature of group membership, as well as some open issues like the realizability of Pareto-optimal fair classifiers.

Jiahao Chen
New York City Office of Technology and Innovation
jiahchen@oti.nyc.gov

MS88

A Simple, Statistically Robust Test of Discrimination

In observational studies of discrimination, the most common statistical approaches consider either the rate at which decisions are made (benchmark tests) or the success rate of those decisions (outcome tests). Both tests, however, have well-known statistical limitations, sometimes suggesting discrimination even when there is none. Despite the fallibility of the benchmark and outcome tests individually, here we prove a surprisingly strong statistical guarantee: under a common non-parametric assumption, at least one of the two tests must be correct; consequently, when both tests agree, they are guaranteed to yield correct conclusions. We present empirical evidence that the underlying assumption holds approximately in several important domains, including lending, education, and criminal justice – and that our hybrid test is robust to the moderate violations of the assumption that we observe in practice. Applying this approach to 2.8 million police stops across California, we find evidence of widespread racial discrimination.

Johann Gaebler
Harvard University
me@jgaeb.com

Sharad Goel
Kennedy School of Government
Harvard University
sgoel@hks.harvard.edu

MS88

Auditing the Algorithmic Pipeline for Disparate Impact

Many algorithms applied in real-world situations result in disparate impact; however, it is often challenging to determine which different parts of the algorithmic pipeline contributed to these biases. In this talk, I present four vignettes of auditing different pieces of the algorithmic pipeline: in data collection, data processing, method assumptions, and model choices. First, I discuss how lack of diversity in underlying training datasets can lead to downstream biases in the case study of Xbox recommendation systems. Second, I discuss how choices about data normalization and variable selection can negatively impact marginalized groups in the case study of the CalEnviroScreen algorithm (which currently guides public funding towards disadvantaged neighborhoods in California). Third, I discuss how certain statistical methods assuming homogeneity can mask differences in heterogeneous populations in the case study of early stopping rules for clinical trials and A/B experiments. Finally, I discuss how modeling choices including use of generative AI can amplify biases for individuals with speech impairments in the case study of OpenAIs Whisper speech-to-text model.

Allison Koenecke
Cornell University
koenecke@cornell.edu

MS88

On the Mathematics of Transportation Equity: Concepts, Models, and a Couple Tenets

In this talk I will share my perspective on the mathematics of transportation equity, with a focus on public transit network design. First, I will outline different conceptualizations of the basic purpose of a transportation system and showcase how these might be reflected in mathematical models for transportation planning. Then, I will present some examples of how these abstractions might, at times unexpectedly, fail to truly advance transportation equity in practice. Far from being discouraging, it is my view that these failures represent opportunities for mathematical innovation. I will conclude with a couple (subjective) tenets to help guide innovation in this area, their connections to data science, and a brief highlight of my current research along these lines.

Juan Carlos Martinez Mori
Georgia Institute of Technology
H. Milton Stewart School of Industrial and Systems Engineeri
jcmm@gatech.edu

MS89

Dual Operating Modes of In-Context Learning

In-context learning (ICL) exhibits dual operating modes: task learning, i.e., acquiring a new skill from in-context samples, and task retrieval, i.e., locating and activating a relevant pretrained skill. Recent theoretical work investigates various mathematical models to analyze ICL, but existing models explain only one operating mode at a time. We introduce a probabilistic model, with which one can explain the dual operating modes of ICL simultaneously. We extend existing models for pretraining data by introducing multiple task groups and task-dependent input distributions, and obtain a quantitative understanding of the two operating modes of ICL. Furthermore, we shed light on an unexplained phenomenon observed in practice: under certain settings, the ICL risk initially increases and then decreases with more in-context examples. Our model offers a plausible explanation for this "early ascent" phenomenon: a limited number of in-context samples may lead to the retrieval of an incorrect skill, thereby increasing the risk, which will eventually diminish as task learning takes effect with more in-context samples. We also theoretically analyze ICL with biased labels, e.g., zero-shot ICL, where in-context examples are assigned random labels. Lastly, we validate our findings and predictions via experiments involving Transformers and large language models.

Kangwook Lee
University of WisconsinMadison
kangwook.lee@wisc.edu

MS89

Theory on Training Dynamics of Transformers

Transformers, as foundation models, have recently revolutionized many machine learning (ML) applications such as natural language processing, computer vision, robotics, etc. Alongside their tremendous experimental successes, there arises a compelling inquiry into the theoretical foun-

dations of the training dynamics of transformer-based ML models; particularly, why transformers trained by the common routine of gradient descent can achieve desired performance. In this talk, I will present our recent results along this direction on two case studies: linear regression in in-context learning and masked image modeling in self-supervised learning. For both problems, we analyze the convergence of the training process over one-layer transformers and characterize the optimality of the attention models upon convergence. Our numerical results further corroborate these theoretical insights. Lastly, I will discuss future directions and open problems in this actively evolving field.

Yingbin Liang
Ohio State University
yingbinliang@gmail.com

MS89

On the Algorithmic Bias of Aligning Large Language Models with RLHF: Preference Collapse, Matching Policy, Regularization

The alignment of large language models (LLMs) using RLHF has emerged as a promising approach to ensure LLMs behave in accordance with human preferences. However, this paper uncovers a significant challenge: the inherent algorithmic bias in RLHF that leads to preference collapse, where LLMs converge to absolute 0-1 preferences dictated by the reference model, disregarding the learned reward model. We establish the necessary and sufficient conditions for achieving a preference matching policy and propose a novel approach, Preference Matching RLHF (PM-RLHF), which provably aligns LLMs with human preferences on the meaningful response space. Our theoretical analysis and empirical results demonstrate that PM-RLHF effectively mitigates the algorithmic bias in RLHF, paving the way for more equitable and aligned LLMs.

Weijie Su
University of Pennsylvania
suw@wharton.upenn.edu

MS89

Uncovering Hidden Geometry in Transformers Via Disentangling Position and Context

Transformers are neural networks that underpin the recent success of large language models. They are often used as black-box models and building blocks of complex AI systems. Yet, it is unclear what information is processed through layers of a transformer, which raises the issue of interpretability. In this talk, I will present an empirical study of transformers by examining various pretrained transformer models via a factor-analysis-inspired decomposition. A surprisingly consistent geometry pattern emerges in hidden states (or intermediate-layer embeddings) across layers, models, and datasets. Our study underscores two statistical notions—smoothness and incoherence—as two key reasons behind the success of LLMs.

Yiqiao Zhong
University of WisconsinMadison
yiqiao.zhong@wisc.edu

MS90

Geometric Neural-Operators for Manifold Point-Cloud Representations: PDE Solvers and Bayesian Inverse Problems

Geometric Neural Operators (GNPs) are introduced for data-driven deep learning of operators for tasks in non-euclidean settings. The approaches allow for handling manifolds of general shape, including with point-cloud representations. We show how GNPs can be used (i) to estimate geometric properties, such as the metric and curvatures, (ii) to develop solvers for Partial Differential Equations (PDEs) on manifolds, and (iii) to solve Bayesian inverse problems for identifying manifold shapes. The developed GNPs provide general methods for data-driven learning of geometric operators.

Paul J. Atzberger, Blaine Quackenbush
University of California-Santa Barbara
atzberg@math.ucsb.edu, bquackenbush@umail.ucsb.edu

MS90

Model-Constrained Uncertainty Quantification for Scientific Deep Learning of Inverse Problems

While Bayesian neural networks facilitate uncertainty quantification (UQ) for neural network predictions, the uncertainty is questionable using Gaussian priors on the weights and biases. Such weights and biases are artificial quantities and parameters, and thus Gaussian priors are a matter of convenience rather than rationale. In this work we present our attempt to circumvent the issue of an interpretable prior for the weights and biases, along with a concise description of our framework for a UQ-enabled, model-constrained, deep-learning-based inverse solver. To justify our approach we provide comprehensive numerical results in problems for 2D inverse heat conductivity, 2D inverse initial conditions for the time-dependent Burgers' equation, and 2D inverse initial conditions for the time-dependent Navier-Stokes equations.

Russell Philley
University of Texas at Austin, U.S.
rsphilley@gmail.com

Tan Bui-Thanh
The University of Texas at Austin
tanbui@oden.utexas.edu

MS90

Using Mixed-Precision Randomized SVD or GKB Algorithms for Low Rank Kronecker Product Approximations of Large Matrices

The quality of the singular value approximation of a matrix A when using the RSVD can be tuned by the use of an oversampling parameter, p, that depends on the problem size and the degree of ill-conditioning of A. On the other hand, it is not standard with the Golub Kahan Bidiagonalization (GKB) algorithm for finding an approximate SVD of a matrix A to also use oversampling, which is equivalent to extending the Krylov subspace that is obtained through the GKB algorithm. In our work we consider both the RSVD and GKB algorithms with oversampling to estimate the SVD of a matrix A. For this presentation our focus is on the approximation of A using a sum of Kronecker Products (KP), leading to a KP rank R approximation to A. Typically, the SVD is required to find the KP rank R approximation to A. We propose, instead, to use the RSVD and GKB algorithms for the needed SVD approximation of the matrix $\tilde{A}$ which is a reordering of the entries of A. We also investigate the use of low precision in the estimate of

the approximation for the SVD and its impact on the accuracy of the KP approximation. Hence we compare both the RSVD and GKB with standard double precision and single precision on the rank R KP approximation, as well as the eventual impact of using this lower precision approximation for the solution of ill-conditioned inverse problems.

Rosemary A. Renaut
Arizona State University
School of Mathematical and Statistical Sciences
renaut@asu.edu

Abdulmajeed Alsubhi
Arizona State University
ahalsubh@asu.edu

MS90

Using Data-Consistent Inversion to Build Population-Informed Priors for Bayesian Inference

In many applications of interest, such as digital twins and additive manufacturing, the aim is to model individual assets that belong to a population of related assets. The uncertain parameters comprising the model of an individual are often calibrated through Bayesian inference; however, if data on the individual is sparse, one must encode more information into the Bayesian prior to constrain the inference problem. This work explores using data-consistent inversion to leverage data/experiments on related assets to construct population-informed priors. This population-informed prior is then used in the Bayesian inference problem allowing us to estimate properties of the individual asset despite the limited availability of individual data. For linear Gaussian systems, we demonstrate how population-informed priors better inform the inference problem, resulting in increased information gain. We then show extensions of this approach applied to more complex systems governed by differential equations.

Rebekah White
Sandia National Laboratories
rebwhit@sandia.gov

John D. Jakeman
Sandia National Labs
jdjakem@sandia.gov

Tim Wildey
Optimization and Uncertainty Quantification Dept.
Sandia National Laboratories
tmwilde@sandia.gov

Troy Butler
University of Colorado Denver
troy.butler@ucdenver.edu

MS91

Topological Characterization and Uncertainty Visualization of Atmospheric Rivers

Atmospheric rivers (ARs) are long, narrow regions of water vapor in the Earths atmosphere that are often associated with extreme weather events. ARs contribute significantly to water supply and flood risk. Characterizing ARs has been a major challenge due to the lack of a universal definition and their structural variations. Existing AR detection tools (ARDTs) produce distinct AR boundaries for the same event, making the risk assessment of ARs a difficult task. Understanding these uncertainties is crucial to improving the predictability of AR impacts including their landfall areas and associated precipitation that could cause catastrophic flooding and landslides over the coastal regions. In this work, we develop an uncertainty visualization framework that captures boundary and interior uncertainties, i.e., structural variations, of an ensemble of ARs that arise from a set of ARDTs. We first provide a statistical overview of the AR boundaries using contour boxplots of Whitaker et al.. We then introduce the topological skeletons of ARs based on Morse complexes that characterize the interior variation of an ensemble of ARs. We propose an uncertainty visualization of these topological skeletons, inspired by MetroSets of Jacobsen et al. that emphasizes the agreements and disagreements across the ensemble members. Through case studies, we demonstrate that the two approaches complement each other, and they provide a more confident outlook on an AR's shape, area, and onshore impact.

Fangfei Lan
University of Utah
fangfei.lan@gmail.com

MS91

Topology and InfoVis: Scatterplot, Line Charts, and Graphs

Insights from data are highly influenced by the shape perceived in the data. However, two problems exist. First, data are often high dimensional, making their shape difficult to visualize. Second, once visualized, they often suffer from scalability and readability issues, even with modest amounts of data. By applying topology-based descriptors to these problems, approaches can utilize the shape for presenting and interacting with data in ways that are mathematically robust and correspond to human perception and cognition. This talk will discuss the applications of topology-based descriptors, namely persistent homology, contour trees, and mapper, in several commonly used visualization types, including scatterplots, line charts, and node-link diagrams.

Paul Rosen
University of Utah
prosen@sci.utah.edu

MS91

MapperType Algorithms for Complex Data and Relations

Mapper and Ball Mapper are Topological Data Analysis tools used for exploring high dimensional point clouds and visualizing scalarvalued functions on those point clouds. Inspired by open questions in knot theory, new features are added to Ball Mapper that enable encoding of the structure, internal relations and symmetries of the point cloud. Moreover, the strengths of Mapper and Ball Mapper constructions are combined to create a tool for comparing high dimensional data descriptors of a single dataset. This new hybrid algorithm, Mapper on Ball Mapper, is applicable to high dimensional lens functions. Applications include cancer genomics, knot and game theory.

Radmila Sazdanovic
North Carolina State University

rsazdan@ncsu.edu

MS91

Topological Dimensionality Reduction Via Persistent Cohomology

The circular coordinates algorithm of de Silva-Vejdemo-Johansson-Morozov, later strengthened by Perea, extracts circular parameterizations of data from persistent cohomology calculations. More precisely, one can obtain coordinate functions from a data set to the unit circle via fist-degree persistent cohomology, providing a valuable tool for data visualization. Since these results were first published a little over a decade ago, the question of whether spherical coordinates, i.e., maps from data to the unit sphere, could be similarly (and directly) obtained from second-degree persistent cohomology remained open. In this talk, we will begin by reviewing the circular coordinates algorithm of de Silva et al., and then discuss work in which we obtain spherical parameterizations of data from second-degree persistent cohomology, describing examples of our algorithm on various data sets. Time permitting, we may briefly discuss additional applications to data sets with multiple circular features. This is joint work with Stefan Schonsheck.

Nikolas Schonsheck
University of Delaware
nischon@udel.edu

MS92

Sample-Efficient Active Learning Strategies for Deep Learning in Scientific Computing

Active learning is an important topic in machine learning for scientific computing in which learning algorithms can query ground truth data selectively to enhance model accuracy. This is increasingly vital in science applications where data acquisition is costly. This talk introduces a broad framework for active learning in regression problems that extends beyond traditional pointwise data samples to include various practical data types such as data from transform domains (e.g., Fourier data), vector-valued data (e.g., gradient-augmented data), data along continuous curves, and multimodal data (i.e., involving combinations of different types of measurements). This framework uses random sampling from a finite number of sampling measures and accommodates arbitrary nonlinear approximation spaces (model classes). We then introduce generalized Christoffel functions to optimize these sampling measures and discuss how this leads to near-optimal sampling strategies for various important problems of interest. The focus will be on applications in scientific computing, highlighting the efficacy of this framework in gradient-augmented learning with polynomials, Magnetic Resonance Imaging (MRI) with generative models, adaptive sampling for solving PDEs using Physics-Informed Neural Networks (PINNs), and operator learning.

Ben Adcock
Simon Fraser University
ben_adcock@sfu.ca

Juan Cardenas
University of Colorado Boulder
juca7891@colorado.edu

Nick Dexter
Florida State University, U.S.
nick.dexter@fsu.edu

Sebastian Moraga
Simon Fraser University
sebastian_moraga_scheuermann@sfu.ca

MS92

Leveraging Randomness for Efficient Active Learning and Adaptive Sampling: An Overview

This talk serves as an introduction to the main topics and trends in leveraging randomness for active learning and adaptive sampling in data-efficient machine learning. We will discuss the common structure underlying various randomized methods for reducing labeling and data acquisition costs in classification, regression, and low-rank approximation tasks. A summary of sample complexity results, algorithmic components of these methods, and open problems will be presented to introduce mini-symposium attendees to the topic and pose the rest of the mini-symposium presentations in a common framework.

Kevin Miller
Department of Mathematics
Brigham Young University, Provo
ksmiller@math.byu.edu

MS92

Dirichlet Active Learning

One basic context for active learning involves selecting unlabeled data points for which label information is likely to improve classification performance. However, many common rules used to select such points are ad hoc or only appropriate in specific contexts. We propose a data-adaptive, Bayesian-inspired approach to active learning called Dirichlet Active Learning (DiAL). This approach models feature conditional probabilities as a Dirichlet Random Field, which provides a clear conceptual foundation for active learning tasks, and which smoothly adapts to both low and high data regimes. Theoretical guarantees and computational experiments will both be described. This is joint work with Kevin Miller.

Ryan Murray
North Carolina State University
rwmurray@ncsu.edu

Kevin Miller
Department of Mathematics
Brigham Young University, Provo
ksmiller@math.byu.edu

MS92

Novelty Sampling for Fast, Effective Data Reduction

We present a new algorithm for reducing a large data set to a small number of landmark data points. The landmarks are randomly selected, yet they account for nearly all the "novelty" in the data. To generate landmarks, we randomly propose data points and accept/reject with probabilities depending on the previous selections. After the generation step, the landmarks can be used to quickly make predictions and find clusters in the data. Landmark-based learning has a memory footprint which is independent of the data size, so the approach is suitable for distilling large data sets with $N >= 10^9$ data points.

Robert Webber

California Institute of Technology
rwebber@ucsd.edu

MS93

Proximal Optimal Transport Divergences and Their Connections with Neural Operators

Measures of discrepancy between probability distributions have become extremely important in Data Science and Machine Learning. Tasks ranging from sampling unnormalized distributions to generative modeling heavily lean on these measures. In this talk, we introduce an information-theoretic divergence termed proximal Optimal Transport divergence. It is the infimal convolution of f-divergences and optimal transport distances, leveraging the computational efficiency and flexibility of the former, while having the ability to compare distributions which are not absolutely continuous. Finally, we discuss its connections with mean-field games, generative flows and neural operators.

Panagiota Birmpa
Herriot-Watt University
P.Birmpa@hw.ac.uk

MS93

The Dynamics of Learning, Sampling and Dimension Reduction

We discuss optimization, sampling, and dimension reduction algorithms as dynamical systems in discrete-time – maps. These three different contexts provide a fertile ground for new analyses and algorithms that marry dynamical systems theory with statistical learning. In nonconvex optimization, the ergodic properties influence the traditional notion of generalization, the performance of the learner on unseen data. We make this connection precise and derive a new dynamics-aware generalization bound based on a statistical notion of algorithmic stability. Next, we introduce a new transport-based sampling algorithm. When the score (gradient of log density) of a target distribution is known, we may write a transport map – a function from samples of a reference distribution to those of the target – as a zero of a score-residual operator. We describe a Newton-Raphson method on function spaces, and as a result, a discrete-time dynamical system that leads to fast sampling algorithms. Finally, we study dimension reduction in nonlinear dynamical systems via the lens of statistical learning.

Nisha Chandramoorthy
Georgia Institute of Technology
nishac@uchicago.edu

Andreas Loukas
Genentech, Roche
loukasan@gmail.com

Youssef M. Marzouk
Massachusetts Institute of Technology
ymarz@mit.edu

Elizabeth Qian
Georgia Institute of Technology
eqian@gatech.edu

MS93

Differentiable Programming for Spectral Approximation of Koopman Operators

We discuss methods for consistent spectral approximation of Koopman evolution operators in continuous-time, measure-preserving dynamical systems. These methods approximate the Koopman generator (which typically has non-trivial continuous spectrum) by an operator with compact resolvent (and thus discrete spectrum) that spectrally converges to the generator in a limit of vanishing regularization. The regularized generators are built by composition with kernel smoothing operators in a manner that allows incorporating known dynamical equations of motion (physics) through automatic differentiation of kernel functions. We illustrate this approach with applications to low-dimensional chaotic systems.

Dimitrios Giannakis
Dartmouth College
Dimitrios.Giannakis@dartmouth.edu

MS93

Learning Dynamical Systems with Optimal Transport

Optimal transportation (OT) theory, initially developed for the transportation problem of goods and other materials, has found numerous applications in PDE, fluid dynamics, geometric inequalities, probability theory, economics, and so on. Moreover, the recent explosion of interest in data science and machine learning surged an immense interest in OT-based techniques mainly because OT provides a notion of distance and geometry on the space of probability measures. In this talk, I will discuss the applications of OT geometry in improving and developing novel computational methods for parameter identification and related problems. In particular, I will discuss parameter identification in dynamical systems.

Levon Nurbekyan
Department of Mathematics
Emory University
levon.nurbekyan@emory.edu

Yunan Yang
Cornell University
yunan.yang@cornell.edu

Robert Martin
U.S. Army Research Office
robert.s.martin163.civ@army.mil

Elisa Negrini
University of California, Los Angeles
enegrini@math.ucla.edu

Mirjeta Pasha
Massachusetts Institute of Technology
In Fall 2024 Virginia Tech
mpasha@mit.edu

MS94

Randomized and Sparse Cholesky for Kernel Methods in High-Dim Scientific Computing

Kernel methods and their probabilistic counterpart, Gaussian processes, have seen extensive use in scientific computing and machine learning. However, the computational challenge often arises from the need to handle dense kernel matrices. Low-rank and sparse structures have become

popular solutions to address this. In this talk, we discuss how to use randomized and multiscale ordering to construct low-rank Cholesky factorization. Additionally, we show how to employ sparsification techniques, leveraging screening effects in spatial statistics,to further sparsify the Cholesky factors. We showcase practical implementations of these techniques in tackling high-dimensional scientific computing tasks. These include solving heterogeneous partial differential equations (PDEs), addressing kernel ridge regression problems prevalent in quantum chemistry, and tackling kernel spectral clustering challenges encountered in molecular biophysics. The randomized part is based on joint work with Ethan Epperly, Rob Webber, Joel Tropp. The multiscale and sparse part is based on joint work with Houman Owhadi and Florian Schaefer.

Yifan Chen
NYU
yifan.chen@nyu.edu

MS94

Preconditioned NFFT-Accelerated Additive Gaussian Processes

Gaussian processes (GPs) are an essential tool in machine learning given their ability to quantify the uncertainty in the model. The covariance matrices of GPs arise from kernel functions and are typically dense and large-scale. Depending on their dimension even computing all their entries is challenging and the cost of matrix-vector products scales quadratically with the dimension, if no customized methods are applied. We present a matrix-free approach that exploits the computational power of the non-equispaced fast Fourier transform (NFFT) and is of linear complexity for fixed accuracy. With this, we cannot only speed up multiplications with the covariance matrix but also take care of the derivatives needed for the gradient method avoiding the costly Hadamard products of the Euclidean distance matrix and the kernel matrix. Our method utilizes a derivative kernel which is then well suited for multiplying with the Hadamard product. We propose to work with an additive kernel where each sub-kernel captures lower order feature interactions only. This potentially provides better accuracy for many real-world datasets and employs the full efficiency of the NFFT approach that pays off most for small input dimensions. Accordingly we introduce a preconditioner for this additive procedure. By applying our preconditioned NFFT-accelerated multiplication technique, fitting the kernel and the derivative kernel will allow for fast tuning of the hyperparameters.

Theresa Wagner
TU Chemnitz
theresa.wagner@math.tu-chemnitz.de

Tianshi Xu, Yuanzhe Xi
Emory University
xuxx1180@gmail.com, yxi26@emory.edu

Martin Stoll
Chemnitz University of Technology
martin.stoll@mathematik.tu-chemnitz.de

MS94

Efficient Preconditioned Unbiased Estimators in Gaussian Processes

Hyperparameter tuning is crucial in Gaussian Process (GP) modeling for achieving accurate predictions. Existing methods often face a trade-off between bias and variance, with traditional approaches introducing bias and randomized-truncated CG (RT-CG) suffering from high variance. In this talk, we introduce the Preconditioned Single-Sample CG (PredSS-CG) estimator, designed to reduce variance while maintaining unbiasedness, thus allowing GP models to handle more complex datasets. We demonstrate the effectiveness of PredSS-CG in accurately estimating Log Marginal Likelihood (LML) and its gradient on several real-world datasets. This research was collaboratively executed with Hua Huang, Shifan Zhao, Edmond Chow, and Yuanzhe Xi.

Tianshi Xu
Emory University
xuxx1180@gmail.com

Hua Huang
Georgia Institute of Technology
huangh223@gatech.edu

Shifan Zhao
Emory University
shifan.zhao@emory.edu

Edmond Chow
School of Computational Science and Engineering
Georgia Institute of Technology
echow@cc.gatech.edu

Yuanzhe Xi
Department of Mathematics
Emory University
yuanzhe.xi@emory.edu

MS95

SE(3) Synchronization by Eigenvectors of Dual Quaternion Matrices

In synchronization problems, the goal is to estimate elements of a group from noisy measurements of their ratios. A popular estimation method for synchronization is the spectral method. It extracts the group elements from eigenvectors of a block matrix formed from the measurements. The eigenvectors must be projected, or "rounded", onto the group. The rounding procedures are constructed ad hoc and increasingly so when applied to synchronization problems over non-compact groups. We developed a spectral approach to synchronization over the non-compact group SE(3), the group of rigid motions of $\mathbb{R}^3$. We based our method on embedding SE(3) into the algebra of dual quaternions, which has deep algebraic connections with the group SE(3). These connections suggest a natural rounding procedure considerably more straightforward than the current state-of-the-art for spectral SE(3) synchronization, which uses a matrix embedding of SE(3). We show by numerical experiments that our approach yields comparable results to the current state-of-the-art in SE(3) synchronization via the spectral method. Thus, our approach reaps the benefits of the dual quaternion embedding of SE(3), while yielding estimators of similar quality. Associated Poster: SE(3) Synchronization by Dual Quaternions

Ido Hadi
Tel Aviv University

idohadi@mail.tau.ac.il

MS95

Bilipschitz Invariant Theory

Machine learning algorithms are designed for Euclidean space, and the first step of many applications is to represent data in a Hilbert space V. Frequently, many elements of V represent the same data point, and furthermore, the equivalence classes are orbits of a group G of unitaries on V. For instance, a point cloud may be represented as columns of a matrix, but the columns could be permuted in any order. To properly model such data, we seek a Euclidean embedding of the quotient metric space V/G. Any such embedding is induced by some G-invariant function $V \to H$ into another Hilbert space, but not just any invariant function will do. With applications in mind, we need the embedding $V/G \to H$ to preserve distance as much as possible. Ideally, it should be bilipschitz. In this talk, we share some bad news and some good news about such bilipschitz invariant functions. The bad news is that classical invariant theory is **not** good enough, as polynomials usually cannot deliver bilipschitz embeddings. The good news is that we offer a large and flexible class of bilipschitz invariants for finite groups acting on finite-dimensional Hilbert spaces.

Jameson Cahill
University of North Carolina Wilmington
cahillj@uncw.edu

Joseph W. Iverson
Iowa State University
jwi@iastate.edu

Dustin Mixon
Air Force Institute of Technology
dustin.mixon@gmail.com

Daniel Packer
The Ohio State University
packer.61@osu.edu

MS95

Bispectrum Unbiasing for Dilation-Invariant Multi-Reference Alignment

Motivated by modern data applications such as cryo-electron microscopy, the goal of classic multi-reference alignment (MRA) is to recover an unknown signal $f : \mathbb{R} \to \mathbb{R}$ from many observations that have been randomly translated and corrupted by additive noise. We consider a generalization of classic MRA where signals are also corrupted by a random scale change, i.e. dilation. We propose a novel data-driven unbiasing procedure which can recover an unbiased estimator of the bispectrum of the unknown signal, given knowledge of the dilation distribution. Lastly, we invert the recovered bispectrum to achieve full signal recovery, and validate our methodology on a set of synthetic signals.

Anna Little
University of Utah
little@math.utah.edu

MS95

Maximum Likelihood and Moment Matching for High-Noise Group Orbit Estimation

Motivated by applications to single-particle cryo-electron microscopy (cryo-EM), we study a problem of group orbit estimation where samples of an unknown signal are observed under uniform random rotations from a rotational group. In high-noise regime, we describe a stratification of the Fisher information eigenvalues according to transcendence degrees in the algebra of group invariants. We relate the critical points of the log-likelihood optimization landscape to those of a sequence of moment matching problems. Some examples including a simplified model of cryo-EM will be discussed.

Sheng Xu
Princeton University
sxu21@princeton.edu

MS96

Singular Layer Physics-Informed Neural Network Methods for Some Fluid Flows

Singular perturbations occur when a small coefficient affects the highest-order derivatives in a system of partial differential equations. From the physical point of view, singular perturbations lead to the formation of narrow regions close to the boundary of a domain, known as boundary layers, where numerous crucial physical processes occur. This presentation explores the analysis of viscous boundary layers and interior layers of some fluid flows and their utilization in developing efficient Physics-Informed Neural Networks (PINNs).

Gung-Min Gie
University of Louisville
USA
gungmin.gie@louisville.edu

MT1

BE: Fundamentals of Deep Learning 1

Xi Chen
University of California, Santa Cruz
billchenxi@uky.edu

MT2

Identifying the Multidisciplinary Competencies for Data Science

Jamie Haddock
Harvey Mudd College
jhaddock@g.hmc.edu

MT4

BE: Fundamentals of Deep Learning 2

Xi Chen
University of California, Santa Cruz
billchenxi@uky.edu

MT5

Introduction to Geometric Data Science

Vitaliy Kurlin

University of Liverpool
vitaliy.kurlin@liverpool.ac.uk

MT6

Panel Discussion: Funding Opportunities in the NSF Division of Mathematical Sciences

Troy Butler
University of Colorado Denver
tdbutler@nsf.gov

MT6

Panel Discussion: Funding Opportunities in the NSF Division of Mathematical Sciences

Yuliya Gorb
National Science Foundation
ygorb@nsf.gov

MT6

Panel Discussion: Funding Opportunities in the NSF Division of Mathematical Sciences

Jodi Mead
Boise State University
jmead@nsf.gov

MT7

Practical Applications of Data Science in Climate Research: Insights from Nasa Internship and Future Work at Oak Ridge National Lab on Adaptive Mesh Refinement and High-Performance Computing

Wiktoria Zielinska
University of Illinois at Urbana-Champaign, U.S.
wiktoria41412@gmail.com

MT8

BE: Mathematics of Generative Modeling: Sampling via Transport

Aimee Maurais, Matthew T. Li
Massachusetts Institute of Technology
maurais@mit.edu, mtcli@mit.edu

MT9

A Fast Algorithm for Single-Pass Online Computation of Higher-Order Bivariate Statistics

Ilya Stupakov
Novosibirsk State Technical University, Russia
istupakov@gmail.com

MT10

Score-Based Generative Models for Inverse Problems

Yuejie Chi
Carnegie Mellon University
yuejiechi@cmu.edu

MT11

Software for Tensor-Based Analysis of General Hypergraphs

Sinan G. Aksoy
Pacific Northwest National Laboratory
sinan.aksoy@pnnl.gov

MT11

Software for Tensor-Based Analysis of General Hypergraphs

Ilya Amburg
Pacific Northwest National Laboratory
ilya.amburg@pnnl.gov

MT12

How Does a Machine Learn Sequences: an Applied Mathematician's Guide to Transformers, State-Space Models, Mamba, and Beyond

Annan Yu
Cornell University
Center for Applied Mathematics
ay262@cornell.edu

MT13

Recent Advance of Statistical and Computational Analysis for Physics-Informed Machine Learning

Yi Pin Lu
Northwestern University
yipin.lu@northwestern.edu

MT14

PYOMO.DOE 2.0: A Tutorial on Model-Building Using Design of Experiments and Parameter Estimation in Pyomo

Alexander Dowling
University of Notre Dame
adowling@nd.edu

MT14

PYOMO.DOE 2.0: A Tutorial on Model-Building Using Design of Experiments and Parameter Estimation in Pyomo

Daniel J. Laky
University of Notre Dame, U.S.
dlaky@nd.edu

MT15

Application of Topological Data Analysis to Bring in the Medicine Side: The Convergence of Uncer-

tainty, Topology, and Deep Learning

Gunnar Carlsson
Stanford University
carlsson@stanford.edu

MT15

Application of Topological Data Analysis to Bring in the Medicine Side: The Convergence of Uncertainty, Topology, and Deep Learning

Colleen Farrelly
Post Urban Ventures, United Kingdom
cfarrelly@med.miami.edu

MT15

Application of Topological Data Analysis to Bring in the Medicine Side: The Convergence of Uncertainty, Topology, and Deep Learning

Quincy Hathaway
West Virginia University
qathathaway@mix.wvu.edu

MT15

Application of Topological Data Analysis to Bring in the Medicine Side: The Convergence of Uncertainty, Topology, and Deep Learning

Jennifer Rozenblit
University of Texas
jrozenblit@utexas.edu

MT15

Application of Topological Data Analysis to Bring in the Medicine Side: The Convergence of Uncertainty, Topology, and Deep Learning

Yashibir Singh
Mayo Clinic, U.S.
yashbir143@gmail.com

MT16

Unifying Spectral and Spatial Graph Neural Networks

Zhiqian Chen
Mississippi State University
zchen@cse.msstate.edu

MT16

Unifying Spectral and Spatial Graph Neural Networks

Lei Zhang
Virginia Tech
zhanglei@vt.edu

MT16

Unifying Spectral and Spatial Graph Neural Networks

Liang Zhao
Emory University
liang.zhao@emory.edu

MT17

Explicit Construction of Global Minimizers in Deep Learning Networks

In this talk, we present some recent joint results with Patricia Munoz Ewald (UT Austin) where we present an explicit construction of global minimizers of the cost function in underparametrized deep Learning networks for the context of supervised learning, with ReLU activation functions, and sequentially linearly separable training data; no use of gradient descent is made here. Moreover, for the overparametrized situation, a geometrically adapted gradient descent flow is presented which we show to be homotopy equivalent to the standard gradient flow; under a rank condition, we prove exponential convergence to zero loss at a uniform rate, and that it is equivalent to linear interpolation in output space under reparametrization of the time variable. The latter implies neural collapse. Links to sub-Riemannian geometry are pointed out.

Thomas Chen
University of Texas at Austin
tc@math.utexas.edu

MT17

Cubical Topological Data Analysis for 2D Electronic Images

We develop a novel statistical data analysis of 2D images based on Topological Data Analysis (TDA) rooted in cubical persistent homology. This methodology provides best non-Euclidean testing and classification procedures for digital imaging data, while essentially requiring no image processing unlike the competing simplicial homology TDA methods, which require extensive image processing.This is joint work with Robert Paige and Vic Patrangenaru. Paige acknowledges support from award NSF-DMS 2311058 and Patrangenaru acknowledges support from award NSF-DMS 2311059.

Tingan Chen
Florida State University, U.S.
tc19r@fsu.edu

Victor Patrangenaru
Florida State University
vic@stat.fsu.edu

Rob Paige
Missouri University of Science and Technology, U.S.
paigero@mst.edu

MT17

Uniqueness of Frechet Mean

The concept of expectation of a probability measure, also

known as Frechet mean in general settings, is one of the most fundamental concepts in statistical learning and probability theory, on which so many methods and theories have been built. However, all these rely on the requirement that a given probability measure has a unique mean. While such uniqueness is ensured if the measure is defined on a simply connected, complete metric space with non-positive curvature, it becomes a complicated issue when the measure is defined on a metric space with non-negative curvature. In this talk, let us focus on this issue of uniqueness for a probability measure defined on a Riemannian manifold with non-negative sectional curvature, point out the key obstacles in establishing uniqueness, report some new results on uniqueness, discuss general strategies to address the issue, and explain their connections with algebraic and differential topology and with harmonic analysis.

Xiongzhi Chen
Washington State University
xiongzhi.chen@wsu.edu

MT18

Physics Informed Neural Networks for Advanced Modeling: PINA Library and Tutorials

Physics Informed Neural network for Advanced modeling (PINA) is an open-source Python library providing an intuitive interface for solving differential equations using PINNs, NOs or both together. Based on PyTorch and PyTorchLightning, PINA offers a simple and intuitive way to formalize a specific (differential) problem and solve it using neural networks. The approximated solution of a differential equation can be implemented using PINA in a few lines of code thanks to the intuitive and user-friendly interface. We are proposing a general introduction and several tutorials, also in combination with model order reduction.

Nicola Demo
SISSA, Mathematics Area, mathLab
International School for Advanced Studies
ndemo@sissa.it

Dario Coscia
SISSA, Mathematics Area, mathLab
dcoscia@sissa.it

Anna Ivagnes
SISSA
anna.ivagnes@sissa.it

Gianluigi Rozza
SISSA, International School for Advanced Studies,
Trieste, Italy
grozza@sissa.it

MT18

Accelerating Numerical Simulations by Model Reduction with Scientific and Physics-Informed Machine Learning

Partial differential equations (PDEs) are invaluable tools for modeling complex physical phenomena. However, only a limited number of PDEs can be solved analytically, leaving the majority of them requiring computationally expensive numerical approximations. To address this challenge, reduced order models (ROMs) have emerged as a promising field in computational sciences, offering efficient computational tools for real-time simulations. In recent years, deep learning techniques have played a pivotal role in advancing efficient ROM methods with exceptional generalization capabilities and reduced computational costs. In this tutorial we explore how classical ROM techniques can be elevated through the integration of deep learning models. Our discussion encompasses a review of existing approaches based on neural operators to enhancing ROM. We will introduce Physics-Informed Neural Networks (PINNs), highlighting their recent advances in inverse modeling, discretive PINNs, as well as multiphase modeling and application to fluid dynamics.

Gianluigi Rozza
SISSA, International School for Advanced Studies,
Trieste, Italy
grozza@sissa.it

Nicola Demo
SISSA, Mathematics Area, mathLab
International School for Advanced Studies
ndemo@sissa.it

Dario Coscia
SISSA, Mathematics Area, mathLab
dcoscia@sissa.it

Anna Ivagnes
SISSA
anna.ivagnes@sissa.it

MT19

Computing Distances, Similarity Metrics to Enable Analysis of Complex Multimodal Data

Input your abstract, including TeX commands, here. The abstract should be no longer than 1500 characters, including spaces. Only input the abstract text. Don't include title or author information here.

Gunay Dogan
National Institute of Standards and Technology
gunay.dogan@nist.gov

MT19

Computing Distances, Similarity Metrics to Enable Analysis of Complex Multimodal Data

Input your abstract, including TeX commands, here. The abstract should be no longer than 1500 characters, including spaces. Only input the abstract text. Don't include title or author information here.

Zachary Grey
National Institute of Standards and Technology, U.S.
zachary.grey@nist.gov

MT19

Computing Distances, Similarity Metrics to Enable Analysis of Complex Multimodal Data

Input your abstract, including TeX commands, here. The abstract should be no longer than 1500 characters, including spaces. Only input the abstract text. Don't include title or author information here.

Adam Pintar
National Institute of Standards and Technology, U.S

adam.pintar@nist.gov

MT19

Computing Distances, Similarity Metrics to Enable Analysis of Complex Multimodal Data

Input your abstract, including TeX commands, here. The abstract should be no longer than 1500 characters, including spaces. Only input the abstract text. Don't include title or author information here.

Gustavo Rohde
University of Virginia
gustavo.rohde@gmail.com

MT20

Adaptive Annealing Scheduler (AdaAnn) for Improved Distribution Approximation

Approximating probability distributions is an important task in machine learning and statistics. Recently, optimization-based methods in variational inference have gained popularity, such as normalizing flows, to provide approximations which allow both sampling and density estimation. Normalizing flows are invertible mappings used to transform simpler distributions into ones that are more complex through optimizing parameters associated with these mappings. With distributions characterized by complicated geometric shapes, such as multiple modes or dependent parameters, providing an accurate approximation can be a challenging task. This talk will describe an adaptive annealing scheduler (AdaAnn) which facilitates computational efficiency and accurate convergence to the desired distribution by deforming an easier to approximate distribution to the desired distribution through adaptively chosen temperature increments.

Emma Cobian
University of Notre Dame
ecobian@nd.edu

Jonathan Hauenstein
University of Notre Dame
Dept. of App. Comp. Math. & Stats.
hauenstein@nd.edu

Fang Liu, Daniele E. Schiavazzi
University of Notre Dame
fang.liu.131@nd.edu, dschiavazzi@nd.edu

MT20

Efficient Algorithms for Sum-of-Minimum Optimization

We introduce a "sum-of-minimum" problem for optimizing machine learning model performance by finding multiple parameter sets, allowing the model to choose the best parameter set for each data point during inference, enhancing overall performance over single-set parameter approaches. Despite the problem's nonconvex, nonsmooth nature, we demonstrate theoretically and empirically that it can be solved to a small global error. Experimental results show our model significantly outperforms existing smooth models.

Lisang Ding
UCLA
lsding@math.ucla.edu

MT20

Homotopy Training Algorithm for Neural Networks

Machine learning has seen remarkable success in various fields and the success has raised intriguing mathematical questions about optimizing algorithms more efficiently. In this talk, I will discuss the neural network model from a nonlinear computation perspective and present recent work on developing a homotopy training algorithm to train neural networks layer-by-layer and node-by-node.

Wenrui Hao
Pennsylvania State University
wxh64@psu.edu

MT20

Zeroth-order Hessian Estimators: Curvature, Low-rankness, and Robustness

A zeroth-order Hessian estimator aims to recover the Hessian matrix of an objective function at any given point, using as few finite-difference computations as possible. This talk focuses on Hessian estimation methods in high-dimensional spaces, covering robust Hessianestimators and matrix recovery methods for estimating low-rank Hessians.

Tianyu Wang
Fudan University
wangtianyu@fudan.edu.cn

MT21

Landfield - Hands-on HPC for MDS

Ryan Landfield
Oak Ridge National Laboratory, U.S.
landfieldre@ornl.gov

MT21

Koon - Hands-on HPC for MDS

In this student and early career targeted session, we will give an overview of High-Performance Computing topics including programming environments, and Machine Learning before helping students work through a set of guided programming challenges that focus on parallel programming aspects for visualizing and managing large data. In high-performance computing, Python is heavily used to analyze scientific data generated by supercomputers. Students will learn how to manage conda environments and the basics of how to use GPU-enabled and parallel Python in the first two challenge exercises, then they will build on this knowledge while exploring guided exercises about tuning a Pytorch-based convolutional neural network or working with a parallel Python-enabled version of HDF5 to generate a galaxy formation visualization from a larger dataset. This tutorial is beginner-friendly in that each exercise comes with sample codes and a guided workbook. Exercises ask students to modify code that is mostly complete so that learners may focus on one key concept at a time and avoid getting mired in unrelated programming details.These guided challenge exercises will be performed on OLCFs Odo training cluster, which is similar to the

Frontier supercomputer. Students will have continued access to the Odo system for two weeks after the workshop to explore challenges that they did not complete.Participants should have a laptop with an internet browser.

Suzanne Parete-Koon
Oak Ridge National Laboratory, U.S.
paretekoonst@ornl.gov

MT23

A Unified Volume-Optimization Framework for Unsupervised Learning

Jingzhou Hu
University of Florida
jingzhouhu@ufl.edu

MT23

A Unified Volume-Optimization Framework for Unsupervised Learning

Kejun Huang
University of Florida
kejun.huang@ufl.edu

MT23

A Unified Volume-Optimization Framework for Unsupervised Learning

Yuchen Sun
University of Florida
yuchen.sun@ufl.edu

MT24

MamBayes: Model-Free Long-Horizon Time Series Forecasting with Mamba and Bayesian Neural Networks

Julian Antolin Camarena
Arizona State University, U.S.
jantolin@asu.edu

MT25

Enhancing Water Accessibility: Optimizing the Shortest Path for Nomadic Communities in Kenya

Maurine Maraka Wafula
United States International University, Africa (USIU-A)
wmaraka70@gmail.com

MT26

Edge-centric Connectomics

Network neuroscience is built atop a network model in which cells, populations, and regions are linked to one another via anatomical or functional connections. Historically, this model has been approached from a node-centric perspective, emphasizing features of neural elements: the number of connections they make, their centrality, module affiliation, etc. However, brain networks can also be examined from an edge-centric perspective that explicitly focuses on properties of connections: their material and metabolic costs, the generative processes that govern connection formation, and their dynamics across time. In this talk, I will present results from several recent papers and highlight findings and advantages of edge-centric network perspectives compared with traditional node-centric network representations.

Richard Betzel
Indiana University
rbetzel@indiana.edu

MT26

Topological Graph Contrastive Learning

Graph contrastive learning (GCL) has recently emerged as a new concept which allows for capitalizing on the strengths of graph neural networks (GNNs) to learn rich representations in a wide variety of applications which involve abundant unlabeled information. However, existing GCL approaches largely tend to overlook the important latent information on higher-order graph substructures. We address this limitation by bringing the concepts of topological invariance and extended persistence on graphs to GCL. In particular, we propose a new contrastive mode which targets topological representations of the two augmented views from the same graph, yielded by extracting latent shape properties of the graph at multiple resolutions and summarized in a form of extended persistence landscapes (EPL). Our extensive numerical results on molecular and chemical compound datasets show that the new Topological Graph Contrastive Learning approach delivers significant performance gains in unsupervised graph classification and also exhibits robustness under noisy scenarios. This is a joint work with Yuzhou Chen, Temple University and Jose Frias, UNAM.

Yulia Gel
University of Texas at Dallas, U.S.
ygl@utdallas.edu

MT26

Topological Time Series Data Analysis: Optimization and Hierarchy

Time series analysis using persistent homology has recently been developed, with several advantages over classical Fourier analysis. In this talk, we will present our recent works in persistent homology for time series data. We will introduce a new method connected to dynamical systems (Kim & Jung 2023, 2024), which takes a slightly different approach from embedding theory. This method enables filtration learning within a machine learning framework with low computational costs, providing an optimal filtration space. Additionally, analyzing time series, such as music data, as graph data with persistent homology provides rich analytical perspectives in filtration (Tran et al., 2023, 2024; Heo et al., 2024), providing a hierarchical structure in persistent measures. These aspects will also be discussed within the machine learning framework

Jae-Hun Jung
Pohang University of Science and Technology (POSTECH)

jung153@postech.ac.kr

MT26

Topological Data Analysis and Machine Learning

Alice Patania
University of Vermont
alice.patania@gmail.com

MT27

Understanding Double Descent

Pratik A. Chaudhari
UCLA
pratikac@seas.upenn.edu

MT27

Understanding Double Descent

Tyler Jarvis
Brigham Young University
jarvis@math.byu.edu

MT27

A Precise Characterization of SGD Stability Using Loss Surface Geometry

Stochastic Gradient Descent (SGD) stands as a cornerstone optimization algorithm with proven real-world empirical successes but relatively limited theoretical understanding. Recent research has illuminated a key factor contributing to its practical efficacy: the implicit regularization it instigates. Several studies have investigated the linear stability property of SGD in the vicinity of a stationary point as a predictive proxy for sharpness and generalization error in overparameterized neural networks (Wu et al., 2022; Jastrzebski et al., 2019; Cohen et al., 2021). In this paper, we delve deeper into the relationship between linear stability and sharpness. More specifically, we meticulously delineate the necessary and sufficient conditions for linear stability, contingent on hyperparameters of SGD and the sharpness at the optimum. Towards this end, we introduce a novel coherence measure of the loss Hessian that encapsulates pertinent geometric properties of the loss function that are relevant to the linear stability of SGD. It enables us to provide a simplified sufficient condition for identifying linear instability at an optimum. Notably, compared to previous works, our analysis relies on significantly milder assumptions and is applicable for a broader class of loss functions than known before, encompassing not only mean-squared error but also cross-entropy loss.

Rajiv Khanna
Purdue University, U.S.
rajivak@purdue.edu

MT27

Revisiting Optimism and Model Complexity in the Wake of Overparameterized Learning

Common practice in modern machine learning involves fitting a large number of parameters relative to the number of observations. Such overparameterized models often exhibit "double descent' (or even "multiple descent') behavior in the prediction error curve when plotted against the raw number of model parameters, or some similar notion of model complexity. We revisit model complexity from first principles, by first reinterpreting and then extending the classical statistical concept of (effective) *degrees of freedom*. Whereas the classical definition is connected to fixed-X prediction error (in which prediction error is defined by averaging over the same, nonrandom covariate points as those used during training), our extension of degrees of freedom is connected to random-X prediction error (in which prediction error is averaged over a new, random sample from the covariate distribution). The random-X setting more naturally embodies modern machine learning problems, where highly complex models, even those complex enough to interpolate the training data, can still lead to desirable generalization performance under appropriate conditions. We demonstrate the utility of our proposed complexity measures through a mix of conceptual arguments, basic theory, and experiments, and illustrate how they can effectively transform the surprising "multiple descent' generalization behavior from modern machine learning into the more familiar "single descent' behavior from classical statistics.

Pratik Patil
University of California, Berkeley
pratikpatil@berkeley.edu

MT28

Model Reduction for Fluid Flow in Porous Media: Applications to Hydrogen Fuel Cells

Jamal H. Al-Smail
KFUPM
jamalhas@kfupm.edu.sa

MT29

Data Science for Science

Sinan G. Aksoy
Pacific Northwest National Laboratory
sinan.aksoy@pnnl.gov

MT29

Data Science for Science

Bill Kay
Pacific Northwest National Laboratory, U.S.
william.kay@pnnl.gov

MT29

Data Science for Science

Clayton Kerce
Georgia Tech Research Institute
clayton.kerce@gtri.gatech.edu

MT29

Data Science for Science

Stephen J. Young
Pacific Northwest National Laboratory

stephen.young@pnnl.gov

MT30

Koon - Hands-on HPC for MDS - Part II

In this student and early career targeted session, we will give an overview of High-Performance Computing topics including programming environments, and Machine Learning before helping students work through a set of guided programming challenges that focus on parallel programming aspects for visualizing and managing large data. In high-performance computing, Python is heavily used to analyze scientific data generated by supercomputers. Students will learn how to manage conda environments and the basics of how to use GPU-enabled and parallel Python in the first two challenge exercises, then they will build on this knowledge while exploring guided exercises about tuning a Pytorch-based convolutional neural network or working with a parallel Python-enabled version of HDF5 to generate a galaxy formation visualization from a larger dataset. This tutorial is beginner-friendly in that each exercise comes with sample codes and a guided workbook. Exercises ask students to modify code that is mostly complete so that learners may focus on one key concept at a time and avoid getting mired in unrelated programming details.These guided challenge exercises will be performed on OLCFs Odo training cluster, which is similar to the Frontier supercomputer. Students will have continued access to the Odo system for two weeks after the workshop to explore challenges that they did not complete.Participants should have a laptop with an internet browser.

Suzanne Parete-Koon
Oak Ridge National Laboratory, U.S.
paretekoonst@ornl.gov

PP1

Novel Numerical Reconstruction and Isolation of Different Nonlinear Dynamics in Video via Latent Space Disentanglement of An Untrained Generator Network and Applications to Dynamic MRI

Processing different types of dynamics in video data is a highly relevant problem in video analysis particularly in dynamic medical imaging where contrast enhancement, respiratory motion and patient moments poses a great challenge due to its effects on the image reconstruction as well as for its subsequent interpretation. The analysis and further processing of the dynamics of interest is often complicated by additional unwanted dynamics. This work proposes a novel nonlinear approach for the reconstruction and subsequent separation of different types of nonlinear dynamics in a video data via deep learning. The dynamic images are represented as the forward mapping of a sequence of time dependent latent space variables via an untrained generator neural network with no supervision. Different kinds of dynamics are characterized independently from each other via latent space disentanglement using one dimensional prior information, called triggers. Leveraging the triggers, the method successfully reconstruct a video containing different dynamics from highly undersampled data with parallel imaging. The model also detect the unknown dynamic and subsequently freeze any selection of dynamics and obtain accurate independent representations of the other dynamics of interest at any phase of the frozen dynamic. The method is tested on both synthetic data and real MRI datasets where contrast intensity, breathing, respiratory and body motion are separated.

Abdullah Abdullah
University of Graz, Graz, Austria
+ The Chinese University of Hong Kong
abdullahuom1@yahoo.com

PP1

MS7: Learning Mechanical Systems From Data via Structured Barycentric Forms

High-fidelity modeling of complex physical phenomena often leads to large scale models containing too many degrees of freedom for efficient simulation. In the case one has access to the realization of such a model, projection-based methods allow for the construction of accurate surrogates that preserve the physical and differential structures from the original model. However, this structure preservation is typically lost in the cases when one does not have access to the state-space realization of the model but rather input-output data. In this work, we consider data resulting from frequency response measurements. Based on our previous work on structured barycentric forms, we present extensions to data-driven modeling approaches such as the Adaptive Anderson-Antoulas (AAA) algorithm, which now enable us to not only learn accurate models from data but also to preserve second-order differential structures in addition. These approaches are well suited for the structured data-driven modeling of mechanical systems. This poster accompanies the talk "Learning Mechanical Systems via a Structured AAA Algorithm".

Michael S. Ackermann
Virginia Tech
amike98@vt.edu

Steffen Werner
Virginia Tech
U.S.
steffen.werner@vt.edu

Ion Victor Gosea
Max Planck Institute Magdeburg
gosea@mpi-magdeburg.mpg.de

Serkan Gugercin
Virginia Tech
Department of Mathematics
gugercin@vt.edu

PP1

Temporal Dynamics in Spatial Random Field Theory: A Methodological Advance in fMRI Data Analysis

This research enhances fMRI data analysis by integrating temporal dynamics into spatial random field theory. We developed a new test statistic,, within the time-adaptive Scale Space Gaussian Random Field Model, focusing on signal detection in fMRI data. It captures the global maximum across spatial and temporal dimensions. Our methodology, employing the Functional Autoregressive (FAR (1)) model, focuses on temporal dependencies and spatial arrangements in data, significantly contributing to neuroimaging studies. We used a simulation approach to estimate the p-value for testing the signal using X_max and understand its advantages in analyzing spatial-

temporal patterns in fMRI data,

Theophilus B. Acquah, khalil Shafie
University Of Northern Colorado
theophilus.acquah@unco.edu, khalil.shafie@unco.edu

PP1

MS12: Evasion Paths in Mobile Sensor Networks

Suppose ball-shaped sensors are scattered in a bounded domain. Unfortunately the sensors don't know their locations (they're not equipped with GPS), and instead only measure which sensors overlap each other. Can you use this connectivity data to determine if the sensors cover the entire domain? I will explain how tools from topology allow you to address this coverage problem. Suppose now that the sensors are moving; an evasion path exists if a moving intruder can avoid overlapping with any sensor. Can you use the time-varying connectivity data of the sensor network to decide if an evasion path exists? Interestingly, there is no method that gives an if-and-only-if condition for the existence of an evasion path, but I will advertise follow-up questions that remain open!

Henry Adams
University of Florida
henry.adams@ufl.edu

PP1

Clustering in Pure-Attention Hardmax Transformers

We study the behaviour in the infinite-depth limit of a transformer model with hardmax self-attention and normalization sublayers, by viewing it as a discrete-time dynamical system acting on a collection of points. Leveraging a simple geometric interpretation of our transformer connected with ideas of hyperplane separation, we establish convergence to a clustered equilibrium and prove that clusters are completely determined by special points called *leaders*. We apply our theoretical understanding to design a model based on our transformer to solve the sentiment analysis task in an interpretable way: the transformer filters out meaningless words by clustering them towards the leaders, identified with words carrying the sentiment of the text such as 'amazing' or 'terrible'.

Albert Alcalde
Friedrich-Alexander-Universität Erlangen-Nürnberg
albert.alcalde@fau.de

Giovanni Fantuzzi
Department of Data Science
Friedrich-Alexander-Universität Erlangen-Nürnberg
giovanni.fantuzzi@fau.de

Enrique Zuazua
University FAU of Erlangen-Nurnberg, Germany
Friedrich-Alexander Universität
enrique.zuazua@fau.de

PP1

MS10: Diffusion-Based Robust Hybrid Deep MRI Reconstruction

Deep learning (DL) techniques have been extensively employed in magnetic resonance imaging (MRI) reconstruction, delivering notable performance enhancements over traditional non-DL methods. Nonetheless, these models have vulnerabilities during testing such as their susceptibility to worst-case or noise-based measurement perturbations, variations in training/testing settings like acceleration factors, contrast, k-space sampling locations, and distribution shifts stemming from unseen lesions and different anatomies. This paper addresses these robustness challenges by leveraging diffusion models. In particular, we present a robustification strategy that improves the resilience of DL-based MRI reconstruction methods by utilizing pre-trained diffusion models as purifiers. In contrast to conventional robustification methods for DL-based MRI reconstruction, such as adversarial training (AT), our proposed approach eliminates the need to tackle a minimax optimization problem. It only necessitates efficient fine-tuning on purified examples. Our experimental findings underscore the effectiveness of our approach in addressing the mentioned instabilities, outperforming standalone diffusion-based MRI reconstructors and leading robustification methods for deep MRI reconstruction, including AT and randomized smoothing. Our experiments show the adaptability of our approach to multiple DL-based MRI reconstruction models and show its robustness for data with unseen lesions.

Ismail R. Alkhouri
Michigan State University; University of Michigan
ismailal@umich.edu

Shijun Liang, Rongrong Wang
Michigan State University
liangs16@msu.edu, wangron6@msu.edu

Qing Qu
University of Michigan
qingqu@umich.edu

Saiprasad Ravishankar
Michigan State University
ravisha3@msu.edu

PP1

Riccati Equations in Sparse Gaussian Graphical Models: Insights and Applications in Equilibrium Networks

In probabilistic graphical models, a crucial statistical inverse problem is learning edges that link nodes representing random variables. For the Gaussian model, this inverse problem concerns estimating the sparsity patternwhether elements are zero or non-zeroof the inverse covariance or precision matrix. Recent research has concentrated on estimating the precision matrix under various constraints for high-dimensional data, where the number of samples surpasses the graph's size. Many fundamental statistical limits were discovered, and efficient algorithms were championed. We present a novel parameterization for the precision matrix, expressing it as LQL, where Q is known, but the symmetric matrix L remains unknown. Our framework extends the Cholesky-type precision matrix decomposition beyond restrictions to triangular or Toeplitz structures. We devise a maximum likelihood estimator (MLE) to learn the sparsity patterns of L and the precision matrix. Leveraging the theory of Matrix Riccati Equations, we solve the MLE and derive optimal sample complexity results. Furthermore, when L is an M-matrix (nonpositive off-diagonals), we establish conditions under which the underlying Gaussian distribution is total positivity of order two and investigate the associated Markov proper-

ties. Finally, we discuss the implications of our approach in learning equilibrium networks; for preliminary results, see (Rayas, Anguluri, and Dasarathy, vol. 35, pp. 14637-14650, NeurIPS 2022).

Rajasekhar Anguluri
University of Maryland, Baltimore County
rajangul@umbc.edu

PP1

MS17: Scalable Estimation of Conditional Brenier Maps via Entropic Optimal Transport

The task of conditional simulation asks to generate samples from the conditional of a joint distribution described by a collection of samples. Conditional Brenier maps, T_{CB} allow for conditional simulation in a single evaluation, though the estimation of such maps is notoriously difficult. [Carlier et al., 2010] show that optimal transport maps under a rescaled quadratic cost, written $T_{\mathrm{B},t}$, asymptotically converge to T_{CB} as the scaling parameter $t \to 0$. While recent works have proposed methods for conditional simulation based on this framework, they fall short of providing quantitative convergence guarantees of their procedure, nor are they necessarily scalable to large sample sizes. We propose to instead use computational methods based on entropic optimal transport to estimate T_{CB}, which, unlike existing approaches, has the advantage of being both principled and scalable to many training points. We demonstrate the performance on various conditional simulation tasks arising from Bayesian inference problems with intractable likelihood functions.

Ricardo Baptista
Caltech
rsb@caltech.edu

PP1

MS14: Machine Learning Applications to Quantum Many-Body Systems

Developing new approaches to electronic structure calculations and quantum many-body dynamics is of extreme relevance for chemical, physical, and materials applications. Although there exists a variety of approximation methods, in conjunction with formally exact theory, quantum many-body systems often exhibit emergent phenomena that defy traditional analytical techniques. This necessitates innovative computational strategies to capture their nuanced dynamics and complex interactions. In this presentation, we explore how to leverage machine learning (ML) methods to capture the dynamics of highly correlated electronic systems.

Hardeep Bassi
Applied Mathematics
University of California, Merced
hbassi2@ucmerced.edu

PP1

MS18: Optimizing Non-Galerkin Algebraic Multigrid for Asymmetric Finite Element Problems

This study investigates the numerical optimization of sparse operator matrices in the non-Galerkin algebraic multigrid (AMG) smoothed aggregation family of algorithms. AMG is a popular method for solving linear systems, such as those arising from the finite element discretization of partial differential equations (PDEs). However, numerical methods such as AMG and Krylov methods often struggle with asymmetric systems, such as advection-diffusion (AD). In this work, we consider PDEs such as advection-dominated AD and anisotropic diffusion as model problems to be solved by the flexible generalized minimal residual (FGMRES) algorithm. These problems are often solved many times in succession in inverse problems, and therefore reducing the number of iterations to convergence of a single forward problem can be beneficial. Thus, optimizing the graph neural network-like V-cycle for a specific problem may yield an improvement many times over. To this end, we take advantage of deep learning framework PyTorch to optimize AMG cycles for these problems in our numerical experiments, leveraging its automatic differentiation capabilities and its ability to execute computations efficiently on graphical processing units (GPUs). We show that significant improvements in convergence rates can be achieved by improving the action of the coarse-level operator matrices, while maintaining fixed sparsity patterns to avoid expensive sparse matrix operations.

Ido Ben-Yair
Ben-Gurion University of the Negev
idobeny@post.bgu.ac.il

Eran Treister
Ben Gurion University of the Negev
erant@cs.bgu.ac.il

PP1

MS1: Continuous Low-Rank Adaptation (CoLoRa) for Reduced Implicit Neural Modeling of Parameterized Partial Differential Equations

This work introduces reduced models based on Continuous Low Rank Adaptation (CoLoRA) that pre-train neural networks for a given partial differential equation and then continuously adapt low-rank weights in time to rapidly predict the evolution of solution fields at new physics parameters and new initial conditions. The adaptation can be either purely data-driven or via an equation-driven variational approach that provides Galerkin-optimal approximations. Because CoLoRA approximates solution fields locally in time, the rank of the weights can be kept small, which means that only few training trajectories are required offline so that CoLoRA is well suited for data-scarce regimes. Predictions with CoLoRA are orders of magnitude faster than with classical methods and their accuracy and parameter efficiency is higher compared to other neural network approaches.

Jules Berman
NYU
jmb1174@nyu.edu

Benjamin Peherstorfer
Courant Institute of Mathematical Sciences
New York University
pehersto@cims.nyu.edu

PP1

MS17: Sampling with Stochastic Bridges

We introduce a family of methods to sample from unnormalized densities via learned stochastic bridges between an auxiliary prior and the target distribution. This includes recently developed methods based on Schrdinger half-bridges and diffusion models. We show how such

methods can be improved using sequential Monte Carlo methods as well as tools from reinforcement learning and stochastic optimal control. Finally, we demonstrate that the resulting deep learning algorithms achieve state-of-the-art performance on several challenging benchmarks.

Julius Berner
California Institute of Technology
jberner@caltech.edu

PP1

MS15: Certifying Fairness with Incomplete Sensitive Information

As the use of machine learning models in real world high-stakes decision settings continues to grow, it is highly important that we are able to audit and control for any potential fairness violations these models may exhibit towards certain groups. To do so, one naturally requires access to sensitive attributes, such as demographics, biological sex, or other potentially sensitive features that determine group membership. Unfortunately, in many settings, this information is often unavailable. In this work we study the well known equalized odds (EOD) definition of fairness. In a setting without sensitive attributes, we first provide tight and computable upper bounds for the EOD violation of a predictor. These bounds precisely reflect the worst possible EOD violation. Second, we demonstrate how one can provably control the worst-case EOD by a new post-processing correction method. Our results characterize when directly controlling for EOD with respect to the predicted sensitive attributes is–and when is not–optimal when it comes to controlling worst-case EOD. Our results hold under assumptions that are milder than previous works, and we illustrate these results with experiments on synthetic and real datasets.

Beepul Bharti
Johns Hopkins University
bbharti1@jhu.edu

PP1

MS3: A Mean-Field Games Approach to Score-Based Generative Modeling

Advancements in diffusion-based generative modeling have made high quality generative models easily accessible. Recently it has been shown that score-based generative models can be formulated in terms of a mean-field game (MFG). This MFG formulation admits alternate characterizations of the SGM based on its optimality conditions, which are a set of coupled nonlinear partial differential equations (PDE). The first PDE is a controlled Fokker-Planck equation which is equivalent to the denoising process in SGM, while the second PDE is a HJB equation that characterizes the optimal controller, whose solution is related to the score function. Based on this mathematical connection, we propose a new objective function for learning the score function. In addition to the implicit score-matching objective, we introduce two regularizers, the first based on measuring the discrepancy in the HJB equation, the second based on measuring discrepancy in satisfying the terminal condition. By including these regularizers, we are informing the structure of the score function, thereby constraining the search space. This strategy is well-grounded through the theory of mean-field games and HJB equations, and develops a PDE-based theory for explaining and understanding the role of latent space in SGM. Through experiments we will show that the HJB regularizer helps learn the score function in a more stable way, while the terminal condition regularizer is associated with higher quality samples.

Benjamin Burns, Benjamin J. Zhang, Markos Katsoulakis
University of Massachusetts Amherst
bburns@umass.edu, bjzhang@umass.edu, markos@umass.edu

PP1

MS9: Active Learning for the Design of Polycrystalline Materials Using Conditional Normalizing Flows

Generative modeling has opened new avenues for solving previously intractable materials design problems. However, these new opportunities are accompanied by a drastic increase in the required amount of training data. This is in stark juxtaposition to the high expense and scarcity of such large materials datasets. In this work, we propose a novel framework for integrating generative models into an active learning loop. This enables the training of generative models with datasets far smaller than what has traditionally been demonstrated. The framework is demonstrated to the problem of designing polycrystalline textures associated with target anisotropic mechanical properties. Our active learning strategy outperforms random sampling with an $\sim$18x reduction in the number of simulations required for training.

Michael Buzzy, Adam Generale
Georgia Institute of Technology
mbuzzy3@gatech.edu, agenerale3@gatech.edu

David Montes, Hojun Lim
Sandia National Labs
dmonte@sandia.gov, hnlim@sandia.gov

Surya Kalidindi
Georgia Institute of Technology
Materials Informatics for Engineering Design
surya.kalidindi@me.gatech.edu

PP1

MS20: Associated Poster: A Zeroth-Order Block Coordinate Descent Algorithm for Huge-Scale Black-Box Optimization

We consider the zeroth-order optimization problem in the huge-scale setting, where the dimension of the problem is so large that performing even basic vector operations on the decision variables is infeasible. In this paper, we propose a novel algorithm, coined ZO-BCD, that exhibits favorable overall query complexity and has a much smaller per-iteration computational complexity. In addition, we discuss how the memory footprint of ZO-BCD can be reduced even further by the clever use of circulant measurement matrices. As an application of our new method, we propose the idea of crafting adversarial attacks on neural network based classifiers in a wavelet domain, which can result in problem dimensions of over one million. In particular, we show that crafting adversarial examples to audio classifiers in a wavelet domain can achieve the state-of-the-art attack success rate of 97.9% with significantly less distortion.

HanQin Cai
University of Central Florida

hqcai@ucf.edu

PP1

MS9: Generative Multi-Fidelity Modeling and Downscaling via Spatial Autoregressive Gaussian Processes

Computer models are often run at different fidelities or resolutions due to trade-offs between computational cost and accuracy. For example, global circulation models can simulate climate on a global scale, but they are too expensive to be run at a fine spatial resolution. Hence, regional climate models (RCMs) forced by GCM output are used to simulate fine-scale climate behavior in regions of interest. We propose a highly scalable generative approach for learning high-fidelity or high-resolution spatial distributions conditional on low-fidelity fields from training data consisting of both high- and low-fidelity output. Our method learns the relevant high-dimensional conditional distribution from a small number of training samples via spatial autoregressive Gaussian processes with suitably chosen regularization-inducing priors. We demonstrate our method on simulated examples and for emulating the RCM distribution corresponding to GCM forcing using past data, which is then applied to future GCM forecasts.

Alejandro Calle-Saldarriaga
University of Wisconsin-Madison
callesaldarr@wisc.edu

Paul Wiemann, Matthias Katzfuss
University of WisconsinMadison
paul.wiemann@wisc.edu, katzfuss@gmail.com

PP1

User Analysis for Instagram Operational Improvement

Instagram is a popular social media site around the world. Maintaining the social media environment has become a severe problem because of the fake users' intervention leading to some malicious activities in recent years. Due to the growth of users and how to enhance user satisfaction, it's essential to identify the types of fake users to set specific regulatory measures, while the types of authenticated users facilitate business cooperation. Combining identification and cluster analysis into algorithms will correlate high accuracy and differentiate the user types. What's more, an item-based ideal value provided will strengthen the users' account build-up. Instagram can take advantage of the functions to detect both fake users and authenticate users to improve the efficiency of the user detection process. Moreover, the adoption of item-based recommendation analysis will also enable the company to develop a user improvement service, helping the users improve their Instagram account operation.

Yiyi Cao
SUNY Binghamton University
ycao7@binghamton.edu

Sijie Jin, Yuxia Zhang, Xuege Gao, Huiming Jia
Columbia University
School of Professional Studies
sj3109@columbia.edu, yz4096@columbia.edu, xg2259@columbia.edu, hj2597@columbia.edu

PP1

MS2: Inference on Interaction Hypergraphs

While there has been tremendous activity in the area of network inference, hypergraphs have not enjoyed the same attention, on account of their relative complexity and the lack of tractable statistical models. We introduce a hyperedge-centric model for analyzing hypergraphs, called the interaction hypergraph, which models natural sampling methods for hypergraphs in neuroscience and communication networks, and accommodates interactions involving different numbers of entities. We define latent embeddings for the interactions in such a network, and analyze their estimators.

Ga-Ming (Angus) Chan
University of Virginia, Department of Statistics
gc8ev@virginia.edu

Zachary Lubberts
University of Virginia
zlubberts@virginia.edu

PP1

MS14: Stochastic Algorithms As Markov Chains in Min-Max and Variational Inequality Problems

Many reinforcement/machine learning problems involve loss minimization, min-max optimization and fixed-point equations, all of which can be cast as Variational Inequality (VI) problems. Stochastic methods like SGD, SEG, and TD/Q Learning are prevalent, and their constant stepsize versions have gained popularity due to effectiveness and robustness. Viewing the algorithm's iterates as a Markov chain, we study their fine-grained probabilistic behavior. In particular, we establish finite-time geometric convergence of the iterate's distribution and relate the ergodicity properties of the Markov chain to the characteristics of the VI, algorithm and data. Using coupling and drift analysis, we characterize the limit distribution and how its bias depends on stepsize. For smooth problems, exemplified by TD learning and smooth min-max optimization, the bias is proportional to the stepsize. For nonsmooth problems, exemplified by Q-learning and ReLU regression, the bias has drastically different behavior and scales with the square root of the stepsize. This probabilistic characterization allows for variance reduction via tail-averaging and bias reduction via Richardson-Romberg extrapolation. The combination of constant stepsize, averaging, and extrapolation provides a favorable balance between fast mixing and low long-run error. We demonstrate its effectiveness in statistical inference compared to traditional diminishing stepsize schemes.

Yudong Chen
University of Wisconsin, Madison
yudong.chen@wisc.edu

PP1

MS3: Sample Complexity of Probability Divergences under Group Symmetry

We rigorously quantify the improvement in the sample complexity of variational divergence estimations for group-invariant distributions. In the cases of the Wasserstein-1 metric and the Lipschitz-regularized a-divergences, the

reduction of sample complexity is proportional to an ambient-dimension-dependent power of the group size. For the maximum mean discrepancy (MMD), the improvement of sample complexity is more nuanced, as it depends on not only the group size but also the choice of kernel. We provide numerical simulations to verify our theories.

Ziyu Chen, Markos Katsoulakis, Luc Rey-Bellet
University of Massachusetts Amherst
ziyuchen@umass.edu, markos@umass.edu,
luc@math.umass.edu

Wei Zhu
UMass Amherst
weizhu@gatech.edu

PP1

MS20: Associated Poster: Revisiting Zeroth-Order Optimization for Memory-Efficient LLM Fine-Tuning: A Benchmark

In the evolving landscape of natural language processing (NLP), fine-tuning pre-trained Large Language Models (LLMs) with first-order (FO) optimizers like SGD and Adam has become standard. Yet, as LLMs grow in size, the substantial memory overhead from back-propagation (BP) for FO gradient computation presents a significant challenge. Addressing this issue is crucial, especially for applications like on-device training where memory efficiency is paramount. This paper proposes a shift towards BP-free, zeroth-order (ZO) optimization as a solution for reducing memory costs during LLM fine-tuning, building on the initial concept introduced by MeZO. Unlike traditional ZO-SGD methods, our work expands the exploration to a wider array of ZO optimization techniques, through a comprehensive, first-of-its-kind benchmarking study across five LLM families (Roberta, OPT, LLaMA, Vicuna, Mistral), three task complexities, and five fine-tuning schemes. Our study unveils previously overlooked optimization principles, highlighting the importance of task alignment, the role of the forward gradient method, and the balance between algorithm complexity and fine-tuning performance. We further introduce novel enhancements to ZO optimization, including block-wise descent, hybrid training, and gradient sparsity. Our study offers a promising direction for achieving further memory-efficient LLM fine-tuning.

Tianlong Chen
The University of North Carolina at Chapel Hill
tianlong@cs.unc.edu

PP1

MS16: Fractional-Order Dynamics Control using KolmogorovArnold Networks(KAN)

Fractional calculus and KolmogorovArnold networks (KANs) are both adept at encapsulating historical dependencies within dynamic systems. Former utilize fractional calculus to inherently model memory effects, whereas KANs leverage their function approximation capabilities to learn and represent historical data from the systems behavior. A Fractional dynamic system combined with KANs was introduced, and we proposed the solution uniquely existed with numerical experiments for validation.Finally, we apply our theory to the Path Tracking system where the slip effect is involved and learned through KANs.

Haowei(Alice) Chen
University of South Florida
haoweichen@usf.edu

Joel Rosenfeld
University of South Florida
Dept. of Math and Stats
rosenfeldj@usf.edu

PP1

MS5: Scalable and Efficient Lazy Maps Enabled by Derivative-informed Neural Operators

We propose a fast and scalable method for solving high-dimensional PDE-constrained Bayesian inverse problems (BIPs). The method synthesizes a structure-exploiting lazy map, a measure transport technique for solving high-dimensional Bayesian inference, and a derivative-informed neural operator (DINO) surrogate of the forward operator. Training a parametrized lazy map to push prior samples to posterior samples involves minimizing the evidence lower bound. This online training procedure requires repeated evaluations of the log-likelihood function and its gradient (i.e., the score function), which can be prohibitive for PDE-based likelihood models. A DINO surrogate is trained offline via derivative-informed operator learning using samples of the forward operator and its Fréchet derivative. The surrogate can achieve high accuracy inbothlog-likelihood and score evaluations at a much lower cost than conventional operator learning methods. Furthermore, a reduced basis DINO surrogate can be seamlessly integrated into the lazy map structure by sharing a set of reduced bases spanning a derivative-informed subspace found from samples. Numerical results on challenging BIPs suggest that highly accurate DINO-driven lazy maps can be trained as quickly as a few seconds up to a few minutes, depending on the difficulty of the BIPs.

Joshua Chen
The University of Texas at Austin
joshuawchen@utexas.edu

Michael C. Brennan
MIT, CCSE
mcbrenn@mit.edu

Thomas O'Leary-Roseberry
The University of Texas at Austin
Oden Institute for Computational and Engineering Sciences
tom@oden.utexas.edu

Lianghao Cao
California Institute of Technology
lianghao@caltech.edu

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

Youssef M. Marzouk
Massachusetts Institute of Technology
ymarz@mit.edu

PP1

MS19: Variational Framework for Super-Resolution 3D Surface Reconstruction from Limited Inputs in Multimodal Imaging

In many imaging applications, the obtained images are of-

ten in low resolution or even have missing data due to practical and hardware limitations, which can impact subsequent 3D reconstruction. For instance, X-ray imaging carries radiation risks, limiting data collection time. Terahertz (THz) imaging, while safe, is slow and affected by diffraction and noise. Magnetic Resonance Imaging (MRI) struggles with the small region of interest in a high-resolution image. To address these challenges, a new framework using the Euler-Elastica regulariser is presented to reconstruct high-resolution surfaces from a few low-resolution 2D slices, combining mathematical models with local edge features and global smoothness. Two algorithms are developed (a projected gradient descent method and the alternating direction method of multipliers), and quantitative comparisons based on discrete curvatures show superiority over other regularisers. Practical examples in X-ray, MRI, and THz imaging validate its effectiveness, offering potential applications in medical imaging and computer vision. Joint work with Prof Ke Chen (Strathclyde and Liverpool), and Prof Shang-Hua Yang (NTHU, Taiwan).

Ke Chen
University of Strathclyde
k.chen@strath.ac.uk

Yiyao Zhang
Centre for Math. Imag. Techniques and Dept. Math. Sciences
University of Liverpool, UK and National Tsing Hua Uni., TW
Yiyao.Zhang@liverpool.ac.uk

Shang-Hua Yang
Department of Electrical Engineering
National Tsing Hua University
shanghua@ee.nthu.edu.tw

PP1

Understanding Fluid Flux: Topological Data Analysis of Porous Media

This research project utilizes topological data analysis (TDA) to quantify the material properties of porous media samples. The main question driving this study is whether there are correlations between the fluid flux and the topological properties of the considered porous materials; and if so, to identify and quantify such correlations. Our goals and objectives include analyzing the experimental images (CT scans of porous material samples), computing their topological properties, and establishing correlations with the experimental results. For this purpose, we utilize methods based on persistence homology that allow us to quantify the porous material properties in a clear and precise manner. Our poster presentation will discuss the methodology, preliminary results, and the potential impact of this research on understanding fluid behavior in porous media.

Ebru N. Dagdelen
New Jersey Institute of Technology
end@njit.edu

Lou Kondic
Department of Mathematical Sciences, NJIT
University Heights, Newark, NJ 07102
kondic@njit.edu

Linda Cummings, Jonathan Jaquette
New Jersey Institute of Technology
linda.cummings@njit.edu, jonathan.jaquette@njit.edu

Anish Vankadhara
South Brunswick High School
anishv0821@gmail.com

PP1

A Text-Based Classification to Escalate the Predictability of Fake Tweet Actors in Troll Networks

Russian internet trolls utilize fake identities to distribute false information across various social media channels. Drawing from previous research on social media influence networks, we present a novel approach to map these operations. Utilizing Twitter data connected to the Russian influence network, we developed a predictive model to outline these network activities. We categorize accounts based on their authenticity role within a subset. This categorization involves defining logical groupings and training a model to recognize similar behavioral patterns across the network. Since tweets and hashtags form Twitter's core, we analyze these key elements to comprehend the authenticity function of diverse actors and behaviors. We utilize Natural Language Processing techniques to group text data and match tweets to logical categories. Comparing the categories with their respective clusters validates our approach. This comparison demonstrates that text clustering accuracy can exceed 90%, enhancing the predictive model's performance. The prediction and validation results indicate our model can map actors within such networks. Moreover, visualizing and analyzing Twitter activities and patterns aids in understanding Russian troll behaviors. Due to the complexity of activity data, we employ dimensional reduction techniques. This allows us to visualize Twitter actions and analyze patterns, uncovering intriguing relationships and behaviors.

Sachith E. Dassanayaka
Wittenberg University
dassanayakas@wittenberg.edu

PP1

Efficacy of Hybrid Echo State Networks in Forecasting Cardiac Electrophysiological Signals

Understanding cardiac electrophysiological signaling is key to discovering, preventing, and treating cardiac arrhythmias. Much of the modeling research in this area has utilized mechanistic mathematical descriptions of the cardiac action potential to predict the behavior of the electrical potential of cardiac cells. Recently, machine learning, including neural networks, has provided a different approach for this problem. Echo State Networks, a data-driven machine learning technique, show promise for learning from past input signals to predict future ones. An Echo State Network is a type of Recurrent Neural Network with sparsely connected neurons, where only the output layer weights are subject to adjustment, and other parameters remain untrained. The special structure of Echo State Networks provides them with high memory capacity and the ability to capture the features of input signals properly. The hybrid approach of introducing a traditional knowledge-based model to the data-driven Echo State Network may outperform strategies that rely only on mathematical models or the reservoir approach. In this poster, we explore the performance of the Hybrid Echo State Network when applied to synthetic and real-world cardiac data.

Afrouz Delshad, Darby Cairns, Elizabeth Cherry

Georgia Institute of Technology
delshad@gatech.edu, dcairns6@gatech.edu, echerry30@gatech.edu

PP1

Constrained Tensor Decompositions for Low-Rank Modeling of Multiphysics Simulation Data

High-dimensional data is ubiquitous across scientific computing disciplines, including plasma physics, fluids, earth systems, and mechanics. Such data is naturally represented as a tensor, consisting of the value of each simulation variable at each point in space and time. Tensor decomposition methods, akin to matrix factorization methods for two-dimensional data, facilitate powerful analysis/reduction of such data, including data compression, surrogate modeling, pattern identification, and anomaly detection. However, existing tensor decomposition methods target simple statistical error metrics, most commonly least-squares loss, resulting in low-rank models of the data that fail to faithfully represent important physics quantities of interest or invariants arising from conservation principles. In this work, we explore new formulations of two common tensor decomposition methods, the Canonical Polyadic (CP) and Tucker decompositions, that attempt to better preserve these quantities by incorporating them directly in the optimization problems that define the resulting low-rank models. We then explore solving these optimization problems and investigate their ability to preserve these quantities compared to their overall reconstruction accuracy. Computational results of applying this approach to CP and Tucker decomposition of data arising from simulation of plasma physics and combustion will be presented.

Danny Dunlavy
Sandia Labs
dannydunlavy@gmail.com

Eric Phipps
Sandia National Laboratories
Center for Computing Research
etphipp@sandia.gov

Hemanth Kolla
Sandia National Laboratories
hnkolla@sandia.gov

John Shadid, Edward Phillips
Sandia Labs
jnshadi@sandia.gov, egphill86@gmail.com

PP1

Beyond the Shadow of a Doubt: Certifying Dnn Training Performance with Real-Time Data-Aware Guarantees

Non-convex optimization is the new frontier in machine learning, particularly in the application of deep neural networks (DNNs). Due to the difficulties encountered in finding the global optima of non-convex problems, local optimization algorithms have been a key component of the non-convex literature. Recently, there has been a spurt of interest in non-convex optimization methods that can guarantee the quality of their approximate solutions. When it comes to guaranteed non-convex optimization, one may look for certain conditions in which a non-convex problem can be solved optimally in polynomial-time. However, a more useful approach in the context of deep learning is to establish data-aware guarantees for a rather large class of problem instances. Such data-aware guarantees can contribute to machine learning on two different levels: i) guarantees for optimal training and ii) guarantees for the effectiveness of DNN models. Within this framework, in this paper, we will explore the first guarantee type, namely developing the data-aware training guarantees for DNNs. This methodology not only allows us to evaluate the efficiency of training in relation to the highest potential training outcomes dictated by the data but also offers a mechanism to certify the proficiency of the training process.

Arian Eamaz, Farhang Yeganegi, Mojtaba Soltanalian
University of Illinois - Chicago, IL
University of Illinois - Chicago, IL
aeamaz2@uic.edu, fyegan2@uic.edu, msol@uic.edu

PP1

MS7: HdR Roms: Hodge-DeRham Reduced Order Models

The Hodge deRham complex and it's compatible discretization is a fundamental tool in the development of structure-preserving numerical methods, that are widely used in full order models. For example, it underlies the stability of mixed FEM methods for elliptic problems. However, existing methods for reduced order models (ROM) methods, such as proper orthogonal decomposition (POD), are not able to obtain a discrete version of the deRham complex. In this poster we will discuss the development of a POD-based ROM with a discrete version of the deRham complex. This is based on exploiting the discrete Hodge decomposition of the underlying FOM discretization to obtain a reduced model in terms of exact, coexact and harmonic components. We will present results for HdR ROM applied to a variety of elliptic problems, and illustrate it's advantages compared to standard POD ROM approaches.

Christopher Eldred, Anthony Gruber, Paul Kuberry
Sandia National Laboratories
celdred@sandia.gov, adgrube@sandia.gov, pakuber@sandia.gov

PP1

MS18: Feature Transportation Improves Graph Neural Networks

Graph neural networks (GNNs) have shown remarkable success in learning representations for graph-structured data. However, GNNs still face challenges in modeling complex phenomena that involve feature transportation. In this paper, we propose a novel GNN architecture inspired by Advection-Diffusion-Reaction systems, called ADR-GNN. Advection models feature transportation, while diffusion captures the local smoothing of features, and reaction represents the non-linear transformation between feature channels. We provide an analysis of the qualitative behavior of ADR-GNN, that shows the benefit of combining advection, diffusion, and reaction. To demonstrate its efficacy, we evaluate ADR-GNN on real-world node classification and spatio-temporal datasets, and show that it improves or offers competitive performance compared to state-of-the-art networks.

Moshe Eliasof
University of Cambridge
eliasof@post.bgu.ac.il

Eldad Haber

Department of Mathematics
The University of British Columbia
haber@eoas.ubc.ca

Eran Treister
Ben Gurion University of the Negev
erant@cs.bgu.ac.il

PP1

MS9: Sparse Recovery of Poincare-Steklov Operators from Matrix-Vector Products

Poincare-Steklov operators arise from elliptic partial differential equations, representing transformations between different types of boundary conditions. Poincare-Steklov operators and their discretizations are similar in structure to integral and differential operators, and the Neumann-to-Dirichlet map in particular resembles an elliptic solution operator. Previous work by Schaefer and Owhadi has shown that the Cholesky factors of elliptic solvers in d spatial dimensions can be approximated to accuracy ε in $O(\log(N)\log^d(N/\varepsilon))$ matrix vector products. In this work, we give an algorithm that aims to approximate the Cholesky factor of a Neumann-to-Dirichlet operator to accuracy ε in $O(\log(N)\log^{d-1}(N/\varepsilon))$ matrix-vector products with the operator, accessed only as a black box using carefully chosen right-hand sides. This cost-accuracy tradeoff is taken from Schaefer and Owhadi's algorithm, and we have empirically observed results in line with this bound for the Neumman-to Dirichlet map. Additionally, we have found that our approximation can be applied to inverse problems arising from elliptic PDEs. We have seen promising results when using our approximation as a stand-in for the exact operator when solving these inverse problems, typically with better performance compared to an optimal low-rank approximation derived from an equal number of matrix-vector products.

Christian X. Engman, Florian Schaefer
Georgia Institute of Technology
cengman3@gatech.edu, fts@gatech.edu

PP1

Community-Size Biases in Statistical Inference in Temporal Networks

Researchers often consider the evolution of communities — sets of densely connected nodes that are connected sparsely to nodes in other communities — in temporal networks. We study the performance of a class of statistical inference methods for such community detection. We show that models that generate community assignments via either a uniform distribution on community assignments or node-wise evolution — with the community assignment of a node in a given layer arising from its community assignments in previous layers — are biased against generating communities with large or small numbers of nodes. We then demonstrate that statistical inference methods that use such generative models tend to poorly identify community structure in networks with large or small communities. To rectify this, we introduce a novel model to generate community assignments that uses all of the community assignments in the previous layer to generate the community assignments for a given layer. Our model greatly mitigates the bias against large and small communities. Consequently, statistical inference that uses our generative model is beneficial for identifying community structure in networks with such communities.

Theodore Y. Faust
University of California, Los Angeles
tfaust@math.ucla.edu

Mason A. Porter
UCLA
mason@math.ucla.edu

PP1

MS16: Learning Divergence-Free Vector Fields from Data via the Occupation Kernel Method

The occupation kernel method (OCK) has proven itself as a robust and efficient method for learning nonparametric systems of ordinary differential equations from trajectories in arbitrary dimensions. Using an implicit formulation provided by vector-valued reproducing kernel Hilbert spaces, we aim to show how the OCK method can be adapted to learn vector fields satisfying physical constraints. In this presentation, we focus on how particular choices of kernel can ensure that the learned vector fields are analytically divergence-free. We validate the proposed method through experiments on a variety of simulated and real datasets. It is shown that the added constraints often lead to better approximations in these application specific problems.

Nicholas Fisher
Portland State University
nicholfi@pdx.edu

PP1

MS4: Wasserstein-Based Explainability for Fairness of ML Models

The objective of this work is to introduce a fairness interpretability framework for measuring and explaining the bias in classification and regression models at the level of a distribution. In our work, we measure the model bias across sub-population distributions in the model output using the Wasserstein metric. To properly quantify the contributions of predictors, we take into account the favorability of both the model and predictors with respect to the non-protected class. The quantification is accomplished by the use of transport theory, which gives rise to the decomposition of the model bias and bias explanations to positive and negative contributions. To gain more insight into the role of favorability and allow for additivity of bias explanations, we adapt techniques from cooperative game theory.

Ryan Franks
Discover Financial Services
ryanfranks1996@gmail.com

Alexey Miroshnikov
Discover Financial Services
U.S
amiroshn@terpmail.umd.edu

Konstandinous Kotsiopoulos, Arjun Ravi Kannan
Discover Financial Services
kkotsiop@gmail.com, arjun.kannan@gmail.com

PP1

Approximation of Solution Operators for High-

Dimensional PDEs

We propose a finite-dimensional nonlinear model to approximate solution operators for evolutional partial differential equations (PDEs), particularly in high-dimensions. By employing a general reduced-order model, such as a deep neural network, we connect the evolution of the model parameters with trajectories in a corresponding function space. Using the computational technique of neural ordinary differential equation, we learn the control field over the parameter space such that from any initial starting point, the controlled trajectories closely approximate the solutions to the PDE. Approximation accuracy is justified for a general class of second-order nonlinear PDEs. Numerical results are presented for several high-dimensional PDEs, including real-world applications to solving Hamilton-Jacobi-Bellman equations. These are demonstrated to show the accuracy and efficiency of the proposed method.

Nathan Gaby
Georgia State University
ngaby1@gsu.edu

PP1

MS1: Optimisation Algorithms on Riemannian Manifolds with Application to Neural Networks

We consider neural networks (NN) as discretizations of continuous dynamical systems. There are two relevant systems: the NN architecture on one side and the gradient flow for optimizing the parameters on the other. In both cases, stability properties of the discretization methods can be relevant e.g. for adversarial robustness. Moreover, to prevent the problem of exploding or vanishing gradient, it is common to consider NNs whose feature space and/or parameter space is a Riemannian manifold. We therefore investigate the stability of numerical one-step integrators defined on Riemannian manifolds. We focus on the analysis of geodesic versions of the Implicit Euler (GIE) and Explicit Euler (GEE) schemes. We apply these methods to non-expansive systems to discuss whether they reflect their contractivity or not.

Marta Ghirardelli
Norwegian University of Science and Technology
marta.ghirardelli@ntnu.no

Brynjulf Owren
NTNU
brynjulf.owren@math.ntnu.no

PP1

Tree-Based Learning for High-Fidelity Prediction of Chaos

Data-driven forecasting of chaotic systems using machine learningis highly soughtafter in many applications, such as forecasting climate and weather phenomena, stock market indices, and predicting pathological activity in biomedical signals. However, existing solutions, like neural networks-based reservoir computing (RC) and long short-term memory (LSTM), contain numerous model hyperparameters that must be tuned, often requiring high computational resources and large training datasets. Here, we propose a chaotic time series prediction technique with single-digit hyperparameters that employs time delayoverembeddingand regression tree ensembles. Furthermore, we derive optimal values for these hyperparameters through extensive statistical testing based on training data, eliminating the need for users to tune hyperparameters. Lastly, we demonstrate our proposed approach's state-of-the-art performance on various benchmark tasks, including the discrete Henon map, the continuous Lorenz system, the spatiotemporal Kuramoto-Shivashinsky system, and Southern Oscillation Index (SOI), a crucial but noisy climate time series with few samples, to demonstrate its effectiveness in real-world settings. Additionally, we demonstrate not only the ability of our proposed approach to perform accurate forecasts with less training volume than alternative methods but also the ability of our hyperparameter prescriptions to balance the accuracy and speed of the proposed approach.

Adam Giammarese, Kamal Rana
Rochester Institute of Technology
amg2889@rit.edu, kr7843@rit.edu

Erik Bollt
Clarkson University
ebollt@clarkson.edu

Nishant Malik
Rochester Institute of Technology
nxmsma@rit.edu

PP1

MS4: Mechanistic Interpretability of Financial Large Language Models

Large Language Models (LLMs) like GPT have significant power in generating human-like output for a variety of tasks. However, due to their nature, these models pose significant challenges for interpreting their underlying decision-making processes. This is even more crucial when it comes to financial institutions, where concerns regarding bias, fairness, and reliability are highly critical. Mechanistic interpretability (MechInterp) [Olah, 2022] deals with reverse engineering complex AI models such as transformers and DNNs. Particularly, there have been recent advances in making the algorithms that these models implement legible to humans, namely: Neel, et al. 2023 providing an explanation for grokking using MechInterp for a small transformer; Kevin, et al. 2022 finding a circuit responsible for an IOI task; and more recently, Callum, et al. 2023 finding an attention head with a functionality known as Copy Suppression. This poster aims at leveraging MechInterp in order to demystify the inner workings of financial LLMs. Specifically, we start with creating algorithmic tasks tailored to understanding open source LLMs outputs when faced with prompts related to the financial industry. Using a causal intervention method known as path patching (see [Kevin, et al. 2022]), we try to find a circuit and attention heads responsible for completing various tasks that investigate bias and fairness in LLMs. Our work is a preliminary step in leveraging MechInterp to understand financial LLMs.

Ashkan Golgoon, Khashayar Filom, Arjun Ravi Kannan
Discover Financial Services
agolgoon3@gatech.edu, khashayar.1367@gmail.com, arjun.kannan@gmail.com

PP1

New Family of Refinable Functions, Easily Learned by Neural Networks

Two-scale difference equations, also entitled the refinable

functions, are the pillar stones in many fields, among them are approximation theory, signal, and image processing. By applying different refinement rules, it's possible to generate subdivision schemes, B-splines, Wavelets, and fractals. Even though earlier work proved that all refinable functions can be implemented, up to arbitrary high precision, by ReLu-based Neural Networks, it was far from clear how such functions could be learned from data. We propose a different type of refinement that involves not only translation and rescaling but also mirroring; functions satisfying the resulting reflecto-refinable equations still generate multiresolution hierarchies that provide an excellent approximation for many functional spaces of interest, yet are also adapted to ReLu networks. We will illustrate the proposed methodology to create new function families.

Shira Golovin
Duke
shirafaigen@gmail.com

Ingrid Daubechies
Duke
Department of Mathematics
ingrid@math.duke.edu

PP1

MS5: Derivative-Informed Neural Operators for Optimization under Uncertainty

In this work we investigate the application of Derivative-Informed Neural Operators (DINOs) to PDE-constrained optimization under uncertainty (OUU). Building on DINO, we propose an operator learning framework for parametric PDE problems, with additional training on derivatives of the input-output map. Through this, we attain an accurate approximation of the PDE solution operator and its derivatives for the optimization problem. We present numerical examples such as flow control and aerodynamic shape optimization problems to test our method and demonstrate its performance. Through the numerical experiments, we show that DINOs can be orders of magnitude more sample-efficient than non-derivative informed operator learning strategies.

Xindi Gong
Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin
xindi.gong@utexas.edu

Dingcheng Luo
Oden Institute for Computational Engineering and Sciences,
The University of Texas at Austin
dc.luo@utexas.edu

Thomas O'Leary-Roseberry
The University of Texas at Austin
Oden Institute for Computational and Engineering Sciences
tom@oden.utexas.edu

Ruanui Nicholson
Department of Engineering Science
University of Auckland
ruanui.nicholson@auckland.ac.nz

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

PP1

MS18: Physics-Guided Full Waveform Inversion Using Encoder-Solver Convolutional Neural Networks (Poster)

Full Waveform Inversion (FWI) is an inverse problem for estimating the wave velocity distribution in a given domain, based on observed data on the boundaries. The inversion is computationally demanding because we are required to solve multiple forward problems, either in time or frequency domains, to simulate data that are then iteratively fitted to the observed data. We consider FWI in the frequency domain, where the Helmholtz equation is used as a forward model, and its repeated solution is the main computational bottleneck of the inversion process. To ease this cost, we integrate a learning process of an encoder-solver preconditioner that is based on convolutional neural networks (CNNs). The encoder-solver is trained to effectively precondition the discretized Helmholtz operator given velocity medium parameters. Then, by re-training the CNN between the iterations of the optimization process, the encoder-solver is adapted to the iteratively evolving velocity medium as part of the inversion. Without retraining, the performance of the solver deteriorates as the medium changes. Using our light retraining procedures, we obtain the forward simulations effectively throughout the process. We demonstrate our approach to solving FWI problems using 2D geophysical models with high-frequency data.

Matan Goren
Ben-Gurion University of the Negev
gorenm@post.bgu.ac.il

PP1

MS14: A Latent Linear Model for Nonlinear Coupled Oscillators on Graphs

A system of coupled oscillators on an arbitrary graph is locally driven by the tendency to mutual synchronization between nearby oscillators, but can and often exhibit nonlinear behavior on the whole graph. Understanding such nonlinear behavior has been a key challenge in predicting whether all oscillators in such a system will eventually synchronize. In this paper, we demonstrate that, surprisingly, such nonlinear behavior of coupled oscillators can be effectively linearized in certain latent dynamic spaces. The key insight is that there is a small number of latent dynamics filters, each with a specific association with synchronizing and non-synchronizing dynamics on subgraphs so that any observed dynamics on subgraphs can be approximated by a suitable linear combination of such elementary dynamic patterns. Taking an ensemble of subgraph-level predictions provides an interpretable predictor for whether the system on the whole graph reaches global synchronization. We propose algorithms based on supervised matrix factorization to learn such latent dynamics filters. We demonstrate that our method performs competitively in synchronization prediction tasks against baselines and black-box classification algorithms, despite its simple and interpretable architecture. This poster is associated with the minisymposium entitled Probabilistic Methods in Machine Learning and Complex Systems.

Agam Goyal
University of Wisconsin-Madison
agoyal25@wisc.edu

Zhaoxing Wu
University of Washington, Seattle
zwu363@uw.edu

Richard Yim
University of California, San Francisco
richard.yim@ucsf.edu

Binhao Chen
Brown University
binhao_chen@brown.edu

Zihong Xu
University of Wisconsin-Madison
zxu455@wisc.edu

Hanbaek Lyu
University of Wisconsin Madison
hlyu@math.wisc.edu

PP1

MS3: Lipschitz Regularized Gradient Flows and Latent Generative Particles

We developed a generative particle algorithm(GPA) that solves particle systems for Wasserstein gradient flows minimizing a regularized variant of f-divergences which are called Lipschitz regularized f-divergences. The Lipschitz regularized f-divergence is formulated as an f-divergence infimally convolved with Wasserstein-1 proximal, forming a robust framework. In our algorithm the f-divergence is estimated only using samples by solving its dual formulation in a space of Lipschitz continuous neural network functions. The optimizer function equals the first variation of the Lipschitz regularized f-divergence, and therefore its gradient produces a vector field for particle dynamics. Lipschitz regularization imposes a speed limit on the vector field and enhances stability of the dynamics. We also show that a specific selection of f enables stable generation of heavy-tailed data distributions. As a generative model, GPA excels in generating samples from scarce target data and is scalable to high-dimensional problems such as MNIST image generation. Since GPA offers the freedom to select the initial distribution, we show an application of GPA to gene expression data integration.

Hyemin Gu
University of Massachusetts Amherst
gu@math.umass.edu

Panagiota Birmpa
Herriot-Watt University
P.Birmpa@hw.ac.uk

Yannis Pantazis
Foundation for Research and Technology - Hellas
Institute of Applied and Computational Mathematics
pantazis@iacm.forth.gr

Luc Rey-Bellet, Markos Katsoulakis
University of Massachusetts Amherst
luc@math.umass.edu, markos@umass.edu

PP1

Sequential Gradient Descent

In biological neural systems the concept of local interactions plays a crucial role. Inspired by the way biological neurons communicate primarily with their immediate neighbors, we propose a neural network training approach based on sequential parameter updates. Heres a concise overview: We begin by updating the parameters of the rightmost layer (typically the output layer). Subsequently, we move layer by layer toward the input layer. At each step, we re-evaluate the loss function and adjust the parameters of the current layer using (stochastic) gradient descent. This local context mirrors the biological principle of neurons being aware only of nearby influences. There are expected benefits from this approach, such as fine-grained control, interpretability, as the approach aligns with our understanding of biological neural systems and their local influence among layers, which can help capture local patterns. The challenge include delayed learning, as the updates may transfer slowly to early layers, computational overhead, with frequent loss function evaluations, and sensitivity to initialization, with poorly initialized layers impacting subsequent ones. Our goal is to compare this method with the usual stochastic gradient descent method for ANN and CNN.

Gianluca Guadagni
School of Data Science
University of Virginia
gg5d@virginia.edu

PP1

Adjoint-Free Operator Learning and Matrix Recovery

There is a mystery at the heart of operator learning: how can one recover a non-self-adjoint operator from data without probing the adjoint? Practical approaches to operator learning suggest that one can accurately recover an operator while only using data generated by the forward action of the operator, without access to the adjoint. However, while existing theory justifies low sample complexity in operator learning, it relies on access to the adjoint. We prove that without querying the adjoint, one can approximate a family of non-self-adjoint infinite-dimensional compact operators via projection onto a Fourier basis. We apply this result to Green's functions to derive the first rigorous adjoint-free sample complexity bound for elliptic PDE learning. Still, we also show that for some fundamental tasks, such as low-rank matrix approximation and norm estimation, sampling the adjoint is essential.

Diana Halikias
Cornell University
dh736@cornell.edu

PP1

MS12: A Compositional Framework for First-Order Optimization (Poster)

Optimization decomposition methods are a fundamental tool to develop distributed solution algorithms for large scale optimization problems arising in fields such as machine learning and optimal control. In this poster, we present an algebraic framework for hierarchically composing optimization problems defined on hypergraphs and automatically generating distributed solution algorithms that respect the given hierarchical structure. We present an implementation of this framework in a Julia package called AlgebraicOptimization.jl and use our implementation to empirically demonstrate that hierarchical dual decomposition outperforms standard dual decomposition on problems

with hierarchical structure.

Tyler Hanks
University of Florida
GATAS Lab
t.hanks@ufl.edu

Matthew Klawonn
Air Force Research Lab
Information Directorate
matthew.klawonn.2@us.af.mil

Evan Patterson
Topos Institute
evan@epatters.org

Matthew Hale
Georgia Institute of Technology
mhale30@gatech.edu

James P. Fairbanks
University of Florida
fairbanksj@ufl.edu

PP1

MS1: Variational Structure-Preserving Neural Networks for Forced Lagrangian Systems

Structure preserving machine learning has recently become an active area of research. Popular models in this field, such as Hamiltonian neural networks, would typically require data on the system's momentum and this can be a limitation of the approach. Instead, we consider a method for learning the differential equations describing the dynamics of a forced Lagrangian system. The method requires time-series measurements of the system's position only and can learn external forces, e.g., dissipative frictional forces. We investigate how learning external forces impact the training of the network, and discuss regularization approaches.

Martine Dyring Hansen
Norwegian University of Science and Technology
martindh@stud.ntnu.no

PP1

MS9: Dataset Design to Accelerate Training Process-Structure Surrogate Models of Microstructure Evolution

Surrogate models for the evolution of material microstructures spatial fields that characterize the material are needed to perform important engineering tasks like manufacturing optimization. However, traditional strategies face serious bottlenecks in collecting sufficient training data, primarily relying on simulating entire evolution pathways. Computational costs restrict the total number of simulations, producing highly correlated and redundant datasets. In this study, we propose a novel approach that integrates microstructure generation protocols with active learning (AL) to address these challenges: (1) AL identifies points in the microstructure space of high model uncertainty, (2) representative microstructures are generated, (3) microstructure evolution simulations produce multiple frames of interest, (4) the model is trained with new data, and (5) the process repeats until a robust model is achieved. By incorporating microstructure generation into the AL loop, our approach enables simulations to start from diverse initial conditions, enhancing the robustness of surrogate models and eliminating the need for exhaustive dataset generation. Our research aims to provide a more efficient and reliable solution for surrogate modeling of microstructure evolution simulation tools in the field of materials science.

Grayson Harrington
Georgia Institute of Technology
grayson.harrington@gatech.edu

Surya Kalidindi
Georgia Institute of Technology
Materials Informatics for Engineering Design
surya.kalidindi@me.gatech.edu

PP1

MS6: Curvature Estimation for Point Clouds, Graphs, and Finite Metric Spaces

Real-world data sets frequently turn out to have low-dimensional manifold structure that can be studied using ideas from differential geometry. In particular, the curvature of a manifold is an important invariant that characterizes the extent to which the manifold deviates from being flat. We introduce an estimator for the scalar curvature of a data set presented as a finite metric space (e.g., a distance matrix, a point cloud, or a graph with the shortest-path metric). Our estimator depends only on the metric structure of the data (not on an embedding in Euclidean space), and it converges to the ground-truth scalar curvature as the number of points increases.

Abigail Hickok
Columbia University
aghickok@gmail.com

PP1

MS9: Mbius Inversion Meets Tensors: Inference by Edgeworth Series

Spatial statisticians, physicists, and data scientists all make Gaussian approximations of non-Gaussian phenomena of interest. Statistical inference in Gaussian random variables is highly tractable as operations like conditioning, evaluation of the (log-)likelihood, and sampling all involve numerical computation with the covariance matrix which can be accelerated with the Cholesky factor. We generalize these results to non-Gaussian random variables through the formalism of measure transport. The Cholesky factor becomes the Knothe-Rosenblatt rearrangement, a unique transport map whose Jacobian is lower triangular with positive diagonal. In the non-Gaussian case, we cannot hope to match only the mean and covariance, as in Gaussians. Instead, we will use higher-order cumulant tensors to characterize probability distributions, which can be estimated directly from samples. Expansions based on cumulants such as the Edgeworth series and Cornish-Fisher expansion give simple polynomial expressions for approximate densities, log-densities, and transport maps through Mbius inversion. We can also estimate functionals of the distribution such as entropy directly from samples.

Stephen Huan
Georgia Tech
shuan@gatech.edu

Andreas Robertson
Georgia Institute of Technology

arobertson38

Youssef M. Marzouk
Massachusetts Institute of Technology
ymarz@mit.edu

Florian Schaefer
Georgia Institute of Technology
florian.schaefer@cc.gatech.edu

PP1

MS19: 3D Manifold Topology Based Medical Image Data Augmentation

Data augmentation is an effective and universal technique for improving the generalization performance of deep neural networks. Current data augmentation implementations usually involve geometric and photometric transformations. However, none of them considers the topological information in images, which is an important global invariant of the three-dimensional manifold. In our implementation, we design a novel method that finds the generator of the first homology group, i.e. closed loops cannot shrink to a point, of 3D image and erases the bounding box of a random loop. To the best of our knowledge, it is the first time that data augmentation based on the first homology group of the three-dimensional image is applied in medical image augmentation. Our numerical experiments demonstrate that the proposed approach outperforms the state-of-the-art method.

Jisui Huang
Capital Normal University, China
University of Liverpool, UK
jisui@liverpool.ac.uk

Ke Chen
University of Strathclyde
k.chen@strath.ac.uk

Na Lei
Dalian University of Technology
nalei@dlut.edu.cn

PP1

MS9: Virtual X-Ray Diffraction Signals in Simulations of Dislocated Crystals: a Data Generation Pipeline

The relationship between dislocationsthe line defects responsible for plastic deformation in metalsand the macroscopic behavior of a material is complex. While machine learning would lend itself to identifying relationships between the dislocation configuration and macroscopic responses of the crystal, this would require three components: 1) an output macroscopic response which is determined by the dislocation configuration, 2) an input metric of the dislocation configuration which is in some way causally linked to this macroscopic response, 3) a training dataset encompassing a variety of dislocation configurations and their respective macroscopic response. An existing library of line bundle dislocation dynamics simulation could serve as an appropriate dataset if paired with a post-processing pipeline which could extract from this data the macroscopic response and dislocation configuration metrics. This poster will present one such pipeline. Simulation of the virtual response of the simulated domain in an X-ray diffraction experiment will serve as the macroscopic response, with the two-point statistic of the Krner-Nye tensor (a dislocation density) serving as the input metric. Typical time series behavior of both will be discussed.

Joseph P. Anderson, Hiroki Imura
Carthage College
Department of Physics and Astronomy
janderson19@carthage.edu, himura@carthage.edu

Khaled SharafEldin
Purdue University
School of Materials Engineering
kabdelaz@purdue.edu

Anter El-Azab
Purdue University
aelazab@purdue.edu

PP1

Parametric SympNets

The numerical integration of Hamiltonian systems is a central topic in computational physics and theoretical chemistry. SympNets [P. Jin, Z. Zhang, A.Zhu, Y. Tang, and G. E. Karniadakis., SympNets: intrinsic structure-preserving symplectic networks for identifying Hamiltonian system, Neural Networks, 132:166179, dec 2020] are a widely used tool when it comes to learning symplectic integrators for Hamiltonian systems from data. More often then not Hamiltonian systems of interest depend on a set of parameters. Currently, SympNets do not respect parameters. Therefore, SympNets have to be retrained for every new parameter combination. To eliminate this necessity, we propose to extend the SympNet architecture to handle parameters explicitly, making it suitable for parametric Hamiltonian systems. This approach is coined "ParSympNets". We perform different numerical experiments to investigate the approximation and generalization capabilities of ParSympNets. Furthermore, we present a new universal approximation theorem, which shows that ParSympNets can approximate arbitrary parametric symplectic maps.

Konrad Janik
Max Planck Institute for Dynamics of Complex Technical Systems
janik@mpi-magdeburg.mpg.de

Peter Benner
Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
benner@mpi-magdeburg.mpg.de

PP1

MS16: Learning Differential Equations from Data: Evaluating the Occupation Kernel Algorithm Using the Common Task Framework of the AI Institute for Dynamical Systems

As essential tools of statistical learning theory, kernel methods exploit the power of reproducing kernel Hilbert spaces (RKHS). Kernel methods excel in modeling nonlinear relationships by implicitly mapping data into high-dimensional or infinite feature spaces. This work employs occupation kernel (OCK) functions to learn vector fields driving ordinary differential equations (ODEs). The OCK algorithm captures snapshots of ODE trajectories as data-points, recovering the vector field governing the ODE and employing penalized least squares. The solutions

are expressed using OCK functions and RK4 numerical integration. We evaluate the OCK algorithm using the Common Task Framework (CTF) of the AI Institute for Dynamical Systems [1], with model training on six datasets using various Matérn kernels. Bayesian optimization optimizes hyper-parameters before employing the OCK algorithm for prediction. Performance is assessed using the CTF forecasting and reconstruction scores. The OCK algorithm performs competitively in other contexts [2], and the CTF allows for further evaluation of this algorithm. [1] https://maths4dl.ac.uk/wp-content/uploads/2024/04/Kutz_CTF.pdf [2] Lahouel, K., Wells, M., Rielly, V., Lew, E., Lovitz, D., and Jedynak, B.M., 2024. Learning nonparametric ordinary differential equations from noisy data. Journal of Computational Physics, p.112971

Bruno Jedynak, Colin Gibson, Anthony Kolshorn
Portland State University
bruno.jedynak@pdx.edu, gibson25@pdx.edu, kolshorn@pdx.edu

Naganda Kyatsandra Gurukumar
Portland State University, U.S.
kyat@pdx.edu

Chau Nguyen, Arielle Soumare
Portland State University
min25@pdx.edu, soumare@pdx.edu

PP1

MS9: Higher Order Cumulant Truncation for Improved Moment Closure

Moment methods are extensively used to compute the evolution of probability distributions over time. These methods can involve nonlinear transformations where changes in lower order moments depend on higher order moments, resulting in an unclosed system. As a result, a method to express higher order moments in terms of lower order ones is required for system closure. A common closure method is the Gaussian, which approximates higher order moments using only in terms of the first and second cumulants, disregarding other relevant higher order information that could have been used. To address this, we propose higher order cumulant truncation methods that incorporate higher order information. Although only the Gaussian possess a finite number of non-zero cumulants, our approximation leverages the property of higher order cumulants decaying faster than lower order ones, suggesting its potential effectiveness. To obtain moments that correspond to feasible probability distributions, we project our higher order moment approximations onto the moments obtained from the set of feasible probability distributions using the set of polynomial sum-of-squares. We tested our methods through function approximations on various distributions, simulated particle collisions, moment closure applications on the Duffing oscillator, and moment closure applications on microbial consortium dynamics. We have identified situations where higher order truncation performs better than the Gaussian.

Shreya Jha
Georgia Institute of Technology
sjha75@gatech.edu

PP1

MS14: Training Neural Operators to Preserve Invariant Measures of Chaotic Attractors

Chaotic systems make long-horizon forecasts difficult because small perturbations in initial conditions cause trajectories to diverge at an exponential rate. In this setting, neural operators trained to minimize squared error losses, while capable of accurate short-term forecasts, often fail to reproduce statistical or structural properties of the dynamics over longer time horizons and can yield degenerate results. We propose an alternative framework designed to preserve invariant measures of chaotic attractors that characterize the time-invariant statistical properties of the dynamics. Specifically, in the multi-environment setting (where each sample trajectory is governed by slightly different dynamics), we consider two novel approaches to training with noisy data. First, we propose a loss based on the optimal transport distance between the observed dynamics and the neural operator outputs. This approach requires expert knowledge of the underlying physics to determine what statistical features should be included in the optimal transport loss. Second, we show that a contrastive learning framework, which does not require any specialized prior knowledge, can preserve statistical properties of the dynamics nearly as well as the optimal transport approach. In various chaotic systems, our method is shown empirically to preserve invariant measures of chaotic attractors. This poster is associated with the minisymposium Probabilistic Methods in Machine Learning and Complex Systems.

Ruoxi Jiang
the University of Chicago
roxie62@uchicago.edu

Peter Lu, Elena Orlova, Rebecca Willett
The University of Chicago
lup@uchicago.edu, eorlova@uchicago.edu, willett@uchicago.edu

PP1

MS9: Hybrid Data-Driven Solvers for Statistical Continuum Mechanics

In this work we propose a new class of data-driven models which predict local deformations over heterogeneous material microstructures. This problem – termed the localization problem – is a core component of numerous open challenges in multiscale materials design. The inherent stochasticity of manufacturing means that constructing forward (process -¿ property) or inverse (property -¿ process) UQ models requires solving a variable-coefficient PDE repeatedly for many microstructure instantiations. Moreover, the discontinuity and disorder of the PDE coefficients leads to a high-dimensional, poorly-conditioned system of equations. Our work combines traditional numerical solvers and data-driven methods to construct a hybrid approximation for the coefficient-to-solution map. In particular, we utilize the Lippmann-Schwinger formulation of localization to guide the design of a thermodynamically-informed implicit (Deep Equilibrium) neural operator. Applied to both two-phase composites and polycrystalline materials, our methodology shows improved stability and extrapolation capabilities compared to existing machine learning methods. Moreover, we obtain an improved error-accuracy tradeoff for both local and homogenized properties compared to both traditional solvers and feedforward ML methods. Finally, we find that embedding thermodynamic encodings into the architecture provides improved data efficiency and gener-

alizability.

Conlain Kelly
Georgia Tech
ckelly84@gatech.edu

Surya Kalidindi
Georgia Institute of Technology
Materials Informatics for Engineering Design
surya.kalidindi@me.gatech.edu

PP1

MS12: Mining Sheaf Theoretic Narratives (Poster)

Querying time-series data or finding motifs therein is well understood for certain data types but remains challenging in the general case. For various data types or patterns sought, it is sometimes unclear what the correct definition of a temporal pattern is even if the analogous static pattern is well studied. One can look to various theories of temporal graphs, for instance, to see how time adds nuances and complexity to the study of common graph theoretic artifacts like paths. In this talk we recapitulate recent work done by collaborators to define temporal data via a sheaf theoretic construct that is agnostic to the object of study, captures richer structure than mere sequences of data, and leads to natural temporal versions of static patterns, among other benefits. We further show how this sheaf theoretic definition of temporal data can be leveraged for finding/counting occurrences of temporal patterns within a time-series by way of a homomorphism search algorithm. We will discuss empirical results of experiments grounded in Air Force planning applications.

Matthew Klawonn
Air Force Research Lab
Information Directorate
matthew.klawonn.2@us.af.mil

PP1

MS7: Data-Driven Estimation of Stability Guarantees for Nonlinear Dynamical Systems (Poster)

Analyzing the stability of a nonlinear dynamical system is central to understanding system behavior and designing controllers, and we use a Lyapunov function for this purpose. It is possible to guarantee the stability of a system if one can find a Lyapunov function that is positive definite and decreasing over time along the orbit of the system, thus providing a sufficient condition for stability. A Lyapunov function also characterizes an estimate of the domain of attraction, which indicates the region under which the system states asymptotically converge to equilibrium. The construction of a Lyapunov function is done analytically and ad hoc for certain nonlinear systems. However, doing so for systems with high nonlinearities and different dimensions is a challenging task. To address this problem, we present a data-driven method for discovering Lyapunov functions, called Lyapunov function inference (LyapInf). This new method fits a quadratic Lyapunov function to the state trajectory data of the dynamics via optimization, where the process of inferring a Lyapunov function is based on the non-intrusive model reduction method of Operator Inference. This method learns one of many possible Lyapunov functions that ensures stability and estimates the domain of attraction with or without access to the system model. In this work, we demonstrate this new method on several numerical examples.

Tomoki Koike, Elizabeth Qian
Georgia Institute of Technology
tkoike3@gatech.edu, eqian@gatech.edu

PP1

MS4: Approximation of Group Explainers with Coalition Structure Using Monte Carlo Sampling

In recent years, many Machine Learning (ML) explanation techniques have been designed using ideas from cooperative game theory. These game-theoretic explainers suffer from high complexity, hindering their exact computation in practical settings. In our work, we focus on a wide class of linear game values, as well as coalitional values, for the marginal game based on a given ML model and predictor vector. By viewing these explainers as expectations over appropriate sample spaces, we design a novel Monte Carlo sampling algorithm that estimates them at a reduced complexity that depends linearly on the size of the background dataset. The advantage of this approach is that it is fast, easily implementable, and model-agnostic.

Konstandinos Kotsiopoulos
Discover Financial Services
kkotsiop@gmail.com

Alexey Miroshnikov
Discover Financial Services
U.S
amiroshn@terpmail.umd.edu

Khashayar Filom, Arjun Ravi Kannan
Discover Financial Services
khashayar.1367@gmail.com, arjun.kannan@gmail.com

PP1

MS2: Modeling Time-Varying Networks

Modeling dynamic networks over time is a crucial task in understanding how systems change. Previous research often relied on either parametric models or nonparametric models that assume stationarity, both of which limit flexibility in modeling. In this work, we introduce a smooth time-varying dynamic graphon model and propose a method to fit this model using double smoothing. Particularly, we address the problem of change point detection without assuming stationarity, significantly expanding the applicability of previous methods in network settings. We demonstrate the effectiveness of our approach through theoretical results and extensive numerical studies.

Jeonghwan Lee
University of Minnesota
junghwn1408@gmail.com

PP1

MS10: Flow-Based Distributionally Robust Optimization

We present a computationally efficient framework, called `FlowDRO`, for solving flow-based distributionally robust optimization (DRO) problems with Wasserstein uncertainty sets while aiming to find continuous worst-case distribution (also called the Least Favorable Distribution, LFD) and sample from it. The requirement for LFD to be continuous is so that the algorithm can be scalable to prob-

lems with larger sample sizes and achieve better generalization capability for the induced robust algorithms. To tackle the computationally challenging infinitely dimensional optimization problem, we leverage flow-based models and continuous-time invertible transport maps between the data distribution and the target distribution and develop a Wasserstein proximal gradient flow type algorithm. In theory, we establish the equivalence of the solution by optimal transport map to the original formulation, as well as the dual form of the problem through Wasserstein calculus and Brenier theorem. In practice, we parameterize the transport maps by a sequence of neural networks progressively trained in blocks by gradient descent. We demonstrate its usage in adversarial learning, distributionally robust hypothesis testing, and a new mechanism for data-driven distribution perturbation differential privacy, where the proposed method gives strong empirical performance on high-dimensional real data.

Chen Xu
Georgia Institute of technology
cx9711@gatech.edu

Jonghyeok Lee
Georgia Institute of Technology
jlee4177@gatech.edu

Xiuyuan Cheng
Duke University
xiuyuan.cheng@duke.edu

Yao Xie
Georgia Institute of Technology
yao.xie@isye.gatech.edu

PP1

MS19: Discrete Ricci Curvature Based Level Set Model

The level set method has played a critical role among many image segmentation approaches. Several edge detectors, such as the gradient, have been applied to its regularisation term. However, traditional edge detectors lack high-order information and are sensitive to image noise. To tackle this problem, we introduce a method to calculate the Ricci curvature, a vital curvature in three-dimensional Riemannian geometry. In addition, we propose incorporating the curvature into the regularisation term. Experiments suggest that our method outperforms the state-of-the-art level set methods and achieves a comparable result with deep learning.

Jisui Huang
Capital Normal University, China
University of Liverpool, UK
jisui@liverpool.ac.uk

Ke Chen
University of Strathclyde
k.chen@strath.ac.uk

Emil Saucan
ORT Braude College, Israel
semil@braude.ac.il

Na Lei
Dalian University of Technology
nalei@dlut.edu.cn

PP1

MS14: Stochastic Block Majorization-Minimization Methods on Riemannian Manifolds

Block majorization-minimization (BMM) is a simple iterative algorithm for nonconvex optimization that sequentially minimizes a majorizing surrogate of the objective function in each block coordinate while the other block coordinates are held fixed. Recently, the Riemannian counterpart of BMM has been studied, where each block is a Riemannian manifold. In this project, we consider an online version of Riemannian BMM, which proceeds by sampling new data points and recursively minimizing the empirical loss function. We analyze the convergence and complexity of the proposed algorithm and provide numerical experiments to validate its effectiveness. This poster is associated with the mini-symposium entitled Probabilistic Methods in Machine Learning and Complex Systems.

Yuchen Li
University of Wisconsin-Madison
yli966@wisc.edu

Hanbaek Lyu
University of Wisconsin Madison
hlyu@math.wisc.edu

PP1

MS10: Decoupled Data Consistency with Diffusion Purification for Image Restoration

Diffusion models have recently gained traction as a powerful class of deep generative priors, excelling in a wide range of image restoration tasks due to their exceptional ability to model data distributions. To solve image restoration problems, many existing techniques achieve data consistency by incorporating additional likelihood gradient steps into the reverse sampling process of diffusion models. However, the additional gradient steps pose a challenge for real-world practical applications as they incur a large computational overhead, thereby increasing inference time. They also present additional difficulties when using accelerated diffusion model samplers, as the number of data consistency steps is limited by the number of reverse sampling steps. In this work, we propose a novel diffusion-based image restoration solver that addresses these issues by decoupling the reverse process from the data consistency steps. Our method involves alternating between a reconstruction phase to maintain data consistency and a refinement phase that enforces the prior via diffusion purification. Our approach demonstrates versatility, making it highly adaptable for efficient problem-solving in latent space. Additionally, it reduces the necessity for numerous sampling steps through the integration of consistency models.

Xiang Li
University of Michigan
forkobe@umich.edu

PP1

MS2: Improved Curvature Gaps for Stochastic Block Model Graphs

Discrete notions of curvature have recently been used to study the structural properties of networks from a new perspective, modeling properties such as geodesic disper-

sion and volume growth that are not directly captured by spectral analysis. Clustering algorithms based on graph curvatures rely on the property that edges within communities tend to be positively curved, while bridges between communities tend to be negatively curved. In the current work, we improve bounds on the "curvature gap", which quantifies the separation between these two edge types, by analyzing the distributional limits of Ollivier's Ricci curvature in stochastic blockmodel graphs for these two kinds of edges.

Chang Li
University of Virginia
xue6bq@virginia.edu

PP1

MS11: Exploiting Low-Dimensional Data Structures in Deep Learning

In the past decade, deep learning has made astonishing breakthroughs in various real-world applications. It is a common belief that deep neural networks are good at learning various geometric structures hidden in data sets. One of the central interests in deep learning theory is to understand why deep neural networks are successful, and how they utilize low-dimensional data structures. In this talk, I will present some statistical learning theory of deep neural networks where data are concentrated on or near a low-dimensional manifold. The learning tasks include regression, classification, feature representation and operator learning. When data are sampled on a low-dimensional manifold, the sample complexity crucially depends on the intrinsic dimension of the manifold instead of the ambient dimension of the data. These results demonstrate that deep neural networks are adaptive to low-dimensional geometric structures of data sets.

Wenjing Liao
Georgia Institute of Technology
wliao60@gatech.edu

PP1

MS3: Wasserstein Proximal Operators Describe Score-Based Generative Models and Resolve Memorization

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.

Siting Liu
University of California, Los Angeles, U.S.
siting6@math.ucla.edu

Benjamin J. Zhang
University of Massachusetts Amherst
bjzhang@umass.edu

Wuchen Li
University of South Carolina
wuchen@mailbox.sc.edu

Markos Katsoulakis
University of Massachusetts Amherst
markos@umass.edu

Stanley J. Osher
UCLA
sjo@math.ucla.edu

PP1

MS20: Associated Poster: Towards Constituting Mathematical Structures for Learning to Optimize

Learning to Optimize (L2O), a technique that utilizes machine learning to learn an optimization algorithm automatically from data, has gained arising attention in recent years. A generic L2O approach parameterizes the iterative update rule and learns the update direction as a black-box network. While the generic approach is widely applicable, the learned model can overfit and may not generalize well to out-of-distribution test sets. In this paper, we derive the basic mathematical conditions that successful update rules commonly satisfy. Consequently, we propose a novel L2O model with a mathematics-inspired structure that is broadly applicable and generalized well to out-of-distribution problems. Numerical simulations validate our theoretical findings and demonstrate the superior empirical performance of the proposed L2O model.

Jialin Liu
University of Central Florida
jialin.liu@ucf.edu

PP1

On a Latent-Variable Formulation of the Poisson Canonical Polyadic Tensor Model

The Poisson Canonical Polyadic (PCP) tensor model is a popular tool for the analysis of high-dimensional tensors composed elementwise of counts, but many of its statistical properties are unknown due to a loglikelihood that is challenging to differentiate. We revisit the PCP tensor model as a latent variable model, and leverage the missing information principle to obtain the Fisher information matrix, which can be used to (1) gauge the well-posedness or identifiability of the model, (2) derive second-order Fisher scoring algorithms for maximum likelihood estimation and to (3) derive Cramer-Rao lower bounds on the variance of unbiased estimators. We also re-discover existing nonnegative matrix and tensor factorization algorithms within our latent-variable framework as expectation-maximization algorithms, and study our derived properties in extensive

simulation studies.

Carlos Llosa
Sandia National Laboratories
cjllosa@sandia.gov

PP1

The Average Spectrum Norm for Tensor Completion Analysis

We introduce the average spectrum norm to study sample complexity of tensor completion problems based on the canonical polyadic decomposition (CPD). Our novel approach significantly reduces the provable sample rate for CPD-based noisy tensor completion, providing the best bounds to date on the number of observed noisy entries required to produce an arbitrarily accurate estimate of an underlying mean value tensor. Under Poisson and Bernoulli multivariate distributions, we show that an N-way CPD rank-R parametric tensor $\boldsymbol{M} \in \mathbb{R}^{I\times\cdots\times I}$ generating noisy observations can be approximated by large likelihood estimators from $\mathcal{O}(IR^2\log^{N+2}(I))$ revealed entries. Furthermore, under nonnegative and orthogonal versions of the CPD we improve the result to depend linearly on the rank, achieving the near-optimal rate $\mathcal{O}(IR\log^{N+2}(I))$.

Oscar Lopez
Harbor Branch Oceanographic Institute
Florida Atlantic University
lopezo@fau.edu

Richard B. Lehoucq, Carlos Llosa, Daniel M. Dunlavy, Arvind Prasadan
Sandia National Laboratories
rblehou@sandia.gov, cjllosa@sandia.gov, dmdunla@sandia.gov, aprasad@sandia.gov

PP1

MS5: Dimension Reduction for Derivative Informed Operator Learning

Neural operators—neural network approximations of mappings between function spaces—have shown promise as computationally efficient approximations of PDE solution operators, and can be deployed as surrogates in many query-intensive tasks such as optimization and uncertainty quantification, bayesian inverse problems, optimal experimental design. Conventional operator learning typically focuses on the accuracy of the evaluations, i.e. accuracy of the predicted PDE solutions over some input distribution. However, the derivative of the solution operator also plays an important role in the accuracy and efficiency of algorithms for many of the tasks described above. In this work, we present theoretical analysis and universal approximation results for neural operator in Sobolev classes over Gaussian measures. That is, we quantify both the function and derivative approximation errors for input-output maps between infinite dimensional spaces. Specifically, we consider reduced basis architectures, and how compare different choices of dimension reduction, including principal component analysis as well as derivative informed dimension reduction strategies, affect the approximation errors.

Dingcheng Luo
Oden Institute for Computational Engineering and Sciences,
The University of Texas at Austin
dc.luo@utexas.edu

Thomas O'Leary-Roseberry
The University of Texas at Austin
Oden Institute for Computational and Engineering Sciences
tom@oden.utexas.edu

Peng Chen
Georgia Institute of Technology
pchen402@gatech.edu

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

PP1

MS13: Iterative Approaches for Tensor Linear Systems

In this poster presentation, we introduce an iterative method for approximating the solution of large-scale multilinear systems, represented in the form A*X=B under the tensor t-product. Unlike previously proposed randomized iterative strategies, such as the tensor randomized Kaczmarz method (row slice sketching) or the tensor Gauss-Seidel method (column slice sketching), which are natural extensions of their matrix counterparts, our approach delves into a distinct scenario utilizing frontal slice sketching. In particular, we explore a context where frontal slices, such as video frames, arrive sequentially over time, and access to only one frontal slice at any given moment is available. We will present our novel approach, shedding light on its applicability and potential benefits in approximating solutions to large-scale multi-linear systems.

Anna Ma
UC Irvine
anna.ma@uci.edu

Hengrui Luo
Rice University
hrluo@rice.edu

PP1

MS19: Medical Image Segmentation Using the Equivariance under Diffeomorphisms Framework

Medical image segmentation plays a crucial role in medical image analysis, computer-aided detection and diagnosis, treatment planning, and etc. However, it is still challenging to obtain an accurate segmentation result due to irregularity of organ contours or lack of labeled dataset. In this work, we formulate an important property for image segmentation: equivariance under diffeomorphisms, that is, the segmentation results are independent of the small diffeomorphic deformations. Based on this property, we propose a novel Equivariance under Diffeomorphism (ED) framework for medical image segmentation using the optimal transport maps. Experiments are carried out to evaluate the proposed method on two publicly available datasets, including a colon dataset and a hepatic vessels dataset. Results show that the proposed method outperforms the existing method in terms of two metrics Jaccard and Dice, respectively.

Jisui Huang
Capital Normal University, China
University of Liverpool, UK
jisui@liverpool.ac.uk

David Gu
State University of New York at Stony Brook
gu@cs.sunysb.edu

Ming Ma
Winona State University in Minnesota
ming.ma@winona.edu

Na Lei
Dalian University of Technology
nalei@dlut.edu.cn

PP1

An Online, Sampling-Based Convergence Criterion for Symmetric Matrices

When a matrix is streamed or if both its dimensions are enormous, it may not be possible to store the entire matrix in memory. When dimensionality reduction is required, Incremental SVD (iSVD) circumvents the need for costly data storage. However, computing residual information that is necessary for early termination in iterative solvers requires access to all of the matrix, which is not possible. Additionally, we often observe empirically that the solver tolerance requested for each window does not guarantee that the residuals of the final solution will satisfy the same tolerance. In this work, we demonstrate how sampling the matrix at runtime can be used to accurately estimate the residual error that would be achievable if the matrix were fully available. With an estimate of the residual error, we develop a sampling-based convergence criterion for an iterative eigenmethod. We provide numerical evidence on challenging datasets, both dense and sparse, that our convergence criterion reduces cumulative costs as the matrix is streamed. We utilize this criterion only for symmetric matrices since the right singular vectors approximated by iSVD coincide with the eigenvectors of a symmetric matrix.

Jeremy Myers
The College of WIlliam and Mary
jermyer@sandia.gov

Andreas Stathopoulos
College of William & Mary
Department of Computer Science
andreas@cs.wm.edu

PP1

MS17: Conditional Generative Models in Data Assimilation

In this poster we consider the probabilistic data assimilation problem and describe how conditional generative models can be used to approximate the Bayes filter. The advantages of this approach include the ability to work with nonlinear dynamics, nonlinear measurement operators, and non-Gaussian approximations for model error and measurement noise.

Assad A. Oberai
University of Southern California
aoberai@usc.edu

Agnimitra Dasgupta
Department of Civil and Environmental Engineering
University of Southern California
adasgupt@usc.edu

Allen Umali, Bryan Shaddy
University of Southern California
aumali@usc.edu, bshaddy@usc.edu

Brianna Binder
University of Southern Carolina
bjbinder@usc.edu

PP1

MS8: Performance Guarantees for Solving Linear Inverse Problems with Implicit Neural Representations

Implicit neural representations (INRs) have emerged as a powerful tool inverse for solving inverse problems in computer vision and computational imaging. Rather than representing images as a discrete collection of pixels, INRs represent images as a continuous domain function via a neural network taking spatial coordinates as inputs. However, unlike discrete image representations, little is known about the sample complexity of estimating images using INRs in the context of linear inverse problems. Towards this end, we derive necessary and sufficient conditions under which an image is exactly recoverable from its low-pass Fourier coefficients when fitting a two-layer (i.e., single hidden-layer) INR with a Fourier features layer. In particular, we relate the sample complexity to the minimum effective width needed to realize the ground truth image as an INR.

Gregory Ongie
Marquette University
gregory.ongie@marquette.edu

PP1

MS14: Deep Stochastic Mechanics

This work introduces a novel deep-learning-based approach for numerical simulation of a time-evolving Schrdinger equation inspired by stochastic mechanics and generative diffusion models. Unlike existing approaches, which exhibit computational complexity that scales exponentially in the problem dimension, our method allows us to adapt to the latent low-dimensional structure of the wave function by sampling from the Markovian diffusion. Depending on the latent dimension, our method may have far lower computational complexity in higher dimensions. Numerical simulations verify our theoretical findings and show a significant advantage of our method compared to other deep-learning-based approaches used for quantum mechanics. This poster is associated with the minisymposium entitled Probabilistic Methods in Machine Learning and Complex Systems.

Elena Orlova
The University of Chicago
eorlova@uchicago.edu

Aleksei Ustimenko
ShareChat
research@aleksei.uk

Ruoxi Jiang
the University of Chicago
roxie62@uchicago.edu

Peter Lu, Rebecca Willett
The University of Chicago

lup@uchicago.edu, willett@uchicago.edu

PP1

Enhancing Covid-19 Endemic/Epidemic Forecasting Precision Through the Integration of Deterministic Model, Machine Learning, Daily Data, and Cycle Threshold Values

In this study, we propose an overview of the COVID-19 situation using both deterministic and machine learning models on the transmission dynamics of the COVID-19 pandemic. This is achieved by integrating a hybrid model with daily empirical data and cycle threshold (Ct) values from different countries. We partitioned this dataset into before and after vaccination to understand the influence of vaccination on disease dynamics. We used the deterministic model to present some mathematical analyses. Also, the model was fitted to real data to validate the deterministic model's precision. Furthermore, we used a machine learning model to validate our model by performing principal component analysis (PCA) to predict the evolution of the spread of the COVID-19 outbreak on some statistical predictor indicators from time series modeling on a 14-day moving window for detecting which of these indicators capture the dynamics of the disease spread across the epidemic curve. The results of the PCA, the calculated index of dispersion, the fitted deterministic model, and the deterministic model simulation are all in agreement with the dynamics of the disease before and after vaccination started. Conclusively, our approach has been able to capture the dynamics of the pandemic at different phases of the disease outbreak, and the result presented will be useful to understand the evolution of the disease and future and emerging epidemics.

Kayode Oshinubi
Université Grenoble Alpes
Université Grenoble Alpes
oshinubik@gmail.com

PP1

MS8: Effect of Linear Layers in ReLu Networks: Learning Single-/Multiple-Index Models

Why do deeper neural networks tend to outperform shallow ones? We study the inductive bias of neural networks in the simplified case where most layers have a linear activation. Despite these models having the same capacity at different depths, they do not all have the same representation cost. Specifically, minimizing the ℓ_2 penalty when training a neural network with many linear layers followed by a single ReLu layer using weight decay is equivalent to a function-space penalty that encourages the network to select a function with low mixed variation. That is, the function has limited variation in directions orthogonal to a low-dimensional subspace. This means that the trained model will approximately be a single- or multiple- index model. Our experiments show that when this active subspace structure exists in the data, adding linear layers can improve generalization and result in a network that is well-aligned with the true active subspace.

Suzanna Parkinson
University of Chicago
sueparkinson@uchicago.edu

PP1

MS12: The Synthetic Theory of Probability and Statistics

Probability theory and statistics are usually conceived analytically, grounded in the standard measure-theoretic foundation. In recent years, significant progress has been made toward a synthetic account of probability and statistics that treats many of the common definitions and theorems in an axiomatic and purely algebraic setting. The synthetic theory is based on Markov categories and other category-theoretic ideas. Besides widening the scope of probability theory to nonstandard settings, the synthetic theory can be used to give a formal semantics to probabilistic programs, a structuralist account of statistical models and the relationships between them, and a language for specifying statistical models richer than the familiar graphical models. In this talk, we give an introduction to the synthetic theory of probability and statistics.

Evan Patterson
Topos Institute
evan@epatters.org

PP1

Sparse Nystrom Approximation of Currents, Varifolds and Normal Cycles

Comparing and computing distances between shapes is a fundamental task in computer vision, geometric learning, and computational anatomy. When building models of shape variation, one requires a fidelity metric on shapes for model fitting. Ideally shape data is available in parametric correspondence. In applications, it is far from the case, with shape data acquired with inconsistent parametrizations and resolutions across a dataset. For submanifolds of $\mathbb{R}^d$ one may compare shapes in a geometric manner without correspondence, using concepts from geometric measure theory (GMT), such as currents, varifolds and normal-cycles. Shapes are embedded into the dual of a Hilbert Space of differential forms, and compared through the dual metric on this space, independent of parametrization. However, practical metric computation is costly. Comparison between triangulations with M, N triangles respectively, scales as $\mathcal{O}(MN)$; undesirable for practical statistical shape modelling (SSM) when $M, N > 10^5$. In this work, we derive a randomized algorithm for compression of the GMT representations of shapes, using the Nystrom approximation. Our method is significantly faster than existing compression methods and has theoretical convergence bounds. Post compression, one can compute distances and gradients in closed form between shapes in $\mathcal{O}(mn)$, with $m, n << \min(M, N)$. The compressions are shown to be useful for down-line geometric learning tasks.

Allen Paul, Tony Shardlow, Neill Campbell
University of Bath
ap2746@bath.ac.uk, t.shardlow@bath.ac.uk, n.campbell@bath.ac.uk

PP1

MS4: Continuum Limits for Dimension Reduction

Dimension reduction is a common preprocessing step in many data science pipelines, which enables data visualization, feature construction, and greater intuition and interpretability of datasets. However, basic properties of many dimension reduction algorithms, such as statistical consistency, efficiency of parametrization, and structural properties of the learned embeddings, are not well-understood. This presentation discusses basic properties

of the population-level dimension reduction problem from a variational viewpoint: results include Monge-type existence results, parametric representation results, and a counterintuitive result about the fact that in several natural cases the learned embeddings always contain discontinuities. These results raise questions about inherent biases in dimension reduction techniques, and in particular about the validity of the scientific conclusions drawn from interpretations of cluster plots in applications. For example, in genomics, artificial clustering could erroneously separate essential genetic subpopulations into different groups, potentially resulting in ineffective or even harmful treatments.

Adam Pickarski, Ryan Murray
North Carolina State University
appickar@ncsu.edu, rwmurray@ncsu.edu

PP1

MS6: Spectral Families of Persistent Rank Invariants

Using the fact that the persistent rank invariant determines the persistence diagram and vice versa, we introduce a framework for constructing families of continuous relaxations of the persistent rank invariant for persistence modules indexed over the real line. Like the rank invariant, these families obey inclusion-exclusion, are derived from simplicial boundary operators, and encode all the information needed to construct a persistence diagram. Unlike the rank invariant, these spectrally-derived families enjoy a number of stability and continuity properties typically reserved for persistence diagrams, such as smoothness and differentiability. By leveraging its relationship with combinatorial Laplacian operators, we find the non-harmonic spectra of our proposed relaxation encode valuable geometric information about the underlying space, prompting several avenues for geometric data analysis. Moreover, as trace-class operators, we also find these Laplacian operators can be efficiently approximated using just O(n) space with a randomized algorithm based on the stochastic Lanczos quadrature method, suggesting the potential for improving the scalability of the persistence computation. We investigate the utility of our relaxation with applications in topological data analysis and machine learning, such as parameter optimization, shape classification, and mesh simplification.

Matt Piekenbrock
Northeastern University
piekenbrock.m@northeastern.edu

Jose Perea
Northeastern University, U.S.
j.pereabenitez@northeastern.edu

PP1

MS14: Stochastic Optimization with Arbitrary Recurrent Data Sampling

For obtaining optimal first-order convergence guarantees for stochastic optimization, it is necessary to use a recurrent data sampling algorithm that samples every data point with sufficient frequency. Most commonly used data sampling algorithms (e.g., i.i.d., MCMC, random reshuffling) are indeed recurrent under mild assumptions. In this work, we show that for a particular class of stochastic optimization algorithms, we do not need any further property (e.g., independence, exponential mixing, and reshuffling) beyond recurrence in data sampling to guarantee optimal rate of first-order convergence. Namely, using regularized versions of Minimization by Incremental Surrogate Optimization (MISO), we show that for non-convex and possibly non-smooth objective functions with constraints, the expected optimality gap converges at an optimal rate $O(n^{-1/2})$ under general recurrent sampling schemes. Furthermore, the implied constant depends explicitly on the 'speed of recurrence', measured by the expected amount of time to visit a data point, either averaged ('target time') or supremized ('hitting time') over the target locations. We discuss applications of our general framework to decentralized optimization and distributed non-negative matrix factorization. This poster is associated with the section "Efficient and robust optimization techniques for structured data learning" (80608).

William Powell
University of Wisconsin-Madison
Department of Mathematics
wgpowell@wisc.edu

PP1

MS15: Conformal Validity Guarantees Exist for Any Data Distribution (and How to Find Them)

As artificial intelligence (AI) / machine learning (ML) gain widespread adoption, practitioners are increasingly seeking means to quantify and control the risk these systems incur. This challenge is especially salient when such systems have autonomy to collect their own data, such as in black-box optimization and active learning, where their actions induce sequential feedback-loop shifts in the data distribution. Conformal prediction is a promising approach to uncertainty and risk quantification, but prior variants' validity guarantees have assumed some form of "quasi-exchangeability' on the data distribution, thereby excluding many types of sequential shifts. In this paper we prove that conformal prediction can theoretically be extended to *any* joint data distribution, not just exchangeable or quasi-exchangeable ones. Although the most general case is exceedingly impractical to compute, for concrete practical applications we outline a procedure for deriving specific conformal algorithms for any data distribution, and we use this procedure to derive tractable algorithms for a series of AI/ML-agent-induced covariate shifts. We evaluate the proposed algorithms empirically on synthetic black-box optimization and active learning tasks.

Drew Prinster
Johns Hopkins University, U.S.
aprinst1@jhu.edu

PP1

MS13: Scalable Sketching Through Shared Preconditioning

Perhaps no computational problem is more ubiquitous in statistics than that of the least squares problem. In recent years, the increasing scale of datasets has caused computational bottlenecks that classical solvers are ill-suited to overcome. Thus, new methods are needed to overcome these challenges. Iterative Right Random Sketching (IRRS) is one promising technique that iteratively generates updates to a solution using samples of a matrix's columns. These samples limit IRRS's memory load and allows it to solve terabyte-size problems under tight computational constraints. Unfortunately, naive progress tracking methods, like computing the residual of the normal

equations, undermine the benefits of IRRS. In this talk, we will introduce and quantify the uncertainty of a procedure that tracks the progress of IRRS using only the matrix samples already required to update the solution. Further, we use the uncertainty quantification to derive risk-aware stopping criteria, which we integrate into the Gauss-Newton to fit logistic regression models.

Nathaniel Pritchard
University of Wisconsin-Madison
npritchard@wisc.edu

PP1

MS8: Mechanism for Feature Learning in Neural Networks and Backpropagation-Free Machine Learning Models

Understanding how neural networks learn features, or relevant patterns in data, for prediction is necessary for their reliable use in technological and scientific applications. We propose a unifying mechanism that characterizes feature learning in neural network architectures. Namely, we show that features learned by neural networks are captured by a statistical operator known as the average gradient outer product (AGOP). Empirically, we show that the AGOP captures features across a broad class of network architectures including convolutional networks and large language models. Moreover, we use AGOP to enable feature learning in general machine learning models through an algorithm we call Recursive Feature Machine (RFM). Overall, this line of work advances our fundamental understanding of how neural networks extract features from data, leading to the development of novel, interpretable, and effective models for use in scientific applications.

Adityanarayanan Radhakrishnan
Harvard University
aradha@seas.harvard.edu

PP1

Mitigating Hallucinations in Large Language Models Using Meta-Cognition

Large language models (LLMs) can produce human-like, coherent, and contextually appropriate responses to queries. Despite significant advancements in LLMs, they are susceptible to generating incorrect information, posing challenges for their use in safety-critical applications such as autonomous systems. This work investigates the feasibility of leveraging meta-cognition strategies to mitigate hallucinations in LLMs while preserving acquired knowledge. By integrating self-awareness and error detection mechanisms into LLM architectures, we aim to evaluate the effectiveness of meta-cognition in identifying and suppressing hallucination phenomena, which often result in the generation of erroneous sequences.

Sreenivasan Ramasamy Ramamurthy
Bowie State University
sramamurthy@bowiestate.edu

Sumit K. Jha
Florida International University
sumit.jha@fiu.edu

Arvind Ramanathan
Data Science and Learning Division
Argonne National Laboratory
ramanathana@anl.gov

PP1

MS17: Physics-Driven Conditional GAN to Solve Inverse Problems

Recently, a novel conditional Wasserstein GAN was proposed (Ray et al., 2023) to solve physics-based inverse problems. Training this generative model required samples from the joint distribution of the inferred and measured fields. Thus, the GAN implicitly leans the underlying "physics" via the dataset itself. In this poster, we propose an alternate approach for training the conditional GAN to solve PDE-based inverse problems, where the PDE structure is carefully integrated into the training objective function. Such a framework is capable of reducing the data complexity, which can play a critical role in applications where it is not possible to generate a lot of labelled training samples.

Deep Ray
University of Maryland, College Park
deepray@umd.edu

PP1

MS1: Comparison of Weight-Parameterization Methods in Neural ODEs for Surrogate Modeling

Training continuous limits of residual neural networks (ResNets), commonly called neural ordinary differential equations (ODEs), requires determining the neural network weights that define the dynamics. As the weights in ResNets vary from layer to layer, it is natural to model the weights as a function of time and apply techniques from optimal control to perform the training. Although there are some non-parametric approaches, the weights are usually parameterized resulting in an apparent trade-off between the expressiveness and the number of parameters of the model. In this paper, we compare different polynomial parameterizations and highlight the importance of choosing the basis function on supervised machine learning tasks motivated by PDE surrogate modeling tasks. In this application, we aim at achieving high accuracy with as few parameters as possible. Across various types of neural ODEs and two different optimization algorithms our results indicate that restricting the weights to a small dimensional subspace can yield accurate surrogates that generalize well and are efficient to store and evaluate. However, we also observe that the choice of basis functions can have dramatic impact on the achieved accuracy.

Haley Rosso
Emory University
haley.rosso@emory.edu

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

PP1

MS1: Neural ODE/SDE Training in Mixed and Low-Precision

Continuous-dime deep learning can be cast as optimal control of ordinary, partial, or stochastic differential equations. Hence, learning can become computationally expensive regarding the number of operations (e.g., required for time-

stepping) and memory (e.g., to enable gradient computations). One way to lower the computational costs is to resort to lower precision floating or even fixed point arithmetic. This poster will show how to effectively combine the precisions in different parts of the training algorithms to facilitate learning. We will show results for supervised learning, optimal control, and generative models involving various differential equations and discretizations.

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

PP1

MS7: Greedy Construction of Quadratic Manifolds for Nonlinear Dimensionality Reduction and Nonlinear Model Reduction

Dimensionality reduction on quadratic manifolds augments linear approximations with quadratic correction terms. Previous works rely on linear approximations given by projections onto the first few leading principal components of the training data; however, linear approximations in subspaces spanned by the leading principal components alone can miss information that are necessary for the quadratic correction terms to be efficient. In this work, we propose a greedy method that constructs subspaces from leading as well as later principal components so that the corresponding linear approximations can be corrected most efficiently with quadratic terms. Properties of the greedily constructed manifolds allow applying linear algebra reformulations so that the greedy method scales to data points with millions of dimensions. Numerical experiments demonstrate that an orders of magnitude higher accuracy is achieved with the greedily constructed quadratic manifolds compared to manifolds that are based on the leading principal components alone.

Paul Schwerdtner
New York University
ps5136@nyu.edu

Benjamin Peherstorfer
Courant Institute of Mathematical Sciences
New York University
pehersto@cims.nyu.edu

PP1

MS8: Variation Spaces for Multi-Output Neural Networks: Insights on Multi-Task Learning and DNN Compression

This paper introduces a novel theoretical framework for the analysis of vector-valued neural networks through the development of vector-valued variation spaces, a new class of reproducing kernel Banach spaces. These spaces emerge from studying the regularization effect of weight decay in training networks with activations like the rectified linear unit (ReLU). This framework offers a deeper understanding of multi-output networks and their function-space characteristics. A key contribution of this work is a representer theorem for the vector-valued variation spaces. This representer theorem establishes that shallow vector-valued neural networks are the solutions to data-fitting problems over these infinite-dimensional spaces, where the network widths are bounded by the square of the number of training data. This observation reveals that the norm associated with these vector-valued variation spaces encourages the learning of features that are useful for multiple tasks, shedding new light on multi-task learning with neural networks. Finally, this paper develops a connection between weight-decay regularization and the multi-task lasso problem. This connection leads to novel bounds for layer widths in deep networks that depend on the intrinsic dimensions of the training data representations. This insight not only deepens the understanding of the deep network architectural requirements, but also yields a simple convex optimization method for deep neural network compression.

Joseph Shenouda
University of Wisconsin-Madison
jshenouda@wisc.edu

PP1

Use of Data Science in the Field of Energy and Seismic Data Evaluation (Smoothing of Data and Automated Fault Detection Using Ml)

Our proposal focuses on the novel integration of machine learning (ML) techniques in energy systems, specifically addressing the interpretation of complex energy datasets and optimization of energy management strategies. While ML applications in energy have been explored in the past, our session aims to offer unique insights by delving into advanced ML algorithms and their practical implementation in real-world scenarios. While there may be existing tutorials or sessions covering ML in energy systems, our proposal distinguishes itself through its emphasis on cutting-edge techniques, such as deep learning and reinforcement learning, and their application in addressing current challenges in energy data interpretation and management. We do not have specific tutorials or materials available online at this time. However, our team is actively engaged in ongoing research and development in the field of ML for energy systems, and we are committed to delivering a session that offers fresh perspectives and actionable insights for attendees. If desired, we can provide supplementary materials or references to relevant research papers, but we believe that the live presentation format will allow for a dynamic discussion and exchange of ideas that cannot be fully captured in pre-recorded videos or GitHub repositories.

Abhijeet Singh
Rajiv Gandhi Institute of Petroleum Technology
22cd3002@rgipt.ac.in

PP1

MS4: The Easiest Case Is Already Hard: the Complexity of Understanding Decision Trees As Ml Models

Decision trees are traditionally considered one of the most interpretable machine learning models. When looking at decision trees with a couple dozen nodes it is indeed easy to understand whats going on, and why a specific input instance was classified in a certain way. But what about industry size trees having 2000 nodes? It turns out that several natural questions we would hope to be able to answer about a models decision making, like the smallest subset of the input that was actually relevant for a particular decision, are NP-hard over decision trees. Moreover, their probabilistic relaxations turn out to be hard as well, and worse yet, they are hard to approximate. In summary, decisions made by decision trees are not always easy to understand.

Bernardo Subercaseaux

Carnegie Mellon University
bersub@cmu.edu

PP1

MS13: Block Gauss-Siedel Methods for Solving Tensor Regression under the t-Product:

With tensor data becoming more ubiquitous, there is an increasing need for effective randomized algorithms to solve tensor regression problems. While there is a rich body of work exploring row-based Kaczmarz-type methods to tackle this, the literature on column-action methods is limited. In this work, we propose extensions of the Gauss-Siedel method and its variants to the tensor regression regime under the t-product. In this poster, I will present some theoretical convergence guarantees followed by empirical results on synthetic data that highlight the effectiveness of these methods in different settings. Further, the application of these methods to image deblurring will also be discussed.

Shambhavi Suryanarayanan
Princeton University
ss3472@princeton.edu

PP1

MS15: Formal Tests for Semantic Interpretability

Recent works have extended notions of feature importance to semantic concepts that are inherently interpretable to the users interacting with a black-box predictive model. Yet, precise statistical guarantees, such as false positive rate control, are needed to communicate findings transparently and to avoid unintended consequences in real-world scenarios. In this paper, we formalize the global (i.e., over a population) and local (i.e., for a sample) statistical importance of semantic concepts for the predictions of opaque models, by means of conditional independence, which allows for rigorous testing. We use recent ideas of sequential kernelized testing (SKIT) to induce a rank of importance across concepts, and showcase the effectiveness and flexibility of our framework on synthetic datasets as well as on image classification tasks using vision-language models such as CLIP.

Jacopo Teneggi
Johns Hopkins University
jtenegg1@jhu.edu

PP1

MS13: Tensor Completion with BMD Factor Nuclear Norm Minimization

This work is concerned with the problem of recovering third-order tensor data from limited samples. A recently proposed tensor decomposition (BMD) method has been shown to efficiently compress third-order spatiotemporal data. Using the BMD, we formulate a slicewise nuclear norm penalized algorithm to recover a third-order tensor from limited observed samples. We develop an efficient alternating direction method of multipliers (ADMM) scheme to solve the resulting minimization problem. Experimental results on real data show our method to give reconstruction comparable to those of HaLRTC (Liu et al., 2012), a well-known tensor completion method, in about the same number of iterations. However, our method has the advantage of smaller subproblems and higher parallelizability per iteration.

Fan Tian
Tufts University
fan.tian@tufts.edu

Mirjeta Pasha
Massachusetts Institute of Technology
In Fall 2024 Virginia Tech
mpasha@mit.edu

Misha E. Kilmer
Mathematics Department
Tufts University
misha.kilmer@tufts.edu

Eric Miller
Department of Electrical and Computer Engineering
Tufts University
eric.miller@tufts.edu

Abani Patra
Tufts University
abani.patra@tufts.edu

PP1

MS13: Zero-Inflated Poisson Tensor Factorization for Multimodal Genomics Data

Tensor factorizations (TF) are crucial for efficiently representing and analyzing multidimensional data. However, classic TF methods encounter challenges when dealing with zero-inflated count data, where traditional maximum likelihood estimation based techniques struggle. Furthermore, the inherent stochastic nature of TFs introduces variability across repeated runs, posing difficulties in result interpretation and reproducibility. We present Zero Inflated Poisson Tensor Factorization (ZIPTF), an unsupervised approach tailored for high-dimensional count data with excess zeros. To address stochasticity, we propose Consensus Zero Inflated Poisson Tensor Factorization (C-ZIPTF), which combines ZIPTF with a consensus-based meta-analysis approach. Our evaluation on synthetic zero-inflated count data and both synthetic and real multimodal genomics data illustrates the superiority of ZIPTF over baseline matrix and tensor factorization methods in terms of reconstruction accuracy. Additionally, our approach consistently identifies known and biologically relevant gene expression programs in multimodal genomics data. This work showcases the potential of ZIPTF and C-ZIPTF to unlock insights from complex genomics data, paving the way for advancements in understanding biological systems.

Neriman Tokcan
University of Massachusetts Boston
neriman.tokcan@umb.edu

Daniel Chafamo
University of Pennsylvania
daniel.chafamo@pennmedicine.upenn.edu

Vignesh Shanmugam
Broad Institute of MIT and Harvard
vshanmug@broadinstitute.org

PP1

MS18: Multigrid-Augmented Deep Learning Preconditioners for the Helmholtz Equation Using

Compact Implicit Layers

We present a deep learning-based iterative approach to solve the discrete heterogeneous Helmholtz equation for high wavenumbers. Combining classical iterative multigrid solvers and convolutional neural networks (CNNs) via preconditioning, we obtain a learned neural solver that is faster and scales better than a standard multigrid solver. Our approach offers three main contributions over previous neural methods of this kind. First, we construct a multilevel U-Net-like encoder-solver CNN with an implicit layer on the coarsest grid of the U-Net, where convolution kernels are inverted. This alleviates the field of view problem in CNNs and allows better scalability. Second, we improve upon the previous CNN preconditioner in terms of the number of parameters, computation time, and convergence rates. Third, we propose a multiscale training approach that enables the network to scale to problems of previously unseen dimensions while still maintaining a reasonable training procedure. Our encoder-solver architecture can be used to generalize over different slowness models of various difficulties and is efficient at solving for many right-hand sides per slowness model. We demonstrate the benefits of our novel architecture with numerical experiments on a variety of heterogeneous two-dimensional problems at high wavenumbers.

Ido Ben-Yair
Ben-Gurion University of the Negev
idobeny@post.bgu.ac.il

Bar Lerer
Ben Gurion University Negev, Israel
barlere@post.bgu.ac.il

Eran Treister
Ben Gurion University of the Negev
erant@cs.bgu.ac.il

PP1

MS6: Using Persistent Homology to Analyze Access to Resources with Heterogenous Quality

Ideally, all public resources (e.g. parks, grocery stores, hospitals, etc.) should be distributed in a way that is fair and equitable to everyone. However, this is not always the case. Quantifying how much (or little) access individuals have to certain resources is a complex problem. Previous work has shown that tools from topological data analysis (TDA) can be useful in determining "holes" in the locations of resource locations based on geographic locations and travel times [Hickok et al., Persistent homology for resource coverage: a case study of access to polling sites, 2023]. Some resources may necessitate incorporation a notion of quality. As a case study, we look at public parks, which are heterogeneous in many ways. Having access to a park that is hundreds of acres with basketball courts, baseball diamonds, and an aquarium is inherently different than having access to a small patch of grass with an overgrown tennis court. Here we present an exploration of the access to public parks in Chicago using persistent homology, a tool from TDA.

Sarah Tymochko
UCLA
tymochko@math.ucla.edu

Abigail Hickok
Columbia University
aghickok@gmail.com

Gillian Grindstaff
University of Oxford
grindstaff@maths.ox.ac.uk

Jiajie Luo
Department of Mathematics
University of California, Los Angeles
jerryluo8@math.ucla.edu

Mason A. Porter
UCLA
mason@math.ucla.edu

PP1

MS15: Correlated Privacy Mechanisms for Differentially Private Distributed Mean Estimation

Differentially private distributed mean estimation (DP-DME) is a fundamental building block in privacy-preserving federated learning, where a central server estimates the mean of d-dimensional vectors held by n users while ensuring (ϵ, δ)-DP. Local differential privacy (LDP) and distributed DP with secure aggregation (SecAgg) are the most common notions of DP used in DP-DME settings with an untrusted server. LDP provides strong resilience to dropouts, colluding users, and malicious server attacks, but suffers from poor utility. In contrast, SecAgg-based DP-DME achieves an $O(n)$ utility gain over LDP in DME, but requires increased communication and computation overheads and complex multi-round protocols to handle dropouts and malicious attacks. We propose CorDP-DME, a novel DP-DME mechanism that spans the gap between DME with LDP and distributed DP, offering a favorable balance between utility and resilience to dropouts and collusion. CorDP-DME is based on correlated Gaussian noise, ensuring DP without the perfect conditional privacy guarantees of SecAgg-based approaches. Our results demonstrate that (anti) correlated Gaussian DP mechanisms can significantly improve utility in mean estimation tasks compared to LDP – even in adversarial settings – while maintaining better resilience to dropouts and attacks compared to distributed DP.

Sajani Vithana
Harvard University, U.S.
sajani@seas.harvard.edu

PP1

MS10: Interpretable Diffusion Models via Low-Dimensional Data

Recently, diffusion models have emerged as powerful deep generative models, showcasing cutting-edge performance across various applications such as image generation, solving inverse problems, and text-to-image synthesis. These models generate new data (e.g., images) by transforming random noise inputs through a reverse diffusion process. In this work, we study the optimization foundations of diffusion models via low-dimensional modeling. Specifically, under the mixture of low-rank Gaussians model, we show that the training problem of the diffusion model under proper network parameterization is equivalent to the subspace clustering problem. Based on this result, we explain the phenomenon of transition from memorization to generalization in diffusion models. Moreover, we also use it to guide controllable image editions via low-dimensional semantic space.

Peng Wang

University of Michigan
pengwa@umich.edu

PP1

A Modified Extreme Learning Machine Based Diffusion Model in Image Denoising

In this work, we introduce a novel approach to image denoising using an Extreme Learning Machine (ELM) based diffusion model. Inspired by the efficacy of ELM in approximating partial differential equations (PDEs), we adapt this technique to address the problem of image denoising, which can be conceptualized as solving a diffusion process. Traditional ELM methods for solving PDEs typically handle algebraic equations, which may lead to limited accuracy in image denoising. To address this limitation, we propose a modified approach by incorporating a loss function comprising measurement data, equation information, and a carefully designed regularization term. As a result, our method seamlessly integrates a data-driven approach with physical principles. We demonstrate the effectiveness of our proposed method through a series of representative examples. Our results indicate that the method achieves a favorable balance between computational efficiency and denoising accuracy, making it a promising approach for real-world applications.

Yiran Wang
Purdue University
wang2335@purdue.edu

PP1

Exact Recovery in Semi-Supervised Learning: Analysis of Contextual Stochastic Block Model on Gcn and Spectral Method

We delve into the challenge of semi-supervised node classification on the Contextual Stochastic Block Model (CSBM) dataset. Here, nodes from the two-cluster stochastic block model (SBM) are coupled with feature vectors, which are derived from a Gaussian Mixture Model (GMM) that corresponds to their respective node labels. With only a subset of the CSBM node labels accessible for training, our primary objective becomes the accurate classification of the remaining nodes. Venturing into the transductive learning landscape, we, for the first time, pinpoint the information-theoretical threshold for the exact recovery of all test nodes in CSBM. Concurrently, we design an optimal spectral estimator inspired by Principal Component Analysis (PCA) with the training labels and essential data from both the adjacency matrix and feature vectors. We also evaluate the efficacy of graph ridge regression and Graph Convolutional Networks (GCN) on this synthetic dataset. Our findings underscore that, in contrast to graph ridge regression, GCN possesses the ability to achieve the information threshold of exact recovery in a manner akin to the optimal estimator. This disparity highlights the potential role of feature learning in augmenting the proficiency of GCN, especially in the realm of semi-supervised learning.

Zhichao Wang
UCSD
zhw036@ucsd.edu

PP1

MS1: Operator Learning for Hyperbolic Partial Differential Equations

We construct the first rigorously justified probabilistic algorithm for recovering the solution operator of a hyperbolic partial differential equation (PDE) in two variables from input-output training pairs. The primary challenge of recovering the solution operator of hyperbolic PDEs is the presence of characteristics, along which the associated Green's function is discontinuous. Therefore, a central component of our algorithm is a rank detection scheme that identifies the approximate location of the characteristics. By combining the randomized singular value decomposition with an adaptive hierarchical partition of the domain, we construct an approximant to the solution operator using $O(\Psi_\epsilon^{-1}\epsilon^{-7}\log(\Xi_\epsilon^{-1}\epsilon^{-1}))$ input-output pairs with relative error $O(\Xi_\epsilon^{-1}\epsilon)$ in the operator norm as $\epsilon \to 0$, with high probability. Here, Ψ_ϵ represents the existence of degenerate singular values of the solution operator, and Ξ_ϵ measures the quality of the training data. Our assumptions on the regularity of the coefficients of the hyperbolic PDE are relatively weak given that hyperbolic PDEs do not have the "instantaneous smoothing effect' of elliptic and parabolic PDEs, and our recovery rate improves as the regularity of the coefficients increases.

Christopher Wang
Cornell
cyw33@cornell.edu

Alex Townsend
Cornell University
townsend@cornell.edu

PP1

MS2: A Latent Space Approach to Inferring Heterogeneous Reciprocity in Directed Network Data

Reciprocity, or the stochastic tendency for actors to form mutual relationships, is an essential characteristic of directed network data. Existing latent space approaches to modeling direct networks are severely limited by the assumption that reciprocity is homogeneous across subpopulations in a network. In this work, we introduce a new latent space model for directed networks that can quantify sources of heterogeneous reciprocity due to external covariate information and actor-specific effects. To disentangle the factors influencing individual edge formation and reciprocity, we parameterize the model using the marginal edge probabilities and the odds ratio of forming a mutual relationship. We introduce a fast and scalable stochastic variational inference algorithm based on a structured mean-field variational family to approximate the models posterior distribution. Both simulation studies and real-world data examples illustrate that the proposed model effectively uncovers sources of heterogeneous reciprocity in network data.

Xiangyu Wu, Joshua Loyal
Florida State University
xw20bh@fsu.edu, jdl22h@fsu.edu

PP1

MS20: Associated Poster: Large Language Models As Optimizers

Optimization is ubiquitous. While derivative-based algorithms have been powerful tools for various problems, the absence of gradient imposes challenges on many real-world applications. In this work, we propose Optimization

by PROmpting (OPRO), a simple and effective approach to leverage large language models (LLMs) as optimizers, where the optimization task is described in natural language. In each optimization step, the LLM generates new solutions from the prompt that contains previously generated solutions with their values, then the new solutions are evaluated and added to the prompt for the next optimization step. We first showcase OPRO on linear regression and traveling salesman problems, then move on to our main application in prompt optimization, where the goal is to find instructions that maximize the task accuracy. With a variety of LLMs, we demonstrate that the best prompts optimized by OPRO outperform human-designed prompts by up to 8% on GSM8K, and by up to 50% on Big-Bench Hard tasks. Code at https://github.com/google-deepmind/opro.

Chengrun Yang
Google DeepMind
chengrunyang@gmail.com

PP1

MS4: Ms-Umap - Multi-Scale Manifold Learning Through Probabilistic Sampling

Deriving meaningful representations from complex, high-dimensional data in unsupervised settings is crucial across diverse machine learning applications. A prevalent strategy involves nonlinear dimensionality reduction utilizing manifold learning and graph embedding techniques. This study introduces a graph network statistical framework that leverages a diffusion-based approach for sampling network features. Customized for enhancing and guiding the optimization process through stochastic gradient descent (SGD), our framework aims to achieve a more precise representation of the inherent network structure. Our approach introduces a spectrum of sampling techniques, specifically targeting the selection of edges or nodes based on their perceived significance within the underlying graph structure. These techniques encompass the computation of node importance distributions, derived from graph topology measurements like edge betweenness centrality or diffusion wavelets, and subsequent sampling from these distributions. Incorporating multi-scale graph representation into these sampling techniques during graph embedding optimization results in a robust and expressive representation. We validate the effectiveness of our proposed graph embedding through a range of diverse downstream tasks, including clustering.

Shay Deutsch
Discover Financial Services
shaydeu1@gmail.com

Alex Lin
UCLA
atlin271@gmail.com

Lionel Yelibi, Arjun Ravi Kannan
Discover Financial Services
yelibi@spincitylab.com, arjun.kannan@gmail.com

PP1

MS11: Solving High-Dimensional Mean-Field Games with Fictitious Play

Mean-field games study the equilibrium of a game involving infinitely many players. They are shown to have rich connections between reinforcement learning and flow-based generative models. A mean-field game (MFG) system comprises a backward-in-time Hamilton-Jacobi-Bellman (HJB) equation from single players' optimal control problems and a forward-in-time Fokker-Planck (FP) equation describing the evolution of all players. The forward-backward structure makes the computation of MFG difficult, and many existing works in solving high-dimensional MFGs are based on its variational formulation, which is a subset of MFGs. In this work, we explore the effective implementation of fictitious play to solve MFGs. Fictitious play is a concept with a rich history in game theory and was first introduced in the context of MFGs in [Cardaliaguet and Hadikhanloo, 2017]. The idea involves alternately solving the optimal control problem and simulating population evolution. By decoupling the system, we are able to handle a larger class of MFG problems.

Jiajia Yu
Duke University
jiajia.yu@duke.edu

PP1

MS1: Robustifying Long-Memory State-Space Models via Hankel Operator Theory

State-space models (SSMs) that utilize linear, time-invariant (LTI) systems are known for their effectiveness in learning long sequences. However, these models typically face several challenges: (i) they require specifically designed initializations of the system matrices to achieve state-of-the-art performance, (ii) they require training of state matrices on a logarithmic scale with very small learning rates to prevent instabilities, and (iii) they require the model to have exponentially decaying memory in order to ensure an asymptotically stable LTI system. To address these issues, we view SSMs through the lens of Hankel operator theory, which provides us with a unified theory for the initialization and training of SSMs. Building on this theory, we develop a new parameterization scheme, called HOPE, for LTI systems that utilizes Markov parameters within Hankel operators. This approach allows for random initializations of the LTI systems and helps to improve training stability, while also provides the SSMs with non-decaying memory capabilities. Our model efficiently implements these innovations by nonuniformly sampling the transfer functions of LTI systems, and it requires fewer parameters compared to canonical SSMs. When benchmarked against HiPPO-initialized models such as S4 and S4D, an SSM parameterized by Hankel operators demonstrates improved performance on Long-Range Arena (LRA) tasks.

Annan Yu
Cornell University
Center for Applied Mathematics
ay262@cornell.edu

Michael Mahoney
University of California, Berkeley
mmahoney@stat.berkeley.edu

N. Benjamin Erichson
International Computer Science Institute
erichson@icsi.berkeley.edu

PP1

MS2: Preferential Latent Space Models for Net-

works with Textual Edges

Many real world networks contain rich textual information in the edges, such as email networks where an edge (or interaction) between two nodes (or actors) is an email exchange. Other examples include the co-author network and social media networks. The useful textual content is often discarded in most network analyses, resulting in an incomplete view of the interactions between nodes. In this work, we represent the text document between each pair of nodes as a vector that counts the appearances of a set of keywords extracted from the corpus, and propose a new and flexible preferential latent space network model that can offer direct insights on how contents of the textual exchanges modulate the relationships between nodes. We establish identifiability conditions for the proposed model and tackle model estimation using a highly efficient projected gradient descent algorithm. We further derive the non-asymptotic error bound for the estimator from each step of the algorithm. The efficacy of our proposed method is demonstrated through simulations and an analysis of the Enron email dataset.

Maoyu Zhang
Goizueta Business School
Emory University
maoyu.zhang@emory.edu

PP1

MS5: Geometric MCMC for Inference of Coseismic Model for 2011 Tohoku-Oki Earthquake Using Derivative-Informed Neural Operators

We consider the large-scale nonlinear Bayesian inverse problem (BIP) of inferring heterogeneous material properties in coseismic elastic models of earthquakes. Due to the difficulties of (i) the high cost of PDE solutions, (ii) the high dimensionality of the material properties, and (iii) complex posterior geometry, sampling from the posterior distribution using methods such as MCMC has been intractable for this problem. In this work, we propose to overcome the difficulties by training a derivative-informed neural operator (DINO) surrogate of the parametric PDE model, which achieves high accuracy in both approximating the PDE solution map and its Fréchet derivative at a low and tractable training cost compared to conventional operator learning methods. The trained DINO surrogate is employed in a geometric MCMC algorithm to provide scalable and efficient approximate posterior sampling exploiting accurate surrogate posterior local geometry. We demonstrate this approach's efficacy in inferring the earth's heterogeneous material properties (shear modulus) below Japan, using earth surface deformation data from the 2011 Tohoku event.

Ziheng Zhang
The University of Texas at Austin
ziheng@utexas.edu

Lianghao Cao
California Institute of Technology
lianghao@caltech.edu

Thomas O'Leary-Roseberry
The University of Texas at Austin
Oden Institute for Computational and Engineering Sciences
tom@oden.utexas.edu

Umberto Villa
Oden Institute for Computational Engineering & Sciences
uvilla@oden.utexas.edu

Simone Puel
California Institute of Technology
spuel@caltech.edu

Thorsten W. Becker
Department of Earth Sciences
University of Southern California
thorstinski@gmail.com

Omar Ghattas
University of Texas at Austin
omar@oden.utexas.edu

PP1

MS4: Neural Ordinary Differential Equation Enables Radiogenomic Explainable AI for Identifying Post-Radiosurgery Brain Metastasis Radionecrosis

Purpose: Stereotactic radiosurgery(SRS) effectively treats brain metastases(BM) but poses a risk of radionecrosis(RN). A key challenge in managing BM patients post-SRS is the absence of non-invasive diagnostic methods to distinguish RN from true recurrence(TR). We aim to employ a novel neural ordinary differential equation(NODE) model to differentiate post-SRS RN from TR in BM patients. Methods: We designed a model based on heavy ball NODE(HBNODE), enabling tracking of DNN behavior by solving the HBNODE and observing the stepwise derivative evolution. The trajectory of each sample within the Image-Genomic-Clinical(I-G-C) space then becomes traceable. A decision-making field was reconstructed, and a non-parametric model aggregated the optimal solutions to predict TR/RN outcomes. Post-SRS MR image, genomic, clinical features from 90 BMs were used. Performance were compared against 1)a DNN using MR images, and 2)a combined I+G+C features without the HBNODE model. Results: The HBNODE model achieved superior performance of ROCAUC=0.880.04, sensitivity=0.790.02, specificity=0.860.01, and accuracy=0.840.01, outperforming the image-only DNN(AUC=0.710.05, sensitivity=0.660.32) and the 'I+G+C' without HBNODE(AUC=0.810.02, sensitivity=0.580.11). Conclusion: The HBNODE model distinctly identifies RN from TR, aiding explainability in XAI frameworks. Its promising performance encourages further clinical exploration and suggests potential applicability in XAI domains.

Jingtong Zhao
Duke University
jingtong.zhao@duke.edu

Eugene Vaios
Duke University Medical Center
eugene.vaios@duke.edu

Deshan Yang, Kyle Lafata, Ke Lu
Duke University
deshan.yang@duke.edu, kyle.lafata@duke.edu, ke.lu@duke.edu

Hangjie Ji
North Carolina State University
hangjie_ji@ncsu.edu

Peter Fecci

Duke University
peter.fecci@duke.edu

Zhenyu Yang
Duke Kunshan University
zhenyu.yang893@dukekunshan.edu.cn

Chunhao Wang
Duke University
chunhao.wang@duke.edu

PP2

MS30: Operator Learning for Exterior Calculus Surrogate Models

Data-driven surrogate models have proven to be an effective tool for a wide range of physical applications, but their use is limited in instances where structure-preserving guarantees are necessary to match the underlying physics. Additionally, many of such methods must often be retrained for each instance of a problem, particularly if right-hand sides, forcing terms, or spatially-varying coefficients change. We pose a framework for building learning mechanisms for operator learning into a structure-preserving surrogate model framework using exterior calculus, where bespoke Whitney forms are adapted to fit from data and where a nonlinear operator that acts on the modal coefficients of the mimetic surrogate system is learned simultaneously alongside the Whitney forms. We pose examples of such problems for a handful of applications, demonstrating the dimension reduction and structure-preserving capabilities of our approach.

Jonas Actor, Anthony Gruber
Sandia National Laboratories
jaactor@sandia.gov, adgrube@sandia.gov

Nat Trask
University of Pennsylvania
Ntrask@seas.upenn.edu

PP2

Probability Reconstruction for Chemical Kinetics Using a Maximum Entropy Method with Moments

The challenge known as the "curse of dimensionality" has posed significant difficulties for conventional techniques employed in addressing the chemical master equation (CME). This predicament arises when the state space of the Markov chain expands exponentially with the number of species, hindering the computation of the full probability distribution of the CME due to the large data generated. The method of moment provides an efficient alternative to circumvent the challenge, in comparison to other well known approaches such as the stochastic simulation algorithm (SSA) and finite state projection (FSP). Unfortunately, in circumstances where the full marginal probabilities are needed it is necessary to have a process by which to reconstruct them from the moments. In this study, we applied the maximum entropy principle to reconstruct the distribution of certain models. This is accomplished using a finite set of moment constraints, enabling us to gain valuable insights into the underlying probability distribution with increased computational efficiency.

Emmanuel O. Adara
The University of Alabama
eoadara@crimson.ua.edu

Roger Sidje
Department of Mathematics
University of Alabama
roger.b.sidje@ua.edu

PP2

Broader Engagement (BE): What Type of Duck Are You? Identifying Ancestry Based on Limited Number of Snps

Understanding mallard ancestry can help inform conservational efforts for various duck species. This study explores the genetic complexities of various duck breeds using sophisticated machine learning methods. It focuses on breed identification, a common practice in animal genetics and breeding, which is increasingly leveraging artificial intelligence and high-throughput genomic data. Utilizing five duck breeds, comprising 559 individuals and with a total 40,401 SNPs, a challenge lies in identifying the most informative SNPs for optimal breed prediction accuracy and precision. Initial exploration involved the use of PCA components in K-means clustering for breed categorization, which yielded a high silhouette score (0.822), indicating cohesive clusters. A combination of variable selection techniques was employed, including stacking logistic regression with L2 regularization on some filter-based methods of feature selection. The logistic regression selected 15,390 features. These were further reduced to 74 through a combination of three statistical tests: chi-squared test, ANOVA F-test, and mutual information score. The refined feature selection methods significantly improved the accuracy of the Random Forest classifier from 55% to 95.5%. This project provides valuable insights into duck genomics and paves the way for the development of an efficient and user-friendly ancestry analysis framework to analyze a larger high-throughput sequencing project.

Tolulope Adeyina
University of Texas at El Paso, U.S.
tadeyina@miners.utep.edu

PP2

Transformer Learning As a Heat Flow

Transformers are one of the most versatile learning architectures across domains in machine learning and data science. Self-attention matrices lie at the heart of transformer-based architecture. The transformation of input data under consecutive self-attention layers can be interpreted as the evolution of a set of particles over time. Building off the seminal work of Sander et al. (2022), we present a theoretical analysis of this evolution process, its dynamics, and its convergence to a heat flow, within a framework of gradient flows in measure spaces. We also propose a general discretization scheme for gradient flows that involves successfully computing Schrdinger bridges with equal marginals. We prove the uniform convergence of our proposed method to the gradient flow in 2-Wasserstein metric. We present several numerical illustrations of our theoretical results. This is joint work with Garrett Mulcahy, Soumik Pal, and Zaid Harchaoui.

Medha Agarwal
University of Washington
medhaaga@uw.edu

PP2

MS37: Deterministic and Stochastic Moving An-

chor Extragradient Algorithms for Structured Saddlepoint Problems (Poster)

Our work introduces a moving anchor technique to extragradient algorithms for smooth structured minimax problems. First, our moving anchor technique is introduced into the original algorithmic anchoring framework known as EAG. We match the optimal order of convergence in terms of worst-case complexity on the squared gradient norm, $O(1/k^2)$. As many problems of practical interest are nonconvex-nonconcave, the recently developed FEG class of algorithms brings order-optimal anchoring methods developed within EAG to certain nonconvex-nonconcave problem settings. We introduce the moving anchor methods to the FEG class of algorithms and again obtain order-optimal complexity results. Extensions include a preconditioned version of our algorithms, as well as newly developed stochastic moving anchor methods for convex-concave problems. In both convex-concave and nonconvex-nonconcave settings, a variety of numerical examples demonstrate the efficacy and flexibility of the moving anchor framework over its fixed-anchor counterparts. Future directions of the moving anchor framework and other applications are also discussed.

James Alcala
University of Southern California, U.S.
jameska30@gmail.com

Yat Tin Chow
University of California, Riverside
yattinc@ucr.edu

Mahesh Sunkula
Purdue University
msunkula@purdue.edu

PP2

Fundamental Solution-Based RBF Neural Networks for Solving Inverse Boundary Value Problems

The method of fundamental solutions (MFS) has been applied to solve boundary value problems and certain real-world challenges. However, its practical application, especially in solving inverse problems, is hindered by necessity of introducing source points (a fictitious boundary) outside the physical domain. This arises from the significant sensitivity of numerical solutions to placement of source points. To overcome this limitation, we propose a fundamental solution-based RBF neural network for solving inverse boundary value problems with homogeneous equations. In this framework, inspired by MFS, we use fundamental solutions of the homogeneous equations as activation functions and designate source points as centers of neurons. These source points are trained by minimizing the physics-informed loss function. We present numerical experiments in both 2D and 3D domains to demonstrate the effectiveness of the proposed approach.

Wen Li
University of California, Los Angeles
wli198@fordham.edu

Ricardo Araujo Torres
Fordham University
raraujotorres@fordham.edu

PP2

Surrogate-Assisted Parallel Global Optimization Using Sensitivity Analysis Applied to Shaped Charges

Optimization algorithms that use batch sampling, known as parallel optimization, can speed up convergence when the objective function involves evaluating an expensive-to-evaluate black-box simulation. Additionally, sensitivity methods have been successful in speeding up convergence by reducing the dimension of the optimization problem. However, sensitivity analysis coupled with parallel optimization has not been well studied. This work developed several methods to effectively couple sensitivity analysis with parallel optimization. The methods focus on determining the best batch sample selection for a specified surrogate model. To show the effectiveness of the developed methods, an application to maximize the penetration depth of a shaped charge was conducted by optimizing the contours and thicknesses of a multi-material liner as well as shell geometry using a Gaussian process surrogate model. The accuracy and efficiency of the method was compared to other parallel approaches and surrogate models including an artificial neural network and a genetic algorithm.

Matthew Balcer, Derek Armstrong
Los Alamos National Laboratory
mbalcer@lanl.gov, dearmstr@lanl.gov

PP2

MS26: Efficient Wasserstein Splines Through Consecutive Averaging

Capturing data from dynamic processes through cross-sectional measurements is seen in fields such as cell trajectory inference. This inherently involves the challenge of understanding and reconstructing the continuous trajectory of these processes from discrete data points, for which interpolation and approximation play a crucial role. In this work, we propose methods for B-splines and interpolation in the Wasserstein space through consecutive intrinsic averaging. Our methods have the ability to carry out interpolation and approximation with high precision and at a chosen level of refinement, including the capability to accurately infer trajectories in scenarios where particles undergo splitting (division) over time. We rigorously evaluate our method using simulated cell data characterized by bifurcations and merges, comparing its performance against both state-of-the-art trajectory inference techniques and other spline and interpolation methods. The results of our work not only underscore the effectiveness of our method in addressing the complexities of trajectory inference but also highlight its proficiency in performing interpolation and approximation that respects the inherent geometric properties of the data.

Amartya Banerjee
University of North Carolina at Chapel Hill
amartya1@cs.unc.edu

Harlin Lee
University of North Carolina at Chapel Hill
School of Data Science and Society
harlin@unc.edu

Caroline Moosmüller
University of North Carolina at Chapel Hill

Department of Mathematics
cmoosm@unc.edu

PP2

MS23: Toward Consistent Nonlinear Filtering and Smoothing Via Measure Transport

Solving filtering and smoothing problems for geophysical applications involve estimating the hidden states of complex systems and accurately characterizing their uncertainty. Popular algorithms for tackling these problems include ensemble Kalman methods such as the EnKF, EnKS and RTS smoother. While these algorithms yield robust state estimates for high-dimensional models with non-Gaussian statistics, ensemble Kalman methods are limited by linear transformations and are generally inconsistent with the true Bayesian solution. In this presentation, I will discuss how measure transport can be used to consistently transform a prior ensemble into samples from a filtering or smoothing distribution. This approach provides a natural generalization of Kalman methods to nonlinear transformations, thereby reducing the intrinsic bias of classic algorithms with a marginal increase in computational cost. In small-sample settings, I will show how to estimate transport maps for high-dimensional inference problems by exploiting low-dimensional structure in the target distribution. Finally, I will demonstrate the benefit of this framework for filtering and smoothing on chaotic dynamical systems and aerodynamic flows.

Ricardo Baptista
Caltech
rsb@caltech.edu

PP2

MS29: Learning Graph Geometry and Topology via Continuous-Time Message Passing

Graph Neural Networks (GNNs), while transformative, face challenges in capturing the full essence of graph structures due to their reliance on discrete-time message passing. This conventional approach, despite ensuring permutation equivariance, struggles with oversmoothing, underreaching, and computational bottlenecks, thus limiting its effectiveness in capturing graph structure. In this talk, I will introduce Continuous-time Message PASSing Network (COMPASS), a method designed to overcome these limitations by incorporating continuous-time dynamics derived from the heat and wave equations. Based on a solid theoretical foundation, COMPASS distinguishes itself in complex tasks such as the prediction of geometrical and topological features such as Ricci curvature, persistent homology, and the generating parameters of random graphs. Additionally, COMPASS can accurately predict topological attributes of molecular graphs, such as total polar surface area and ring counts, marking a significant improvement over existing discrete-time message passing networks.

Dhananjay Bhaskar
Brown
dbhaskar92@gmail.com

Yanlei Zhang
MILA
yanlei.zhang@mila.quebec

Charles Xu, Xingzhi Sun, Oluwadamilola Fasina
Yale University
charles.xu@yale.edu, xingzhi.sun@yale.edu,
dami.fasina@yale.edu

Guy Wolf
MILA
wolfguy@mila.quebec

Maximilian Nickel
Meta
maxn@meta.com

Michael Perlmutter
Boise State University
mperlmutter@boisestate.edu

Smita Krishnaswamy
Yale University
smita.krishnaswamy@yale.edu

PP2

MS29: Complex Networks with Complex Weights

In many studies, it is common to use binary (i.e., unweighted) edges to examine networks of entities that are either adjacent or not adjacent. Researchers have generalized such binary networks to incorporate edge weights, which allow one to encode nodenode interactions with heterogeneous intensities or frequencies (e.g., in transportation networks, supply chains, and social networks). Most such studies have considered real-valued weights, despite the fact that networks with complex weights arise in fields as diverse as quantum information, quantum chemistry, electrodynamics, rheology, and machine learning. Many of the standard network-science approaches in the study of classical systems rely on the real-valued nature of edge weights, so it is necessary to generalize them if one seeks to use them to analyze networks with complex edge weights. In this paper, we examine how standard network-analysis methods fail to capture structural features of networks with complex edge weights. We then generalize several network measures to the complex domain and show that random-walk centralities provide a useful approach to examine node importances in networks with complex weights.

Lucas Boettcher
University of Florida
Gainesville
lbottcher@ufl.edu

Mason A. Porter
UCLA
mason@math.ucla.edu

PP2

MS26: Comparing Dynamical Systems Using Time-Delayed Invariant Measures

In recent years, several works have compared dynamical systems by studying the discrepancy between their invariant measures under a suitable metric or divergence on the space of probability measures. While the robustness of invariant measures to noisy measurements, uncertain initial conditions, and slow sampling makes them appealing tools for comparing dynamical systems in applications, the approach is also limited by the existence of infinitely many distinct systems all admitting the same invariant measure. To overcome this difficulty, we instead propose studying invariant measures in time-delayed coordinate systems. We present theoretical results which show that, up to a topo-

logical conjugacy, the time-delayed invariant measure can distinguish between dynamical systems, and we provide numerical examples which demonstrate the utility of our proposed method.

Jonah Botvinick-Greenhouse, Yunan Yang
Cornell University
jrb482@cornell.edu, yunan.yang@cornell.edu

PP2

MS28: Efficient Hybrid Learning of Spatial-Temporal Operator Networks

Recent developments in operator-type neural networks, including Fourier Neural Operator (FNO) and Deep Operator Network (DeepONet), have shown encouraging potential for simulating solutions to spatial-temporal Partial Differential Equations (PDEs). Nevertheless, these networks often require significant training resources and may not consistently reach the accuracy needed in various scientific and engineering fields. In this presentation, we introduce a new framework for operator learning that aims to mitigate these challenges. This new approach integrates classical insights from numerical PDE theory and methodologies to enhance the functionality of existing operator neural networks. Our innovative framework specifically targets low-frequency errors through its design, while incorporating a linear layer to manage high-frequency errors. Through numerical tests on a widely recognized 2D Navier-Stokes equations benchmark, we demonstrate notable gains in both computational efficiency and accuracy over current FNO models and conventional numerical methods.

Yuanzhe Xi
Department of Mathematics
Emory University
yuanzhe.xi@emory.edu

Shuhao Cao
Washington University in St. Louis
scao@umkc.edu

Francesco Brarda
Emory University
francesco.brarda@emory.edu

Ruipeng Li
Lawrence Livermore National Laboratory
li50@llnl.gov

PP2

A Mathematical Model at the Wildlife-Livestock-Human Interface to Predict the Implications of Measures to Control Bovine Tuberculosis

The complex and dynamic interactions among wildlife, livestock, and humans create environments conducive to the emergence of new diseases or the reemergence of existing ones. Such outbreaks pose a significant threat to multiple host-species and, in case of cattle infections, farmers face large economic losses. Our study focuses on Mycobacterium bovis that infects domestic and wild animals, as well as humans. In particular, the presence of M. bovis in a wildlife reservoir compromises disease control efforts in the livestock sector and increases the risk of infection in humans involved in wildlife activities (e.g. hunting). To comprehend the transmission dynamics of M. bovis among wildlife, livestock, and human populations, we propose a mathematical model that incorporates cross-species transmission between livestock, wildlife, and humans, as well as contaminated environment transmission routes. The novelty of our model lies in integrating multiple host-species (three distinct populations) plus an external heterogeneous environmental transmission source. The state of Michigan, USA is selected as a case study to estimate the model parameters and disease burden of bovine tuberculosis. Vaccination of livestock and wildlife scenario analysis are performed to quantify the effectiveness of these control strategies to reduce the transmission of bovine tuberculosis among wildlife, livestock, and human populations.

Sarita Bugalia
Central University of Rajasthan, India
saritabugalia44@gmail.com

Liliana Salvador
The University of Arizona
lilianasalvador@arizona.edu

PP2

Broader Engagement (BE): Modeling the Dynamics of Human Trust in Automation

Trust in automation (TiA) is crucial for human-automation interactions, but traditionally measured statically via discrete survey data points. We developed a mathematical model capturing the dynamic nature of TiA, incorporating measurable components across different timescales, and validated it through experimental data. Our findings show that: (1) trust timescales are important for predicting automation usage, emphasizing trust's key role in decision-making; (2) short-timescale trust significantly impacts overall trust dynamics, highlighting the importance of recent interactions; (3) trust levels are generally higher in highly reliable automation scenarios, though this pattern emerges after an initial evaluation period; and (4) automation type, implying varying workloads, does not significantly impact trust dynamics in highly reliable systems but fosters increased trust in lower-reliability systems. This study advances our understanding of trust dynamics in automation, contributing to more intuitive and trustworthy technology development.

Carlos Bustamante Orellana
Arizona State University
cbustam3@asu.edu

Lucero Rodriguez Rodriguez, Lixiao Huang, Nancy Cooke
Arizona State University, U.S.
lucero.rodriguez@asu.edu, lixiao.huang@asu.edu,
nancy.cooke@asu.edu

Yun Kang
Applied Sciences and Mathematics, Arizona State
yun.kang@asu.edu

PP2

Broader Engagement (BE): Turbulence Mitigation on Images Using Gan Models

Turbulence-induced distortions significantly degrade the quality of images in various applications such as astronomical imaging and remote sensing. Generative Adversarial Networks (GANs) have emerged as a powerful tool for image generation and manipulation. This poster presents a study focusing on turbulence mitigation utilizing GAN models. We investigate the effectiveness of GAN-based ap-

proaches in reducing turbulence-induced distortions in images, offering insights into the potential of these models for improving image quality in turbulent environments. The poster discusses the architecture of GAN models employed for turbulence mitigation and presents experimental results demonstrating their performance compared to traditional methods. Furthermore, we highlight the advantages and challenges of using GANs for turbulence mitigation, offering implications for future research and practical applications in imaging systems. Additionally, we recognize that turbulence can manifest instabilities and blurriness simultaneously, further emphasizing the necessity for effective mitigation strategies.

Juan Carrillo Rodriguez
San Diego State University, U.S.
jcarrillo5698@sdsu.edu

PP2

MS31: Equation Discovery Via Symbolic Mathematics

Mathematical equations provide profound insights about complex dynamics across various scientific disciplines. However, discovering such insightful equations in real-world scenarios is faced with significant challenges due to the absence of prior knowledge and sparse and limited data. To address these challenges, we propose a robust framework to discover open-form equations directly from data without prior knowledge. First, symbolic mathematics are utilized to realize the flexible representation of any equations in a binary tree structure. Diverse equations are generated by a reinforcement learning (RL)-guided hybrid generator and their rewards are evaluated by a neural network-based predictive model. The structure of equations is iteratively optimized using a RL strategy and the best-performing equation is finally selected by a parameter-free stability metric. Second, to avoid the overfitting of noise, the initially identified equation is integrated as a physical constraint into the predictive model for robust evaluation and more accurate discovery. According to extensive experiments, our framework is capable of uncovering true equations with complex structures even from limited and highly noisy data. It also opens up new potential for exploring real-world systems with limited understanding.

Yuntian Chen
Eastern Institute of Technology
ychen@eitech.edu.cn

Hao Xu
College of Engineering, Peking University. China
390260267@pku.edu.cn

Mengge Du
Peking University, China
2101111985@stu.pku.edu.cn

PP2

MS33: Ensemble-Based Annealed Importance Sampling

Sampling from a multimodal distribution is a fundamental and challenging problem in computational science and statistics. Among various approaches proposed for this task, one popular method is Annealed Importance Sampling (AIS). In this paper, we propose an ensemble-based version of AIS by combining it with population-based Monte Carlo methods to improve its efficiency. By keeping track of an ensemble instead of a single particle along some continuation path between the starting distribution and the target distribution, we take advantage of the interaction within the ensemble to encourage the exploration of undiscovered modes. Specifically, our main idea is to utilize either the snooker algorithm or the genetic algorithm used in Evolutionary Monte Carlo. We discuss how the proposed algorithm can be implemented and derive a partial differential equation governing the evolution of the ensemble under the continuous time and mean-field limit. We also test the efficiency of the proposed algorithm on various continuous and discrete distributions.

Haoxuan Chen
Institute for Computational and Mathematical Engineering
Stanford University
haoxuanc@stanford.edu

Lexing Ying
Stanford University
Department of Mathematics
lexing@stanford.edu

PP2

Reconstruction of Phonon Relaxation Time Using Boltzmann Transport Equation

In this poster we present a recent completion in addressing an inverse problem associated with the phonon transport equationa fundamental model in material science for heat conductance. Common laboratory practice involves employing temperature measurement data to deduce the relaxation coefficient τ of the material under investigation. Our work approaches this problem from a mathematical perspective, questioning whether the temperature provides adequate information for a successful inversion and exploring methodologies for executing such an inverse problem. To establish the well-posedness of the inversion process, we employ singular decomposition, elucidating the singular component of the solution to trace the information embedded in the equation. Subsequently, PDE-constrained optimization is utilized for the numerical reconstruction where Stochastic Gradient Descent is applied. Moreover, we compare the non-constant heat conductivity in non-diffusive regime compared with that in diffusive regime. This is a joint work with Irene Gamba, Qin Li and Li Wang.

Peiyi Chen
University of Wisconsin-Madison
pchen345@wisc.edu

Irene M. Gamba
Department of Mathematics and ICES
University of Texas
gamba@math.utexas.edu

Qin Li
University of Wisconsin-Madison
qinli@math.wisc.edu

Li Wang
University of Minnesota
wang8818@umn.edu

PP2

MS35: Identification of Differential Equations Via

Weighted Weak Form with Voting

Modern data-driven approaches combining machine learning and numerical methods have provided intriguing possibilities of identifying differential equations from a given data set. Numerical identification of differential equations remains challenging. We explore a weighted weak/integral form of differential equations with respect to a collection of weighted test functions, and introduce a voting strategy to identify the active features for stable identification.

Sung Ha Kang
Georgia Institute of Technology
Mathematics
kang@math.gatech.edu

Jiahui Cheng, Wenjing Liao
Georgia Institute of Technology
jcheng328@gatech.edu, wliao60@gatech.edu

Haomin Zhou
Georgia Institute of Technology
School of Mathematics
hmzhou@math.gatech.edu

PP2

MS21:Mcmc Importance Sampling via Moreau-Yosida Envelopes

Markov chain Monte Carlo (MCMC) is the workhorse computational algorithm employed for inference in Bayesian statistics. Existing approaches require the gradient of the log posterior. In modern parsimonious models, the use of non-differentiable priors is fairly standard, yielding non-differentiable posteriors. Without differentiability, gradient-based MCMC algorithms cannot be employed effectively. Recently proposed proximal MCMC approaches, however, can partially remedy this limitation. These approaches employ the Moreau-Yosida (MY) envelope to smooth the nondifferentiable prior enabling sampling from an approximation to the target posterior. In this work, we leverage properties of the MY envelope to construct an importance sampling paradigm to correct for this approximation error. Specifically, we show that the importance weights are uniformly bounded irrespective of the choice of the smoothing parameter. Leveraging this result, we establish asymptotic normality of the importance sampling estimators of posterior expectations. Our conditions for this result only rely on finiteness of moments under the desired target distribution.

Eric Chi
Rice University
echi@rice.edu

Apratim Shukla, Dootika Vats
Indian Institute of Technology Kanpur
apratims21@iitk.ac.in, dootika@iitk.ac.in

PP2

MS24: Sampling Error Mitigation Through Spectrum Smoothing in Ensemble Data Assimilation

In data assimilation, an ensemble provides a nonintrusive way to evolve a probability density described by a nonlinear prediction model. Although a large ensemble size is required for statistical accuracy, the ensemble size is typically limited to a small number due to the computational cost of running the prediction model, which leads to a sampling error. Several methods, such as localization, exist to mitigate the sampling error, often requiring problem-dependent fine-tuning and design. This work introduces another sampling error mitigation method using a smoothness constraint in the Fourier space. In particular, this work smoothes out the spectrum of the system to increase the stability and accuracy even under a small ensemble size. The efficacy of the new idea is validated through a suite of stringent test problems, including Lorenz 96 and Kuramoto-Sivashinsky turbulence models. This is joint work with Yoonsang Lee at Dartmouth College.

Bosu Choi
University of Texas at Austin
bo.su.choi@dartmouth.edu

Yoonsang Lee
Dartmouth College
Department of Mathematics
yoonsang.lee@dartmouth.edu

PP2

MS24: Details On: Priors for Efficient Three-Dimensional Imaging

The poster will present finer details of the talk "Priors for Efficient Three-Dimensional Imaging", as part of the mini-symposium "Non-intrusive Computational Methods to Incorporate Prior Knowledge for Improved Statistical Accuracy". In particular, the study explores appropriate priors for the challenging problem of 3D image reconstruction from indirect measurements. We apply the proposed techniques for synthetic aperture radar (SAR) imaging and change detection.

Victor Churchill
Trinity College
victor.churchill@trincoll.edu

PP2

MS25: Inferring Network Connectivity Using Reservoir Networks: A Comparative Analysis of Brain Connectivity Metrics for EEG

Electroencephalography (EEG) functional (undirected) and effective (directed) brain connectivity metrics have been used to study communication between brain regions in healthy and diseased conditions, including neurological disorders such as Alzheimer's disease, major depressive disorders, and epilepsy. These metrics capture important aspects of brain connectivity such as linear and nonlinear information flow. However, these different metrics do not converge and provide a significantly different picture of brain connectivity raising the pressing need of resolving this inconsistency and guiding neuroimaging researchers on what connectivity metric to choose. To address this challenge, we systematically compared functional and effective connectivity metrics. We examined two functional connectivity metrics (coherence (coh) and the corrected imaginary part of phase lagged value (ciPLV)) and three established effective connectivity metrics (generalized partial directed coherence (gPDC), direct directed transfer function (dDTF), and pairwise spectral granger prediction (pSGP)), and a new effective connectivity metric based on modeling EEG time-series as a dynamical system (DynSys). We evaluated these metrics using experimental data from an open-access dataset comprising 112 EEG recordings of healthy adults resting with their open (EO) or eyes

closed (EC).

Erik Connerty, Diksha Srishyla, Christian O'Reilly, Vignesh Narayanan
University of South Carolina
erikc@cec.sc.edu, srishyla@email.sc.edu, christian.oreilly@sc.edu, vignar@sc.edu

PP2

MS34: Using Mathematical Models to Identify the Mechanism of Premature Ventricular Complexes in Cardiac Arrhythmias

Premature ventricular complexes (PVCs) are early heartbeats originating in the heart's ventricles, instead of the right atrium. At high frequency, PVCs can impair cardiac function and increase the risk for more serious cardiac arrhythmia. By understanding the mechanism of PVCs, we can improve their treatment by targeting the problem instead of the symptoms. In parasystole, an ectopic pacemaker in the ventricles is responsible for these premature beats. We aim to identify parasystolic mechanisms of premature ventricular complexes from electrocardiograms and determine if they lead to more serious arrhythmias. We built an experimental model of a periodically entrained oscillator using a beating heart cell aggregate coupled with a non-beating aggregate stimulated with blue light. We identified parasystolic rhythms and the phase resetting curve of heart cells from these experiments, which we integrate into computational models. With our physiologically derived models, we evaluate the bifurcation structure of parasystolic dynamics as the stimulus frequency is varied and observe a variety of complex rhythms and bifurcations. These experiments and models aim to improve the identification of PVC mechanisms, which may provide early warning for progression into more severe arrhythmias.

Khady Diagne, Thomas M. Bury, Leon Glass
McGill University
khady.diagne@mail.mcgill.ca, thomas.bury@mcgill.ca, leon.glass@mcgill.ca

Gil Bub
Mcgill University, Department of Physiology, Canada
gil.bub@mcgill.ca

Alvin Shrier
McGill University
Department of Physiology
alvin.shrier@mcgill.ca

Emilia Entcheva
George Washington University
entcheva@email.gwu.edu

PP2

Scikit-Shape: Python Toolbox for Shape Analysis and Segmentation

Many tasks in image processing, e.g. segmentation, surface reconstruction, are naturally expressed as energy minimization problems, in which the free variables are shapes, curves in 2d or surfaces in 3d. We typically express such problems as energies with data (or target) mismatch and geometric regularization components, to be minimized algorithmically to attain the optimal shape. To solve such problems, we have implemented a suite comprising various building blocks of such problems and algorithms to perform the minimization, including geometric regularization, statistical shape priors, adaptive geometric discretization, and fast Newton-type minimization schemes. Moreover, we have developed crucial shape analysis algorithms for statistical analysis and evaluation of the shapes computed, based on elastic shape distance framework. Our main applications are image and data analysis problems, but the infrastructure is quite general, and can be used for problems in other fields as well. All our algorithms are implemented in Python, leveraging on the NumPy/SciPy ecosystem, making them as easy to use as Matlab, also compatible with existing Python tools. Our algorithms is freely available as an open source package for the research community at: http://scikit-shape.org

Gunay Dogan
National Institute of Standards and Technology
gunay.dogan@nist.gov

PP2

Unrolled 1-Bit Foundation Inference Models: Training, Effectiveness, and the Precision-Layer-Complexity Trade-Off

Unrolled deep neural networks have attracted significant attention for their success in various practical applications. We investigate the impact of one-bit quantization on the performance of unrolling techniques when optimizing regularized objective functions, where every weight of the unrolled network is confined to a binary scheme $\{-1, 1\}$. It is demonstrated that by employing one-bit quantization during training, the model can match the performance of the full-precision network. Our approach allows for links to assume a zero value in addition to $\{-1, 1\}$, achieving an effective resolution of approximately 1.58 bits per link, simultaneously allowing for sparse formation of networks, and requiring near-zero multiplication operations, thereby achieving low energy consumption, and superior throughput. We adopt a principled approach to robustness and interpretability of the forward inference model by the algorithm unrolling methodology, which relies on unrolling well-established iterative methods into the layers of a neural network and then learning an enhanced algorithm by leveraging available data. We show that this approach produces speed-ups of at least 100x to 1000x when compared to legacy inference algorithms, is more easily understood in connection to associated physical and naturally creates sparse forward networks. Given the sparse nature of our network, the bit per link rate approaches 1 instead of the 1.58 bits the LLM works.

Arian Eamaz, Farhang Yeganegi, Mojtaba Soltanalian
University of Illinois - Chicago, IL
University of Illinois - Chicago, IL
aeamaz2@uic.edu, fyegan2@uic.edu, msol@uic.edu

PP2

MS24: Maximizing Wasserstein Dependence in Autoencoders

Wasserstein dependence is an optimal transport-based analog of mutual information that has been gaining traction in both statistical inference and representation learning research. We show that (sliced) Wasserstein dependence maximization serves as an effective and computationally efficient training objective for autoencoders, significantly reorganizing the latent space as compared to a Gaussian prior. The proposed method additionally admits a theoretically rigorous Bayesian interpretation that helps clarify the extent to which statistics computed in the latent space

can be decoded into accurate statistical estimates in the original, high-dimensional data space.

Keenan Eikenberry
Dartmouth College
keenan.j.eikenberry@dartmouth.edu

Lizuo Liu
Southern Methodist University
lizuo.liu@dartmouth.edu

Yoonsang Lee
Dartmouth College
Department of Mathematics
yoonsang.lee@dartmouth.edu

PP2

MS32: What's in a Prior? Learned Proximal Networks for Inverse Problems

Proximal operators are ubiquitous in inverse problems, commonly appearing as part of algorithmic strategies to regularize problems that are ill-posed. Modern deep learning models have been applied to these tasks too, as in the framework of plug-and-play or deep unrolling, where they loosely resemble proximal operators. Yet, something essential is lost in these data-driven approaches: there is no guarantee that a general deep network represents the proximal operator of any function, nor is there any characterization of the function for which the network might provide some approximate proximal. This not only makes guaranteeing convergence of iterative schemes challenging but, more fundamentally, complicates the analysis of what has been learned by these networks about the training data. Herein we present learned proximal networks (LPN), prove that they provide exact proximal operators for a data-driven nonconvex regularizer, and show how a new training strategy, dubbed proximal matching, provably promotes the recovery of the log-prior of the true data distribution. Such LPN provide general, unsupervised, expressive proximal operators that can be used for general inverse problems with convergence guarantees. We illustrate our results in a series of cases of increasing complexity, demonstrating that these models not only result in state-of-the-art performance, but provide a window into the resulting priors learned from data.

Zhenghan Fang
Johns Hopkins University
zfang23@jhu.edu

PP2

MS33: Diffusion Models and Flow Matching on General Constrained Domains

Diffusion and Flow Matching models are two related classes of generative models which achieve state-of-the-art results in many domains such as image generation and text-to-speech tasks. Diffusion models consist of a noising process destroying the data and a backward stage defined as the time-reversal of the noising diffusion. Flow matching models work by learning over linear interpolations between two distributions, where the distributions are coupled by an optimal transport mapping. Building on their success, both types of models have recently been generalized to the Riemannian manifold setting. While this setting encompasses many important applications, it does not by default include manifolds defined via a set of inequality constraints, which are ubiquitous in many scientific domains such as robotics, protein design, and quantum state estimation. We discuss methods which attempt to incorporate these constraints in both the diffusion paradigm and in the flow matching paradigm, discussing costs and benefits of the various approaches.

Nic Fishman
Harvard University
njwfish@gmail.com

PP2

MS32: Thermal Radiance Fields: Regularization for Sensor Fusion (Poster)

Thermal imaging has a variety of applications, from agricultural monitoring to building inspection to imaging under poor visibility, such as in low light, fog, and rain. However, reconstructing thermal scenes in 3D presents several challenges due to the comparatively lower resolution and limited features present in long-wave infrared (LWIR) images. To overcome these challenges, we propose a unified framework for scene reconstruction from a set of LWIR and RGB images, using a multispectral radiance field with specialized multispectral regularizers. We calibrate the RGB and infrared cameras with respect to each other, as a preprocessing step using a simple calibration target. We demonstrate our method on real-world sets of RGB and LWIR photographs captured from a handheld thermal camera, showing the effectiveness of our method at scene representation across the visible and thermal spectra. We show that our method is capable of thermal super-resolution, as well as visually removing obstacles to reveal objects that are occluded in either the RGB or thermal channels.

Sara Fridovich-Keil
Stanford University and Georgia Institute of Technology, U.S
sarafk@stanford.edu

Yvette Lin, Xin-Yi Pan
Stanford University, U.S.
yvelin@stanford.edu, xinyipan@stanford.edu

Gordon Wetzstein
Stanford University
gordon.wetzstein@stanford.edu

PP2

Using the Signature Transform for Analysing Reinforcement Learning Agents

Defining differences is a necessary prerequisite to distinguish solutions to a given problem. We aim to investigate the performance and behavior of different reinforcement learning (RL) agents based on trajectories obtained during roll-outs. Here, often the obtained reward of a RL agent is used as a single quantitative evaluation metric, but it can often be seen that different agents obtain the same reward using different behavioral characteristics. The signature transform offers a method to measure distances between paths, where the signature of a path is the collection of specific iterated integrals over the path. We propose and investigate the signature transform obtained from roll-out trajectories as a general behavior descriptor to distinguish solutions of different methods to reinforcement learning problems. Furthermore, we define a robustness score and profile, which allows a structured analysis of the behavior of agents and populations of agents under a changing environ-

ment. The robustness analysis substantiates the abstract diversity induced by distance in signature space through tangible effects and allows a different view on generalization and robustness. We show that these ideas have merit in the analysis of reinforcement learning algorithms. We illustrate that statistical differences between agents can be converted into distances, allowing further analysis, e.g. by dimensionality reduction preserving the distance between the signatures of trajectories.

Jochen Garcke
Institut für Numerische Simulation, Universität Bonn
Fraunhofer SCAI
garcke@ins.uni-bonn.de

Arno Feiden
Fraunhofer SCAI
arno.feiden@scai.fraunhofer.de

PP2

Measuring Entanglement in Physical Networks

Many complex networks, from the brain to the network of atoms or molecules in materials, have true physical manifestation. Hence, the nodes and links of the network cannot cross each other. Because links cannot cross, they force themselves into nontrivial paths, entangling themselves around each other, a phenomena called network entanglement. Previously, network entanglement has been measured using the graph linking number, a computationally costly metric which can only be applied to small systems. Instead, we propose the average crossing number as an alternative measure of network entanglement with significantly lower computational cost. We analytically derive the dependence of the average crossing number on network density, average link length, degree heterogeneity, and community structure. Furthermore, we show that the predictions accurately estimate the entanglement of both network models and of large scale physical networks found in biological and material systems.

Cory Glover, Albert-László Barabási
Northeastern University
glover.co@northeastern.edu, barabasi@gmail.com

PP2

MS30: Reversible and Irreversible Bracket-based Dynamics for Deep Graph Neural Networks

Recent works have shown that physics-inspired architectures allow the training of deep graph neural networks (GNNs) without oversmoothing. The role of these physics is unclear, however, with successful examples of both reversible (e.g., Hamiltonian) and irreversible (e.g., diffusion) phenomena producing comparable results despite diametrically opposed mechanisms, and further complications arising due to empirical departures from mathematical theory. This work presents a series of novel GNN architectures based upon structure-preserving bracket-based dynamical systems, which are provably guaranteed to either conserve energy or generate positive dissipation with increasing depth. It is shown that the theoretically principled framework employed here allows for inherently explainable constructions, which contextualize departures from theory in current architectures and better elucidate the roles of reversibility and irreversibility in network performance. Code is available at the Github repository https://github.com/natrask/BracketGraphs.

Anthony Gruber
Sandia National Laboratories
adgrube@sandia.gov

Kookjin Lee
Arizona State University
kookjin.lee@asu.edu

Nat Trask
University of Pennsylvania
Ntrask@seas.upenn.edu

PP2

MS27: Swarm-Based Gradient Descent Meets Simulated Annealing

In generic non-convex optimization, one needs to be able to pull samples out of local optimal points to achieve global optimization. Two common strategies are deployed: adding stochasticity to samples such as Brownian motion, as is done in simulated annealing (SA), and employing a swarm of samples to explore the whole landscape, as is done in Swarm-Based Gradient Descent (SBGD). The two strategies have severe drawbacks but complement each other on their strengths. SA fails in the accuracy sense, i.e., finding the exact optimal point, but succeeds in always being able to get close, while SBGD fails in the probability sense, i.e., it has non-trivial probability to fail, but if succeeds, can find the exact optimal point. We propose to combine the strength of the two and develop a swarm-based stochastic gradient method with samples automatically adjusting their annealing. Using mean-field analysis and long-time behavior PDE tools, we can prove the method to succeed in both the accuracy sense and the probability sense. Numerical examples verify these theoretical findings.

Zhiyan Ding
University of California, Berkeley
zding.m@berkeley.edu

Martin Guerra, Qin Li
University of Wisconsin-Madison
mguerra4@wisc.edu, qinli@math.wisc.edu

Eitan Tadmor
University of Maryland, College Park
tadmor@umd.edu

PP2

MS23: Control Problems in Wasserstein Space Via Conditional Flow Matching.

The space of squared integrable probability densities in $\mathbb{R}^d$, equipped with the Wasserstein Metric (W_2), has a manifold structure suitable for studying the evolution of densities. This study focuses on optimizing the transportation of a reference density μ, to a target density ν given a dynamic transportation cost. We present an algorithm that uses conditional flow matching and parametric Hamiltonian Flows techniques to approximate solutions to control problems in W_2.

Sebastian Gutierrez Hernandez
Georgia Tech

shern3@gatech.edu

PP2

Broader Engagement (BE): Discovering Phase Boundaries in Equation-of-State Tables: A Machine Learning Approach

Equation of state (EOS) tables are used in hydrodynamic simulations run on high-performance computing (HPC) clusters. However, generating and storing EOS tables for multiphase, multicomponent mixtures over a wide range of pressures and temperatures is computationally infeasible because the input space is high-dimensional. To address this issue, we have developed a neural network-based machine learning model to predict new EOS tables for binary mixtures. In particular, a deep feed-forward neural network (FFNN) trained on a set of 10 EOS tables at particular mixture compositions is able to predict 10 new (hold-out) EOS tables at different mixture compositions. Specifically, the FFNN predicts the phase diagram, including phase fractions and compositions, of the ten hold-out EOS tables with an accuracy of 96.83%. Overall, this approach is computationally efficient and highly accurate, and one of our goals is to extend it to mixtures with more than two components. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-ABS-868002

Kristen Hallas
University of Texas Rio Grande Valley
kristen.hallas01@utrgv.edu

PP2

Machine Learning for System Identification and Parameter Estimation

Incorporating *a priori* physics knowledge into machine learning leads to more robust and interpretable algorithms. In this work, we combine deep learning techniques with numerical methods to solve two problems in dynamical systems theory: dynamics discovery and parameter estimation. We present numerical results of applying these proposed approaches to highly oscillatory and chaotic problems. Finally, we compare the performance of various numerical schemes, such as Runge-Kutta and linear multistep families of methods, in predicting system dynamics and estimating physical parameters for these problems.

Caitlin Ho, Andrea Arnold
Worcester Polytechnic Institute
cjho@wpi.edu, anarnold@wpi.edu

PP2

Convergence Rates for Poisson Learning to a Poisson Equation with Measure Data

Graph-based learning is a field within machine learning that uses similarities between datapoints to create efficient representations of high-dimensional data for tasks like semi-supervised classification, clustering, and dimension reduction. Poisson learning was recently proposed for graph-based semi-supervised learning problems with very few labeled examples, where the widely used Laplacian regularization performs poorly. In contrast to Laplacian regularized learning, where labels are represented as Dirichlet boundary conditions, Poisson learning encodes the labels as point sources and sinks in a graph Poisson equation. In this work, we prove quantitative convergence rates for discrete to continuum convergence for Poisson learning. The problem is challenging since the source term is measure-valued in the continuum, and the continuum Poisson equation does not admit a variational interpretation. This work gives a rigorous mathematical justification for using Poisson learning for semi-supervised learning at low label rates.

Kodjo M. Houssou, Jeff Calder
University of Minnesota
houss001@umn.edu, jwcalder@umn.edu

Leon Bungert
University of Würzburg
leon.bungert@uni-wuerzburg.de

Franca Hoffmann
Caltech
franca.hoffmann@gmail.com

Max Mihailescu
University of Bonn
mihailescu.m@gmx.de

Amber Yuan
Spotify
yuanx290@umn.edu

PP2

MS38: K-Mer Topology for Whole Genome Analysis

DNA is the building block of all life. Similarities and differences between DNA sequences provide insights into evolutionary relationships, mutations, genetic drifts, genome assembly, gene annotation, and more. Having a robust and scalable method that allows for comparisons within species and across species is essential for understanding the foundation of biology. In this poster, we present k-mer topology, an alignment-free sequence analysis method. Utilizing tools from topological data analysis, such as persistent homology, k-mer patterns are extracted from nucleotide sequences. Then, we define a metric on the k-mer topology features, which is utilized for phylogenetic analysis and viral classification.

Yuta Hozumi
Georgia Institute of Technology
hozumiyu@msu.edu

PP2

Separating Capacity of Convolutional Neural Networks

Convolutional neural networks (CNNs) have demonstrated remarkable success in the realm of pattern classification, a core task in the field of machine learning. Particularly notable are their state-of-the-art results in image and sound signal classification. In these applications, CNN-based feature extraction is typically performed first, a trainable classifier such as, e.g., a support vector machine is then applied to the extracted features. Feature extractors generated by CNNs are multi-layer networks, where each layer consists of a convolution with an atom of a frame (i.e., a redundant spanning set), followed by a pointwise nonlinearity and a pooling operator. We explain the reasons behind the success of such feature extractors by studying them through the lens of function-counting theory, a framework for pat-

tern classification established by Cover, 1965. In particular, the notion of separating capacity plays a crucial role in this framework and is based on counting the number of realizable dichotomies (i.e., binary label assignments). We compute the separating capacity of CNN-based feature extractors in terms of their network parameters and architectures (i.e., frames, nonlinearities, pooling operators, widths, and depths). By doing so, we deduce that generic CNNs exhibit a large separating capacity, which explains why such networks excel at feature extraction and hence in classification tasks.

Konstantin Häberle, Helmut Bölcskei
ETH Zurich
haeberlk@ethz.ch, hboelcskei@ethz.ch

PP2

Double Descent and Generalized Aliasing

A central problem in data science is to use noisy samples of an unknown function to predict values for unseen inputs. Traditionally, this error is estimated by using hold-out data from a training procedure and observing the accuracy of this test set. In classical statistics the predictive error for different model classes is seen as a tradeoff between the bias and the variance that balances model simplicity with its ability to fit complex functions. However, in modern machine learning, over-parameterized models often exhibit "double descent' in which models of increasing complexity exhibit decreasing generalization error. We present an alternative paradigm to the bias/variance tradeoff that we call the aliasing/invertibility decomposition. We explain double descent as a systematic "de-aliasing' that occurs in over-parameterized models. In the limit of large models, the contribution due to aliasing vanishes, leaving an expression for the asymptotic total error as the invertibility failure of infinitely large models on (finite) training points. Our decomposition is in terms of general operators acting on a representation, so that the decomposition can be explicitly calculated without seeing any data. This enables us to answer questions related to experimental design and model selection before collecting data or performing experiments.

Tyler Jarvis
Brigham Young University
jarvis@math.byu.edu

PP2

MS28: Conformal Prediction for Multi-Dimensional Time Series by Ellipsoidal Sets

Conformal prediction (CP) has been a popular method for uncertainty quantification because it is distribution-free, model-agnostic, and theoretically sound. For forecasting problems in supervised learning, most CP methods focus on building prediction intervals for univariate responses. In this work, we develop a sequential CP method called MultiDimSPCI that builds prediction regions for a multivariate response, especially in the context of multivariate time series, which are not exchangeable. Theoretically, we estimate finite-sample high-probability bounds on the conditional coverage gap. Empirically, we demonstrate that MultiDimSPCI maintains valid coverage on a wide range of multivariate time series while producing smaller prediction regions than CP and non-CP baselines.

Hanyang Jiang
University of Chicago
scottjhy@gatech.edu

Chen Xu
Georgia Institute of technology
cx9711@gatech.edu

Yao Xie
Georgia Institute of Technology
yao.xie@isye.gatech.edu

PP2

MS35: Parameterized Equations for Nongradient System on Wasserstein Manifold

The Wasserstein manifold presents a powerful framework for modeling probability distributions and their evolution over time. In this poster, we introduce a novel approach for analyzing nongradient systems on the Wasserstein manifold using parameterized equations. Our method leverages the geometric structure of the Wasserstein space to develop a set of equations that capture the dynamics of these systems without relying on gradient information.

Yijie Jin
Georgia Institute of Technology
yijiejin@gatech.edu

PP2

MS38: A Persistent Directed Flag Laplacian

Topological data analysis (TDA) has had enormous success in science and engineering in the past decade. Persistent topological Laplacians (PTLs) overcome some limitations of persistent homology, a key technique in TDA, and provide substantial insight to the behavior of various geometric and topological objects. This work extends PTLs to directed flag complexes, which are an exciting generalization to flag complexes, also known as clique complexes, that arise naturally in many situations. We introduce the directed flag Laplacian and show that the proposed persistent directed flag Laplacian (PDFL) is a distinct way of analyzing these flag complexes. Example calculations are provided to demonstrate the potential of the proposed PDFL in real world applications.

Benjamin D. Jones
Michigan State University
jones657@msu.edu

Guowei Wei
Department of Mathematics
Michigan State University
weig@msu.edu

PP2

MS28: A Deep Learning Method for the Dynamics of Classic and Conservative Allen-Cahn Equations Based on Fully-Discrete Operators

The Allen-Cahn equation is a well-known stiff semilinear parabolic equation used to describe the process of phase separation and transition in phase field modeling of multi-component physical systems, while the conservative Allen-Cahn equation is a modified version of the classic Allen-Cahn equation that can additionally conserve the mass. In this work, we present a novel deep learning method for predicting the dynamics of the classic and conservative Allen-Cahn equations. Specifically, we design two special con-

volutional neural network models, one for each of the two equations, to learn the fully-discrete operators between two adjacent time steps. The loss functions of the two models are defined using the residual of the fully-discrete systems, which result from applying the central finite difference discretization in space and the CrankNicolson approximation in time. This approach enables us to train the models without requiring any ground-truth data. Moreover, we introduce an effective training strategy that automatically generates useful samples along the time evolution to facilitate training of the models. Finally, we conduct extensive experiments in two and three dimensions to demonstrate outstanding performance of our proposed method, including its dynamics prediction and generalization ability under different scenarios.

Lili Ju
University of South Carolina
Department of Mathematics
ju@math.sc.edu

Yuwei Geng, Yuankai Teng
University of South Carolina
ygeng@email.sc.edu, yteng@email.sc.edu

Zhu Wang
Department of Mathematics
University of South Carolina
wangzhu@math.sc.edu

PP2

MS22: Multivariate Quantile Function Forecaster

We propose Multivariate Quantile Function Forecaster (MQF^2), a global probabilistic forecasting method constructed using a multivariate quantile function and investigate its application to multi-horizon forecasting. Prior approaches are either autoregressive, implicitly capturing the dependency structure across time but exhibiting error accumulation with increasing forecast horizons, or multi-horizon sequence-to-sequence models, which do not exhibit error accumulation, but also do typically not model the dependency structure across time steps. MQF^2 combines the benefits of both approaches, by directly making predictions in the form of a multivariate quantile function, defined as the gradient of a convex function which we parametrize using input-convex neural networks. By design, the quantile function is monotone with respect to the input quantile levels and hence avoids quantile crossing. We provide two options to train MQF^2: with energy score or with maximum likelihood. Experimental results on real-world and synthetic datasets show that our model has comparable performance with state-of-the-art methods in terms of single time step metrics while capturing the time dependency structure.

Kelvin Kan
Department of Mathematics, Emory University
kaifung0620@gmail.com

François-Xavier Aubet
Google DeepMind
francois.aubet@gmail.com

Tim Januschowski
Databricks
timoschowski@gmx.de

Youngsuk Park, Konstantinos Benidis
Amazon Research
pyoungsu@amazon.com, kbenidis@amazon.de

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

Jan Gasthaus
Meta
gasthaus@amazon.com

PP2

MS26: Using Linearized Optimal Transport to Predict the Movement of Stochastic Particle Systems

Many problems involve systems of particles evolving over time, typically with the individual particles moving quickly and chaotically. In many applications, however, the behavior of interest is in the overall distribution of particles, which often evolves much slower and more smoothly. To model and predict this slow behavior, we assume the particles form an empirical estimate of a time-dependent distribution and that the evolution of this distribution is smooth. This motivates the use of Linearized Optimal Transport to embed these distributions in an L2-space in order to compute a derivative, and we use this derivative in an Euler-like scheme to predict the distribution's movement. We show that this derivative is exactly the velocity flow vector field which describes the distribution's evolution, and we show that the Euler-like scheme using this vector field gives a prediction which is first-order accurate in Wasserstein distance. By computing a discrete optimal transport map across a sufficiently long time step, we can ignore the stochasticity of the particles and approximately recover this vector field using finite differences. We then develop a prediction scheme which takes an Euler step using this derivative estimate, and we demonstrate its effectiveness on a few example stochastic particle systems.

Nicholas Karris
UC San Diego
nkarris@ucsd.edu

Seungjoon Lee
CSU Long Beach
seungjoon.lee@csulb.edu

Ioannis G. Kevrekidis
Johns Hopkins University
yannisk@jhu.edu

Evangelos Nikitopoulos
UC San Diego
enikitop@ucsd.edu

Alexander Cloninger
University of California, San Diego, USA
acloninger@ucsd.edu

PP2

MS22: Mambev: Vision State Space Models for Birds-Eye-View Generation

Camera-only 3D perception attracts a large amount of attention for autonomous driving systems from academia and industry. CNNs and Transformers have cemented themselves as the backbones of choice for processing vision data

in Birds-Eye View (BEV) object detection tasks. CNNs excel in quick and lightweight feature extraction while Transformer based architectures have shown powerful results in improving mAP across the board, however neither architecture is well equipped to process temporal information in long context settings with low memory, a critical aspect of the problem required for in usage in vehicles. In this work, we present a state space model based framework to learn unified BEV representations with the spatiotemporal transformers in a top-down manner to support multiple autonomous driving perception downstreaming tasks. In a nutshell, we introduce a novel approach to BEV using MAMBA, a recent development in state space models, which offers a significant improvement in both throughput speed and memory usage over transformers while offering a longer context window which can accommodate sequences of arbitrary length.

Hongyu Ke
Georgia state university
hke3@gsu.edu

Jack A. Morris
Georgia State University
jmorris116@gsu.edu

Haoxin Wang
Georgia State University
Department of Computer Science
haoxinwang@gsu.edu

Yi Ding
Georgia State University
yding37@gmail.com

PP2

MS27: Recent Developments in Consensus Based Optimization

Consensus Based Optimization (CBO) algorithms are a recent family of particle methods for solving complex nonconvex optimization problems. In many application settings, the objective function is not available in closed form. Additionally, derivatives may not be available, or very costly to obtain. Consensus Based Optimization makes use of the Laplace principle to circumvent the use of gradients and is well suited for black-box objectives. Most of the available analysis for this recent family of algorithms studies the corresponding mean-field descriptions of the distribution of particles. Especially convergence analysis with explicit rates is of interest to assess algorithm performance and has mostly been done on the level of the mean-field PDEs. However, all results currently in the literature connecting the discrete particle system to the mean-field regime are restricted to finite time domains. In this talk, we present recent advances regarding the CBO algorithm and its variants and discuss uniform-in-time mean field limits. We focus on second-order variants of CBO as they have numerical advantages in terms of convergence and provide a conceptual bridge to Particle Swarm Optimization (PSO), one of the most widely used particle-based optimization methods.

Dohyeon Kim, Franca Hoffmann
California Institute of Technology
dohyeon@caltech.edu, franca.hoffmann@caltech.edu

Seung-Yeal Ha
Seoul National University
syha@snu.ac.kr

PP2

MS26: Tear and Repulsion Enabled Registration of Point Clouds for Manifold Learning

We present a framework for aligning the local views of a possibly closed/non-orientable data manifold to produce an embedding in its intrinsic dimension through tearing. Through a spectral coloring scheme, we render the embeddings of the points across the tear with matching colors, enabling a visual recovery of the topology of the data manifold. The embedding is further equipped with a tear-aware metric that enables computation of shortest paths while accounting for the tear. To measure the quality of an embedding, we propose two Lipschitz-type notions of global distortiona stronger and a weaker onealong with their pointwise counterparts for a finer assessment of the embedding. Subsequently, we bound them using the distortion of the local views and the alignment error between them. We show that our theoretical result on strong distortion leads to a new perspective on the need for a repulsion term in manifold learning objectives. As a result, we enhance our alignment approach by incorporating repulsion. Finally, we compare various strategies for the tear and repulsion enabled alignment, with regard to their speed of convergence and the quality of the embeddings produced.

Dhruv Kohli
UC San Diego
dhkohli@ucsd.edu

Alexander Cloninger
University of California, San Diego, USA
acloninger@ucsd.edu

Gal Mishne
UC San Diego
gmishne@ucsd.edu

PP2

MS33: Convergence of Kinetic Langevin Monte Carlo on Lie Groups

Explicit, momentum-based dynamics for optimizing functions defined on Lie groups was recently constructed, based on techniques such as variational optimization and left trivialization. We appropriately add tractable noise to the optimization dynamics to turn it into a sampling dynamics, leveraging the advantageous feature that the trivialized momentum variable is Euclidean despite that the potential function lives on a manifold. We then propose a Lie-group MCMC sampler, by delicately discretizing the resulting kinetic-Langevin-type sampling dynamics. The Lie group structure is exactly preserved by this discretization. Exponential convergence with explicit convergence rate for both the continuous dynamics and the discrete sampler are then proved under W_2 distance. Only compactness of the Lie group and geodesically L-smoothness of the potential function are needed. To the best of our knowledge, this is the first convergence result for kinetic Langevin on curved spaces, and also the first quantitative result that requires no convexity or, at least not explicitly, any common relaxation such as isoperimetry.

Lingkai Kong
Georgia Institute of Technology
lkong75@gatech.edu

Molei Tao, Molei Tao
School of Mathematics, Georgia Institute of Technology
mtao@gatech.edu, mtao@gatech.edu

PP2

MS33: Uniform Sampling under Isoperimetric Assumptions

We present a new random walk for uniformly sampling high-dimensional convex bodies. It achieves state-of-the-art runtime complexity with stronger guarantees on the output than previously known, namely in Rényi divergence (which implies TV, $\mathcal{W}_2$, KL, χ^2). The proof departs from known approaches for polytime algorithms for the problem — we utilize a stochastic diffusion perspective to show contraction to the target distribution with the rate determined by functional isoperimetric constants of the stationary density.

Yunbum Kook
KAIST
Discrete Mathematics Group, Institute for Basic Science
yb.kook@gatech.edu

Santosh Vempala
Georgia Institute of Technology
vempala@gatech.edu

Matthew Zhang
University of Toronto
matthew.zhang@mail.utoronto.ca

PP2

A Likelihood Based Approach to Distribution Regression Using Conditional Deep Generative Models

In this work, we study the statistical properties of the likelihood-based approach for the distributional regression with conditional deep generative model with full dimensional noise and underlying singular support. To be precise, we use the assume the full dimensional response conditioned on a covariate in concentrated near low dimensional manifold. We provided and proved the desirable convergence rate for the ambient density in relevant metrics (Hellinger and Wasserstein) for our approach with simplified presentation for traditional network classes. We demonstrate that the characterization of the learnable distribution class is broad. It encompasses not only smooth and analytically tractable distributions but also extends to the general manifold case with minimal assumptions. Our analysis emphasizes the importance of introducing a small noise perturbation to the data when they closely align with the manifold. This observation validates the inherent structural challenges encountered in related manifold estimation problems with noisy data, as previously highlighted in the literature. Furthermore, we successfully implement the proposed approach in numerical studies using synthetic and real-world datasets, thereby providing complementary validation to our theoretical findings.

Shivam Kumar
Department of ACMS
University of Notre Dame
skumar9@nd.edu

PP2

Bayesian Optimization for Parameters Search in Radial Basis Function Interpolation

In this contribution, we introduce an innovative approach in the *Radial Basis Function* (RBF) interpolation that facilitates the efficient detection of the optimal shape parameter. In contrast to the classical and widely used *Leave-One-Out Cross-Validation* (LOOCV), which supposedly consists of an exhaustive grid search, we propose the exploitation of *Bayesian Optimization* (BO). BO is a statistical method that involves modeling the error function with a Gaussian process [Snoek J., Larochelle H., Adams R.P., Practical Bayesian Optimization of Machine Learning Algorithms] and, through an iterative procedure, dynamically selects the optimal parameter. Moreover, when the datasets increase their dimensions, the RBF interpolation can be applied in a Partition of Unity setting (RBF-PU) [Fasshauer G.E., McCourt M.J., Kernel-based Approximation Methods Using MATLAB]. This combined technique, on one side, allows the reduction of computational expense by solving small interpolation problems and combining them; on the other, it preserves the estimated error, considering an overlapped decomposition of the initial domain into spherical patches of variable radius. We also propose a bivariate BO search for the shape parameter and the radius of the spherical decomposition of the domain in RBF-PU. Numerical results show that BO sharply reduces the parameter search time with respect to the classical LOOCV technique in RBF and RBF-PU settings.

Sandro Lancellotti
University of Turin
sandro.lancellotti@unito.it

PP2

MS32: Data Adaptive RKHS Tikhonov Regularization for Learning Kernels in Operators

We present DARTR: a Data Adaptive RKHS Tikhonov Regularization method for the linear inverse problem of nonparametric learning of function parameters in operators. A key ingredient is a system intrinsic data adaptive (SIDA) RKHS, whose norm restricts the learning to take place in the function space of identifiability. DARTR utilizes this norm and selects the regularization parameter by the L-curve method. We illustrate its performance in examples including integral operators, nonlinear operators and nonlocal operators with discrete synthetic data. Numerical results show that DARTR leads to an accurate estimator robust to both numerical error due to discrete data and noise in data, and the estimator converges at a consistent rate as the data mesh refines under different levels of noises, outperforming two baseline regularizers using l^2 and L^2 norms.

Quanjun Lang, Fei Lu, Xiong Wang
Johns Hopkins University
quanjun.lang@duke.edu, feilu@math.jhu.edu,
xiong_wang@jhu.edu

Neil Chada
Heriot-Watt University
n.chada@hw.ac.uk

Qingci An
Johns Hopkins University
qan2@jhu.edu

PP2

MS23: Preserving Linear Invariants in Ensemble

Filtering Methods

Data assimilation is an elegant paradigm for estimating evolving state variables from observations. In the filtering setting, we seek to infer the state distribution conditioned on all the observations available up to that time. Ensemble filtering methods tackle this problem by sequentially updating a set of particles to form an empirical approximation for the filtering distribution. For accurate and robust predictions of dynamical systems, discrete solutions must preserve their critical invariants. While modern numerical solvers satisfy these invariants, existing invariant-preserving analysis steps are limited to Gaussian settings and are often not compatible with existing regularization techniques of ensemble filters. The present work focuses on preserving linear invariants, such as mass or stoichiometric balance of chemical species. Using tools from measure transport theory (Spantini et al., 2022, SIAM Review), we introduce a generic class of nonlinear ensemble filters that automatically preserve desired linear invariants in non-Gaussian settings. By specializing this framework to the Gaussian case, we recover a constrained formulation of the Kalman filter. Then, we show how to combine existing regularization techniques for the ensemble Kalman filter (Evensen, 1994, J. Geophys. Res.) with the preservation of the linear invariants. Finally, we assess the benefits of preserving linear invariants for the ensemble Kalman filter and nonlinear ensemble filters.

Mathieu Le Provost, Jan Glaubitz, Youssef M. Marzouk
Massachusetts Institute of Technology
n/a, Glaubitz@mit.edu, ymarz@mit.edu

PP2

MS35: On the Training of Deep Operator Networks

We present a novel training method for deep operator networks (DeepONets), one of the most popular neural network models for operators. DeepONets are constructed by two sub-networks, namely the branch and trunk networks. Typically, the two sub-networks are trained simultaneously, which amounts to solving a complex optimization problem in a high dimensional space. In addition, the nonconvex and nonlinear nature makes training very challenging. To tackle such a challenge, we propose a two-step training method that trains the trunk network first and then sequentially trains the branch network. The core mechanism is motivated by the divide-and-conquer paradigm and is the decomposition of the entire complex training task into two subtasks with reduced complexity. Therein the Gram-Schmidt orthonormalization process is introduced which significantly improves stability and generalization ability. On the theoretical side, we establish a generalization error estimate in terms of the number of training data, the width of DeepONets, and the number of input and output sensors. Numerical examples are presented to demonstrate the effectiveness of the two-step training method, including Darcy flow in heterogeneous porous media.

Sanghyun Lee
Florida State University
slee17@fsu.edu

Yeonjong Shin
North Carolina State University
yeonjong_shin@ncsu.edu

PP2

MS35: Finite Element Operator Network (feonet): Operator Learning for Parametric PDEs Without Labeled Data

Partial differential equations (PDEs) are crucial for modeling diverse phenomena, yet solving them efficiently remains challenging. We propose Finite Element Operator Networks (FEONet), a novel approach combining deep learning and traditional finite element methods to solve parametric PDEs without paired training data. Our method demonstrates robust performance across various benchmarks, emphasizing accuracy, generalization, and computational efficiency. Additionally, we provide a theoretical analysis of FEONet's convergence for second-order linear elliptic PDEs, highlighting the condition number's role in error estimates. Numerical experiments validate our findings, highlighting FEONet's potential in complex domains requiring precise modeling and analysis. This research bridges deep learning and numerical methods, enhancing parametric PDE solutions.

Jaeyong Lee
Chung-Ang University, Korea
jaeyong@kias.re.kr

PP2

MS40: Geometry-Preserving Encoder and Decoder in Latent Diffusion Models (Poster)

In this poster, we conduct a mathematical analysis of a popular machine learning framework for contrastive learning, which has become extensively used in supervised, semi-supervised, and unsupervised learning algorithms. Our focus is on examining the geometrical properties of graphs generated by the contrastive learning algorithm and understanding the impact of neural network optimization on the solutions derived from contrastive learning. Our findings reveal that despite the widespread acclaim for the performance of contrastive learning losses, these methods may be fundamentally ill-posed due to the potential for obtaining ill-posed solutions. However, we demonstrate that neural network optimization plays a crucial role in achieving a robust feature mapping that maintains the clustering structure of the given dataset. To substantiate our theoretical insights, we present numerical results from simple toy examples and real-world datasets, confirming the alignment between our theoretical predictions and empirical observations.

Wonjun Lee
University of Minnesota
University of California, Los Angeles
lee01273@umn.edu

PP2

MS29: Fractional Schrodinger Equation: Eigenfunctions with Physical Potentials, and Fractionally-Enhanced Quantum Tunneling

Fractional evolution equations lack generally accessible and well-converged codes excepting anomalous diffusion. A particular equation of strong interest to the growing intersection of applied mathematics and quantum information science and technology is the fractional Schrdinger equation, which describes sub-and super-dispersive behavior of quantum wavefunctions induced by multiscale media. We derive a computationally efficient sixth-order split-step numerical method to converge the eigenfunctions of the FSE to arbitrary numerical precision for arbitrary fractional order derivative. We demonstrate applications of this code

to machine precision for classic quantum problems such as the finite well and harmonic oscillator, which take surprising twists due to the non-local nature of the fractional derivative. For example, the evanescent wave tails in the finite well take a Mittag-Leffer-like form which decay much slower than the well-known exponential from integer-order derivative wave theories, enhancing penetration into the barrier and therefore quantum tunneling rates. We call this effect fractionally enhanced quantum tunneling. This work includes an open source code for communities from quantum experimentalists to applied mathematicians to easily and efficiently explore the solutions of the fractional Schrdinger equation in a wide variety of practical potentials for potential realization in quantum tunneling enhancement and other quantum applications.

Joshua M. Lewis
Colorado School of Mines
joshuamlewis2@gmail.com

PP2

MS27: Bilevel-Fedcbo: Robust Clustered Federated Learning Through Bilevel Consensus-Based Optimization

Federated learning is an important framework in modern machine learning that seeks to integrate the training of learning models from multiple users, each user with their own local data set, in a way that is sensitive to the users data privacy and to communication cost constraints. In clustered federated learning, one assumes an additional unknown group structure among users, and the goal is to train models that are useful for each group, rather than training a single global model for all users. We propose a novel solution to the problem of clustered federated learning that is inspired by ideas in consensus-based optimization (CBO). Our new CBO-type method is based on a system of interacting particles that is oblivious to group memberships. Our algorithm is accompanied by theoretical justification that is illustrated by real data experiments. Motivated from an additional point of concern in federated learning: the vulnerability of federated learning protocols to backdoor adversarial attacks, we further introduce a modified, improved particle system with enhanced robustness properties that, at an abstract level, can be interpreted as a bi-level optimization algorithm based on interacting particle dynamics. This poster is based on joint works with Nicols Garca Trillos, Konstantin Riedl, and Yuhua Zhu.

Sixu Li
University of Wisconsin
sli739@wisc.edu

Nicolás García Trillos
University of Wisconsin-Madison
Department of Statistics
garciatrillo@wisc.edu

Konstantin Riedl
Technical University of Munich
konstantin.riedl@ma.tum.de

Yuhua Zhu
University of California, Los Angeles
yuhuazhu@ucla.edu

PP2

MS23: Ensemble Score Filter for Tracking Turbulent Atmosphere Dynamics - Generative AI in Data Assimilation

In this study, we introduced a stable and highly efficient implementation of the Ensemble Score Filter (EnSF) for sequential data assimilation (DA) with geophysical systems. As a diffusion model-based generative AI approach for DA, the score filter can effectively store the information of filtering density in the score model, and the EnSF adopts an ensemble approximation scheme to efficiently approximate the filtering density scores. To showcase its advantageous performance, we compared EnSF with a benchmark Local Ensemble Transform Kalman Filter (LETKF) in DA for a surface quasi-geostrophic (SQG) model, and we have demonstrated the superior performance of EnSF over LETKF in scenarios involving incomplete knowledge of the state dynamical model and nonlinear/partial observations.

Siming Liang
Florida State University
sliang@fsu.edu

Feng Bao
Florida State University
Department of Mathematics
fbao@fsu.edu

Hristo G. Chipilski
Florida State University, Department of Scientific Computing
hchipilski@fsu.edu

Guannan Zhang
Oak Ridge National Laboratory
zhangg@ornl.gov

Jeffrey Whitaker
NOAA/Earth System Research Laboratories
Physical Sciences Laboratory
jeffrey.s.whitaker@noaa.gov

PP2

MS31: PROSE: Predicting Multiple Operators and Symbolic Expressions Using Multimodal Transformers

Approximating nonlinear differential equations using a neural network provides a robust and efficient tool for various scientific computing tasks. Observing that families of differential equations often share key characteristics, we seek one network representation across a wide range of equations. Our multimodal approach, called Predicting Multiple Operators and Symbolic Expressions (PROSE), is capable of constructing multiple operators and governing equations simultaneously through a novel fusion structure. In particular, PROSE solves differential equations, predicts future states, and generates the underlying equations of motion by incorporating symbolic "words" through a language model. Experiments with 25600 equations show that PROSE benefits from its multimodal nature, resulting in robust generalization (e.g. noisy observations, equation misspecification, and data imbalance) supported by comparison and ablation studies. PROSE provides a new operator learning framework incorporating multimodal input/output and language models for solving forward and model-discovery problems related to differential equations.

Yuxuan Liu

University of California, Los Angeles
yxliu@math.ucla.edu

Zecheng Zhang
Florida State University
zz23e@fsu.edu

Hayden Schaeffer
Department of Mathematics
University of California, Los Angeles, U.S.
hayden@math.ucla.edu

PP2

MS25: Interpretable Inference of Drift and Control Vector Fields of Stochastic Dynamical Systems

In this presentation, we will discuss a novel approach for inferring drift and control vector fields of a dynamic system independently represented by a stochastic differential equation. First, we propose an optimal perturbation scheme to generate rich data online for identifying the control vector field. In the second step, we formulate the learning problems to unambiguously approximate the vector field without passively relying on the persistency of the excitation condition. We leverage the structural and dynamic properties of the autonomous agents to develop the perturbation scheme for data-driven inference. The optimal number of data points for the inference is determined online to meet the required approximation accuracy. The learning problem is formulated using the pseudospectral method, specifically the Chebyshev polynomial of the second kind. The convergence of the parameter identification scheme is guaranteed using the Lyapunov approach in probability. Finally, simulation results will demonstrate the effectiveness of our interpretable inference scheme.

Avimanyu Sahoo
University of Alabama Huntsville
as0472@uah.edu

Tiffany Lodge
University of Alabama in Huntsville
tbl0009@uah.edu

PP2

MS31: Accelerating Radiation Calculations for Fire Safety Applications Using DeepONet

Computational Fluid Dynamics (CFD) simulations of industrial fires model and simulate all the complex physical phenomena in fires, including combustion, heat transfer, pyrolysis, turbulence, and water suppression. These CFD simulations are computationally expensive, and a moderate-to-large scale simulation takes several days to over two months. One major performance bottleneck is in solving radiative transport equations (RTEs) which contributes up to 90% of total computational time. This work leverages deep operator networks (DeepONets) to develop a fast surrogate to accelerate radiation calculations. The proposed model is trained to learn the nonlinear mapping between radiative absorption and emission and radiative intensity, which corresponds to RTE solutions. We show that the trained model approximates the solution operator of the RTE and can provide accurate radiation predictions at an order-of-magnitude faster rate than traditional numerical models. Special session: Recent Advances of Operator Learning and Foundation-Model-Assisted Multi-Operator Learning Organizers: Zecheng Zhang, and Lu Lu

Xiaoyi Lu
FM Global
xiaoyi.lu@fmglobal.com

PP2

A General Approach for Imputation of Non-Normal Continuous Data Based on Copula Transformation

Dealing with missing data problems for skewed data is a difficult task especially since most of the imputation and data augmentation methodologies assume multivariate normality. The performance of imputation and hence the accuracy of inference on parameters become questionable when the violation of the above assumption occurs. One approach to solve the normality violation is to apply normalizing transformation prior to the imputation phase. However, this approach may introduce new problems such as altering the dependence structure among random variables. We present a general purpose multiple imputation approach based on Copula transformation. The approach is used to effectively transform any continuous multivariate non-normal data to multivariate normal, thereby allowing the imputation using standard normality-based techniques. The method then allows to conveniently back transform the data into original space. Real data sets are used to illustrate the techniques. We then compare the performance of our Copula-based method with other traditional normality based multiple imputation approaches through extensive simulated and real non-normal multivariate datasets. We demonstrate that this method significantly mitigates the problem and hence the practice of making the blind assumption of multivariate normality for non-normal multivariate data under the assumption that data are missing at different mechanisms.

Zhixin Lun
University of Colorado Anschutz Medical Campus
zhixin.lun@cuanschutz.edu

Ravindra Khattree
Oakland University
khattree@oakland.edu

PP2

MS21: Amortized Control of Pathological Neuronal Dynamics under Parametric Variations

Neuronal dynamics, governed by the Hodgkin-Huxley (HH) equations, exhibit complex behaviors influenced by various parameters. Controlling these dynamics, particularly in pathological states, is challenging due to inherent electrophysiological variability and uncertainty in these parameters. Drawing inspiration from numerical approaches for optimal control problems with uncertain parameters, we consider a two-stage approach for tackling this problem. This involves offline amortization of the control solution over a wide parameter range and online adaptation of the control policy to specific settings. We leverage the system dynamics, objective function, and the Hamilton-Jacobi-Bellman equation to learn the value function and employ neural networks to represent policies for managing high-dimensional states and parameters. This approach enables us to devise closed-loop control policies enabling real-time adaptation to new or unseen scenarios without extensive retraining. Numerical experiments demonstrate the efficacy and robustness of our approach in managing perturbations

within HH system dynamics.

Malvern Madondo
Emory University
malvern.madondo@emory.edu

Deepanshu Verma
Department of Mathematics
Emory University
dverma@clemson.edu

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

PP2

Broader Engagement (BE): Sensitivity and Ml Analysis of a Response of Permafrost Soil to Surface Temperature Variations in Variable Topography.

We consider a computational model for energy equation in permafrost soils and its response to surface temperature model [Ling, Zhang' 2003] involving n=9 environmental parameters including the albedo of surface (lake, snow, vegetation, wildfire-affected vegetation). We study the sensitivity of the solutions using Sobol indices sensitivity framework, and set up a ML regression model: these allow to assess the robustness of our computational model as well as to understand the uncertainty associated with the parameters and the model itself. Our simulations and analyses help to determine the response of the soils in the Arctic to the changing climate and to assess the reliability of the model.

Praveeni Mathangadeera
Oregon State University, U.S.
mathangp@oregonstate.edu

Malgorzata Peszynska
Oregon State University
malgo.peszynska@oregonstate.edu

PP2

MS22: HJ-Sampler: a Bayesian Sampler for Inverse Problems of a Stochastic Process by Leveraging Hamilton-Jacobi PDEs and Score-Based Generative Models

The interplay between stochastic processes and optimal control has been extensively explored in the literature. With the recent surge in diffusion models, stochastic processes have increasingly been applied to sample generation. This paper builds on the log transform, known as Cole-Hopf transform in Brownian motion contexts, and extends it to a more abstract framework that involves a linear operator. Within this framework, we found that the well-known relationship between the Cole-Hopf transform and optimal transport is a particular case where the linear operator is the generator of a stochastic process. We also introduce a novel scenario where the linear operator is the adjoint of the generator, linking to Bayesian inference under specific initial and terminal conditions. Leveraging this theoretical foundation, we develop an algorithm, named HJ-sampler, for Bayesian inference for inverse problems of a stochastic differential equation with given terminal observations. The HJ-sampler involves two stages: solving viscous Hamilton-Jacobi (HJ) partial differential equations (PDEs) and sampling from the associated stochastic optimal control problem. Our proposed algorithm naturally allows for flexibility in selecting the numerical solver for viscous HJ PDEs. We introduce two variants of the solver: the Riccati-HJ-sampler, based on the Riccati method, and the SGM-HJ-sampler, which utilizes diffusion models. Numerical examples demonstrate the effectiveness of our proposed methods.

Tingwei Meng
University of California, Los Angeles, U.S.
tingwei@math.ucla.edu

Zongren Zou, Jerome Darbon
Brown University
zongren_zou@brown.edu, jerome_darbon@brown.edu

George E. Karniadakis
Brown University
Division of Applied Mathematics
george_karniadakis@brown.edu

PP2

Machine Learning Parametrization of Sub-Grid Fluxes in 1D Shallow Water Equations

In this work, we consider the 1D shallow-water equations (SWE) and use machine learning to estimate the effect of unresolved sub-grid degrees of freedom. To this end, we first consider a fine-mesh finite-volume discretization of the SWE and define coarse variables as averages in space. Next, we use Machine Learning to estimate effective fluxes on a coarse mesh. Moreover, we also use flux limiters to ensure consistency of the effective coarse-mesh model. We consider statistical steady-state regimes and use numerical simulations to demonstrate agreement between fine-mesh and coarse-mesh discretization in several parameter regimes, including regimes not included in the training dataset.

Md Amran Hossan Mojamder, Ilya Timofeyev
University of Houston
Department of Mathematics
mmojamde@cougarnet.uh.edu, itimofey@central.uh.edu

PP2

MS25: Expressive and Flexible Simulation of Information Spread Strategies in Social Networks Using Planning

In the digital age, understanding the dynamics of information spread and opinion formation within networks is paramount. This research introduces an innovative framework that combines the principles of opinion dynamics with the strategic capabilities of Automated Planning and reinforcement learning. We have developed a menu of numeric PDDL as well as neural network-based reinforcement learning planners tailored for opinion dynamics networks. Our tool empowers users to visualize intricate networks, simulate the evolution of opinions, and strategically influence that evolution to achieve specific outcomes. By harnessing planning techniques, our framework offers a nuanced approach to devise sequences of actions tailored to transition a network from its current opinion landscape to a desired state. This holistic approach provides insights into the intricate interplay of individual nodes within a network and paves the way for targeted interventions. Furthermore, the tool facilitates human-AI collaboration, enabling users to not only understand information spread but also devise

practical strategies to mitigate potential harmful outcomes arising from it.

Protik Nag
Univeristy of South Carolina
pnag@email.sc.edu

Bharath Muppasani, Rae Jones, Biplav Srivastava, Michael Huhns, Vignesh Narayanan
University of South Carolina
bharath@email.sc.edu, sej15@email.sc.edu, biplav.s@sc.edu, huhns@sc.edu, vignar@sc.edu

PP2

Broader Engagement (BE): Classification and Prediction of Heart Diseases Using Machine Learning Algorithms

Heart disease is a serious worldwide health issue because it claims the lives of many people who might have been treated if the disease had been identified sooner.The leading cause of death in the world is cardiovascular disease, usually referred to as heart disease. Creating reliable, effective, and precise predictions for these diseases is one of the biggest issues facing the medical world today. Although there are tools for predicting heart disease, they are either expensive or challenging to apply for determining a patients risk. The aim of this research was to find out some of the best classification models for predicting heart disease. This experiment examined a range of machine learning algorithms, including Logistic Regression, K-Nearest Neighbor, Support Vector Machine, and Artificial Neural Networks, to determine which machine learning algorithm was most effective at predicting heart disease. The UCI heart disease repository provided the dataset for this study. The K-Nearest Neighbor technique was shown to be the most effective machine learning algorithm for the prediction. It will be beneficial to conduct further studies on the application of additional machine learning algorithms for heart disease prediction.

Akua Osei-Nkwantabisa
The University of Texas Rio Grande Valley, U.S.
akua.oseinkwantabisa01@utrgv.edu

PP2

Metric Entropy Limits on Recurrent Neural Network Learning for Lipschitz Fading Memory Systems

Neural networks have attracted significant attention due to their ability to approximate almost any function arbitrarily well. The Kolmogorov-Donoho rate-distortion theory, as introduced in "[Elbrchter et al., Deep neural network approximation theory, 2021]", quantifies the information-efficiency in terms of the number of bits needed to specify the approximating neural network relative to a fundamental limit. This limit is based on the concept of metric entropy which essentially measures the complexity of the approximation problem. We extend this theory to the approximation of sequence-to-sequence maps using recurrent neural networks (RNNs). Specifically, we consider Lipschitz fading memory (LFM) systems, which model dynamical systems that gradually forget long-past inputs. We develop tools to quantify the complexity of these systems in terms of metric entropy and propose a general construction to approximate them using RNNs. We show that RNNs can approximate LFM systems in a metric-entropy optimal manner in the case of exponentially decaying memory and polynomially decaying memory.

Yang Pan, Clemens Hutter
Chair for Mathematical Information Science
ETH Zürich
yangpan@mins.ee.ethz.ch, huttercl@mins.ee.ethz.ch

Helmut Bölcskei
ETH Zurich
hboelcskei@ethz.ch

PP2

MS30: A Novel Ensemble Approach to Uncertainty Quantification in Graph Neural Operators

Neural operators are a recently developed generalization of regression to mappings between functions. They promise to drastically reduce expensive numerical integration of PDEs to fast evaluations of mappings between functional states of a system, i.e., surrogate and reduced-order modeling. Neural operators have already found applications in several areas such as modeling sea ice, combustion, and atmospheric physics. However, even in the ideal case where neural operators recover the training data, there will be uncertainty in interpolatory and extrapolatory regimes. A useful model is one that is transparent about its shortcomings. Single point estimates, hide the broad range of models that reasonably fit training data but present vastly differing predictions. Operator learned models with uncertainty quantification (UQ) would expose users of the models to the full range of surrogate predictions. In this talk, we introduce a novel ensemble approach to UQ in a graph neural operator and compare it to a more standard variational Bayes approach in synthetic benchmarks.

Ravi Patel
Sandia National Laboratories
rgpatel@sandia.gov

PP2

Evolution of Aortic Geometry Modeled with Morphoelastic Growth

Growth is ubiquitous in biological tissues. While growth is normal during development, it is often a hallmark of or response to disease in adults. We study the evolution of the largest blood vessel in the human body, the aorta. Looking at a cohort of 150 normal people ranging from 1 years to 93 years old, we calculate the shape operator on the aortic surface and perform a Gauss mapping of each surface to the unit sphere. This allows us to study changes in shape independently of changes in size. Using constant rate isotropic growth, where the volume of all finite elements increases proportionally to the growth tensor, we can trace out the trajectory of normal aortic growth. A second cohort of patients, with diseased aneurysmal or dissected aortas, was shown to geometrically diverge from the shape-preserving normal growth. Diseased aortic shape evolution is highly sensitive to the initial geometry. Furthermore, the correct shape evolution cannot be captured with a single growth law. Globally the shape evolution of diseased aortas is shown to be similar to the development of amplitude fluctuations on a sphere that can be modeled using spatially heterogenous growth. Aortic surfaces are sub-divided into patches of relatively constant Gaussian curvature. Each individual patch grows with a varying growth rate coupled to the mean Gaussian curvature within the patch, this strategy allows us to reproduce shape evolution of diseased

aortas.

Luka Pocivavsek, Kameel Khabaz, Julius Kim, Nhung Nguyen, Ross Milner
University of Chicago
lpocivavsek@bsd.uchicago.edu, kkhabaz@uchicago.edu, juliuskim1012@uchicago.edu, nhungng@bsd.uchicago.edu, rmilner@bsd.uchicago.edu

Joseph Pugar
Postdoc
jpugar@uchicago.edu

PP2

Broader Engagement (BE): Spatial Heterogeneity Reverses the Evolution of Recombination in Temporally Varying Environments

Recombination is a fundamental biological process that enhances genetic diversity and adaptive potential by generating novel allele combinations. However, the evolutionary dynamics of recombination are complex and influenced by various ecological and genetic factors. This study explores the evolution of recombination in a haploid Wright-Fisher population model under temporally and spatially varying environments. We employed Monte Carlo simulations and analytic approximations to investigate how spatial heterogeneity impacts the evolution of recombination rates between two selected loci within a finite subdivided metapopulation. Our findings reveal that spatial heterogeneity can significantly alter the evolutionary trajectories of recombination. Contrary to previous models predicting an increase in recombination in response to temporally varying selection, our results demonstrate that spatial heterogeneity can reverse this trend, favoring the clustering of genes rather than increased recombination. This clustering preserves advantageous gene combinations under spatially localized selective pressures, thereby enhancing local adaptation but potentially reducing overall genetic diversity. The study extends existing theories by incorporating spatial dimensions into the models of recombination evolution, highlighting the importance of considering both genetic and environmental heterogeneity.

Nawsheen Tarannum Promy
University of New Mexico, U.S.
npromy@unm.edu

PP2

MS30: Learning Dirichlet-to-Neumann Maps for Graphs with Gaussian Processes

We present a novel method for learning Dirichlet-to-Neumann maps on graphs using Gaussian processes, focusing on cases where data obeys some conservation constraint. We aim to construct a data-driven surrogate model and uncertainty estimates for an entire graph, even when data is only available for a subset of nodes and edges. Our approach combines Gaussian processes and discrete exterior calculus to infer a relationship between node and edge values. By optimizing over the reproducing kernel Hilbert space (RKHS) norm and applying a maximum likelihood estimation (MLE) penalty on kernel complexity, we ensure that our surrogate models adhere to conservation laws without excessive complexity. We demonstrate the effectiveness of our method on various datasets, highlighting its potential for scientific applications with limited data.

Adrienne M. Propp
Stanford University
propp@stanford.edu

Jonas Actor
Sandia National Laboratories
jaactor@sandia.gov

Elise Walker
Sandia National Laboratory
eawalke@sandia.gov

Nathaniel Trask
University of Pennsylvania
ntrask@seas.upenn.edu

Houman Owhadi
Applied Mathematics
Caltech
owhadi@caltech.edu

Daniel M. Tartakovsky
Stanford University
tartakovsky@stanford.edu

PP2

MS24: A Statistical-Stochastic Reduced-Order Model for Multiscale Turbulent Systems

The capability of using imperfect stochastic and statistical reduced-order models to capture key statistical features in multiscale nonlinear dynamical systems is investigated. A systematic framework is proposed using a high-order statistical closure enabling accurate prediction of leading-order statistical moments and probability density functions in multiscale complex turbulent systems. A new efficient ensemble forecast algorithm is developed dealing with the nonlinear multiscale coupling mechanism as a characteristic feature in high-dimensional turbulent systems. To address challenges associated with closely coupled spatio-temporal scales in turbulent states and expensive large ensemble simulation for high-dimensional complex systems, we introduce efficient computational strategies using the so-called random batch method. It is demonstrated that crucial principal statistical quantities in the most important large scales can be captured efficiently with accuracy using the new reduced-order model in various dynamical regimes of the flow field with distinct statistical structures. Finally, the proposed model is applied for a wide range of problems in uncertainty quantification, data assimilation, and control.

Di Qi
Purdue University
qidi@purdue.edu

PP2

MS31: Derivative-enhanced Deep Operator Network

Deep operator networks (DeepONets), a class of neural operators that learn mappings between function spaces, have recently been developed as surrogate models for parametric partial differential equations (PDEs). In this work we propose a derivative-enhanced deep operator network (DE-DeepONet), which leverages the derivative information to enhance the prediction accuracy, and provide a more accurate approximation of the derivatives, especially when the training data are limited. DE-DeepONet incorporates

dimension reduction of input into DeepONet and includes two types of derivative labels in the loss function for training, that is, the directional derivatives of the output function with respect to the input function and the gradient of the output function with respect to the physical domain variables. We test DE-DeepONet on three different equations with increasing complexity to demonstrate its effectiveness compared to the vanilla DeepONet.

Yuan Qiu, Nolan Bridges, Peng Chen
Georgia Institute of Technology
yuan.qiu@gatech.edu, bridges@gatech.edu, pchen402@gatech.edu

PP2

MS25: The Effect of Uncertainty on Machine Learning Through Optimal Control

A typical learning process of a neural network can be described by the interaction between the data, the weights and the model architecture. In a standard regime, this interaction involves updating the weights of the neural network model with respect to a series of data batches for a fixed architecture. Multitude of decisions such as the number of layers, the activation function, learning rate, etc; are made during this process of learning. As these decisions are made apriori in the standard setup, there is an inherent uncertainty in the learning process due to the discrepancy between these decisions and the nature of the data. In this talk, we will demonstrate that this uncertainty could be quantified and further leveraged to improve performance in a neural network training process. Towards this end, we will enunciate a novel optimal control-driven mathematical formulation where the dynamics of learning are represented by a differential equation. Subsequently, we will elucidate, how to quantify this uncertainty during learning?" and provide insights into the effect of the this uncertainty. We will end this talk with insights on how to use uncertainty to correct the behavior of the graph neural network while performing anomaly detection?"

Krishnan Raghavan
Mathematics and Computer Science
Argonne National Laboratory
kraghavan@anl.gov

PP2

MS33: Confronting Singularities in Diffusion-Based Generative Modelling

Recently there have been ongoing efforts to apply diffusion-based generative modeling and sampling to physical and chemical systems. One of the challenges in these applications is the singularities that may present in the learned score function. We aim to investigate the impact of these singularities on the performance of diffusion-based algorithms and provide new insights that may alleviate the issue.

Yinuo Ren
Stanford University
yinuoren@stanford.edu

PP2

Using Phase Space Perturbations for Defibrillation in Cardiac Tissue

A voltage wave produced by the cascade of cardiac cells action potentials triggers the contraction of the heart. These voltage waves can be characterized by four level sets: the excitation and refractory fronts and backs. When the level sets intersect in physical space, they form a phase singularity which is characteristic of the tip of a spiral wave. When spiral waves occur in the heart, they sustain fibrillation (an abnormal and fast, irregular heartbeat). Previous studies show that a small stimulus across the refractory wave back allows for the instantaneous movement of the spiral waves tip across physical space which allows for termination of small spiral waves. Here we use the FitzHugh-Nagumo model to show that the elimination of any complex distribution of spiral waves (defibrillation) is accomplishable at any point in time and phase space by an appropriate perturbation performed in phase space. Singularities of spiral waves are located at the center of the limit cycle. To defibrillate, the stimulus requires a minimum region in phase space and the stimulus must move any values in this region across the phase singularity. Such a stimulus is able to eradicate all singularities and return the system to a steady state. This is a different kind of defibrillation than previously shown. These stimulations designed in phase space allow for the instantaneous termination of all spiral waves by removing the phase singularities in physical space.

Evan H. Rheaume, Mikael Toye, Flavio H. Fenton
Georgia Institute of Technology
evan.rheaume@gatech.edu, mtoye@gatech.edu, flavio.fenton@physics.gatech.edu

PP2

Broader Engagement (BE): Advancing Real-Time Risk Stratification: Evaluating the Association Between Comorbidities and Hospitalization Outcomes in Patients with Multimorbidity

Multimorbidity, the presence of multiple chronic diseases, affects over 60% of Arkansans aged 65+, the second-highest rate in the U.S., impacting hospitalization outcomes like length of stay, costs, and mortality. Risk tools like the Elixhauser and van Walraven-ECI (vW-ECI) were developed using data from other states, limiting their accuracy in Arkansas. This study aimed to enhance vW-ECI by integrating Arkansas-specific demographics and clinical factors to improve real-time predictions of hospitalization outcomes. Methods: A retrospective analysis of Arkansas Clinical Data Repository (AR-CDR) data (20142021) was conducted. The study included a test group (=2 chronic diseases, n=5,109) and a control group (< 2 chronic diseases, n=14,094). Data were processed in OpenRefine, and statistical analyses (descriptive stats, Mann-Whitney U, logistic regression, Bayesian models) were done in R. Subgroups analyzed included age, race, sex, ethnicity, and clinical diagnoses. Results: The test group was older (mean=56.87, SD=18.44) than the control (mean=45.72, SD=17.66) ($p < 0.05$). Hypertension was most common (72.07% vs. 16.20%). Significant differences in age and comorbidity scores were noted ($p < 0.001$). Conclusion: Refining vW-ECI by integrating demographic and clinical factors improves risk assessment and hospitalization outcomes for Arkansas and similar regions.

Taiquitha Robins, Christopher P. Wardell, Reid D. Landes, Kevin W. Sexton
University of Arkansas for the Medical Sciences, U.S.
ttrobins@uams.edu, cpwardell@uams.edu, rdlandes@uams.edu, kevin.sexton@uams.edu

Maryam Y. Garza
University of Texas, San Antonio, U.S.

garzam46@uthscsa.edu

Tremaine B. Williams
University of Arkansas for the Medical Sciences, U.S.
tbwilliams@uams.edu

PP2

MS34: FateNet: An Integration of Dynamical Systems and Deep Learning for Cell Fate Prediction

Understanding cellular decision-making, particularly its timing and impact on the biological system such as tissue health and function, is a fundamental challenge in biology and medicine. Existing methods for inferring fate decisions and cellular state dynamics from single-cell RNA sequencing data lack precision regarding decision points and broader tissue implications. Addressing this gap, we present FateNet, a computational approach integrating dynamical systems theory and deep learning to probe the cell decision-making process using scRNA-seq data. By leveraging information about normal forms and scaling behavior near tipping points common to many dynamical systems, FateNet accurately predicts cell decision occurrence and offers qualitative insights into the new state of the biological system. Also, through in-silico perturbation experiments, FateNet identifies key genes and pathways governing the differentiation process in hematopoiesis. Validated using different scRNA-seq data, FateNet emerges as a user-friendly and valuable tool for predicting critical points in biological processes, providing insights into complex trajectories

Mehrshad Sadria
University of Waterloo
msadria@uwaterloo.ca

Thomas M. Bury
McGill University
thomas.bury@mcgill.ca

PP2

AI-Driven Dimensionality Reduction for Scientific Computing with HPC

The extensive adoption of high-performance computing across diverse scientific domains such as high-energy physics, chemistry, and bioinformatics has resulted in a substantial increase in data volumes. To tackle this influx, experimental facilities are increasingly turning to edge data reduction techniques. For instance, detectors and simulations in fields like combustion and fusion generate substantial data, necessitating efficient reduction methods. This abstract presents a novel approach to addressing the challenges posed by high-dimensional data in scientific research, particularly in real-time analysis scenarios. It emphasizes the significance of dimensionality reduction techniques in reducing complex data streams while preserving critical information. The proposed methodology harnesses artificial intelligence (AI) algorithms such as autoencoders, principal component analysis (PCA), and t-distributed stochastic neighbor embedding (t-SNE) to effectively reduce dimensionality without compromising essential insights. By integrating these AI-powered dimensionality reduction techniques into real-time scientific streaming pipelines, researchers can adeptly handle huge datasets, enabling rapid extraction of meaningful patterns. This approach holds promise for advancing our understanding of complex phenomena across various scientific domains in real-time. We validate and assess our proposed approach using specific scientific use cases such as Open Catalyst and NWChem.

Vijayalakshmi Saravanan
University of Texas at Tyler
vsaravanan@uttyler.edu

Sai Karthik N
USD
saikarthik2606@gmail.com

Khaled Z Ibrahim
Lawrence Berkeley National Lab
kzibrahim@lbl.gov

PP2

A Functional Network Criterion for Identifying Seizure Onset Zones from Ieeg Recordings

Epilepsy, affecting millions worldwide, poses significant challenges to patients and healthcare providers due to its debilitating nature and limited treatment options. The critical clinical goal for drug-resistant epilepsy lies in the localization of the initiation site of the seizure (i.e., the seizure focus). It is this localization that allows for successful surgical intervention. Surgical interventions targeting the epileptogenic zone are crucial for those unresponsive to medications, but the lack of precise biomarkers hampers surgical efficacy. To address these difficulties we conduct a retrospective study of ictal iEEG data by reconstructing the functional network using Granger causal analysis. We study the role of functional brain networks in controlling the onset of epileptic seizures and how their unfolding dynamics localize the seizure onset zone. Here we show an effective method for the identification of the seizure onset zone by analyzing the temporal derivative of standard graph theoretic measures. We find that the temporal derivative of node degree is particularly effective at identifying candidate nodes of the seizure focus of Engel class I cases. This method offers an innovative clinical tool that promises real-world benefit.

Smith Kelley
Georgia State University
ksmith55@gsu.edu

Kevin Slote
Department of Mathematics and Statistics
Georgia State University
kslote1@gmail.com

Marrium Shamshad, Mukesh Dhamala
Georgia State University
mshamshad1@gsu.edu, mdhamala@gsu.edu

Igor Belykh
Department of Mathematics and Statistics
Georgia State University
ibelykh@gsu.edu

PP2

A Regularized Cox Hierarchical Model for Incorporating Annotation Information in Predictive Omic Studies

Background: Associated with high-dimensional omics data there are often meta-features such as pathways and functional annotations that can be informative for predicting

an outcome of interest. We developed a regularized hierarchical framework for integrating meta-features, with the goal of improving prediction and feature selection performance with time-to-event outcomes. Methods: A hierarchical framework is deployed to incorporate meta-features. Regularization is applied to the omic features as well as the meta-features so that high-dimensional data can be handled at both levels. The proposed hierarchical Cox model can be efficiently fitted by a combination of iterative reweighted least squares and cyclic coordinate descent. Results: In a simulation study we show that when the external meta-features are informative, the regularized hierarchical model can substantially improve prediction performance over standard regularized Cox regression. Importantly, when the external meta-features are uninformative, the prediction performance based on the regularized hierarchical model is on par with standard regularized Cox regression, indicating robustness of the framework. We illustrate the proposed model with applications to breast cancer and melanoma survival based on gene expression profiles, which show the improvement in prediction performance by applying meta-features, as well as the discovery of important omic feature sets with sparse regularization at meta-feature level.

Dixin Shen
Gilead Sciences
dixinshen@gmail.com

Juan Pablo Lewinger
University of Southern California
lewinger@usc.edu

PP2

MS21: Optimal Allocation in Splitting Methods for Variance Reduction

Splitting and killing schemes (often referred to as just "splitting") are among the oldest methods for enhanced Monte Carlo sampling. In splitting and killing, we simulate many realizations of a Markov chain, and we introduce interactions between the Markov chains that encourage the sampling of targeted events of interest. This poster reviews recent and ongoing work on a particular method, Weighted Ensemble, and reflects work in collaboration with D. Aristoff, J.Copperman, R.J. Webber, and D.M. Zuckerman.

Gideon Simpson
Department of Mathematics
Drexel University
grs53@drexel.edu

PP2

MS29: Slepians on Directed Signed Graphs

Graph slepian is an important concept for analyzing graph signals on a subgraph of the given graph. Graph slepians are the signals that are maximally concentrated within a subset of nodes while at the same time are bandlimited in the graph frequency domain. However, the existing slepian designs on graphs are restricted to only undirected and unsigned graphs with non-negative edges. Negative edges frequently serve as a valuable tool for representing dissimilarity relations. For example, in social networks, users may share or oppose political perspectives, trust or distrust each other's recommendations, or have positive or negative feelings toward one another. These dissimilarities can be effectively depicted using signed graphs, where edges can take both positive and negative values. In this work, we extend the graph slepians to (directed) signed graphs. Our proposed slepian construction is based on recently introduced signed magnetic Laplacian. We further study the proposed graph slepian properties for balanced and unbalanced signed graphs.

Rahul Singh
Yale University
r.singh@yale.edu

Dhananjay Bhaskar
Brown
dbhaskar92@gmail.com

Smita Krishnaswamy
Yale University
smita.krishnaswamy@yale.edu

PP2

Data-Driven Modelling with Limited Data Via Laplacian Kernelized Dynamic Mode Decomposition and Gaussian Random Vectors

Prevalent methods of reduced order modelling, primarily Dynamic Mode Decomposition (DMD) encounters pitfalls when limited data is acquired. Hence, it's been a key challenge faced by numerous practitioners in the wide variety of field such as fluid dynamics, epidemiology, neuroscience and to name a few. In the interest of overcoming this very challenge, we provide an exciting algorithm of executing Kernelized Extended DMD (KeDMD) which employs the Gaussian random vectors to recover the dominant Koopman modes for the standard fluid flow across cylinder experiment. The theoretical perspective of this algorithm raises two critical issues. These issues are observable replacement by Gaussian probability density function and data vector replacement by Gaussian random vectors. In this talk, we study the KeDMD algorithm powered by Gaussian Random Vectors coupled with the Laplacian Kernel Function which is a common kernel choice in machine learning architectures of speech and text recognition. We also settle the much-needed theory of the aforementioned critical issues. Further, we study the reproducing kernel Hilbert space (RKHS) of the Laplacian Kernel and its interface on the Koopman operator. This deep mathematical analysis here demonstrates that the Koopman operator is closable on the RKHS of the Laplacian Kernel Function unlike that of Gaussian Radial Basis Function Kernel which leverages in making it a unique and novel choice of kernel function for such experimental situation.

Himanshu Singh
Los Alamos National Laboratory, U.S.
singh_h@lanl.gov

PP2

The Twitter Effect: How Anti-Regulation Organizations Drive Firearm Acquisitions in the United States

This study seeks to address the causal role in the Weiner-Granger sense of social media and media as drivers of firearm aquisition in the United States. Firearm injuries are a leading cause of death in the United States, where a greater number of people die by firearms than in motor vehicle crashes. Although firearms present a significant public health risk, Americans continue to purchase them in large amounts. Three drivers of firearm acquisi-

tion are commonly cited in literature, namely fear of violent crime, fear of mass shootings, and panic-buying. These drivers have been shown to have causal links on a weekly scale. In addition, the activity of relevant interest groups on social media, which capitalize on emotions such as fear, may impact impulsive firearm acquisition in the short time scales of days. We examine the combined role of all factors in driving firearm purchases has never been tested in the causal framework, PCMCI.

Kevin Slote
Georgia State University
kslote1@gsu.edu

Kevin Daley
Department of Mathematics, Georgia State University
kevin.m.daley5.ctr@us.navy.mil

Rayan Succar
Department of Mechanical and Aerospace Engineering
New York University Tandon School of Engineering
rs7949@nyu.edu

Roni Ventura
New York University
rbv215@nyu.edu

Maurizio Porfiri
Department of Mechanical and Aerospace Engineering
New York University Tandon School of Engineering
mporfiri@nyu.edu

Igor Belykh
Department of Mathematics and Statistics
Georgia State University
ibelykh@gsu.edu

PP2

Under-Parameterized Double Descent for Ridge Regularized Least Squares Denoising of Data on a Line

The relationship between the number of training data points, the number of parameters, and the generalization capabilities has been widely studied. Previous work has shown that double descent can occur in the over-parameterized regime, and believe that the standard bias-variance trade-off holds in the under-parameterized regime. These works provide multiple reasons for the existence of the peak. We present a simple example, ridge regularized least squares denoising with data on a line embedded in high-dimension space, that provably exhibits double descent in the under-parameterized regime and does not seem to occur for reasons provided in prior work. By deriving an asymptotically accurate formula for the generalization error, we observe sample-wise and parameter-wise double descent with the peak in the under-parameterized regime rather than at the interpolation point or in the over-parameterized regime. The location of the peak depends on the strength μ of the ridge regularizer.

Rishi Sonthalia
University of California Los Angeles
rsonthal@math.ucla.edu

Xinyue Li, Bo Chao Gu
UCLA
xinyue05@g.ucla.edu, billgubochao@g.ucla.edu

PP2

Entropic Information Functionals Combined with Neural Networks Explainability to Detect Events in Signals

In this study, we employ a novel methodology combining entropic and informational functionals with neural networks to autonomously detect events in signals. Our feature extraction process involves dividing the signals into non-overlapping windows and employing non-parametric kernel density estimation. This transforms the signals into an ordered sequence of probability density functions, which is the cornerstone of our approach. By applying Shannon, generalized Rnyi and Tsallis entropies, along with complementary measures such as non-parametric Fisher information, our methodology aims to capture the dynamic characteristics of the signals. Moreover, we employ eXplainable AI (XAI) techniques to understand the role of each measure in determining the events' offset. As a case study, we consider epileptic seizure detection in scalp EEG recordings. Epilepsy is a neurological disorder characterised by recurring, unprovoked seizures. Detection of epilepsy is crucial for accurate diagnosis. Traditional approaches typically rely on describing the frequency of seizure occurrences. Our experiments show a significant dimensionality reduction compared to the raw signals, thus aiding in the accurate classification of epileptic seizures. Additionally, the application of XAI methods shows how the classifier assigns importance to specific measures and channels, offering insights not only into the classifiers decision-making process but also into the underlying dynamics driving a seizure.

Antonio Squicciarini
Universidad Politecnica de Madrid
Departamento Matemática Aplicada a la Ingeniería Industrial
a.squicciarini@alumnos.upm.es

Mario Refoyo, David Luengo
Universidad Politécnica de Madrid
Departamento de Ingeniería Audiovisual y Comunicaciones
m.refoyo@upm.es, david.luengo@upm.es

Carlos E. González-Guillén
Universidad Politécnica de Madrid
Departamento Matemática Aplicada a la Ingeniería Industrial
carlos.gguillen@upm.es

Alejandro Zarzo Altarejos
Departamento Matemática Aplicada a la Ingeniería Industrial
Universidad Politécnica de Madrid
alejandro.zarzo@upm.es

PP2

MS31: A Foundation Model for Partial Differential Equations: Multi-Operator Generalization and Extrapolation

Foundation models, such as large language models, have demonstrated success in addressing various language and image processing tasks. In this work, we introduce a multimodal foundation model for scientific problems, named

PROSE-PDE. Our model, designed for bi-modality to bi-modality learning, is a multi-operator learning approach which can predict future states of spatiotemporal systems while concurrently learning the underlying governing equations of the physical system. Specifically, we focus on multi-operator learning by training distinct one-dimensional time-dependent nonlinear constant coefficient partial differential equations, with potential applications to many physical applications including physics, geology, and biology. More importantly, we provide three extrapolation studies to demonstrate that PROSE-PDE can generalize physical features through the robust training of multiple operators and that the proposed model can extrapolate to predict PDE solutions whose models or data were unseen during the training. Furthermore, we show through systematic numerical experiments that the utilization of the symbolic modality in our model effectively resolves the well-posedness problems with training multiple operators and thus enhances our model's predictive capabilities.

Jingmin Sun
Carnegie Mellon University
jingmins@andrew.cmu.edu

Yuxuan Liu
University of California, Los Angeles
yxliu@math.ucla.edu

Zecheng Zhang
Florida State University
zz23e@fsu.edu

Hayden Schaeffer
Department of Mathematics
University of California, Los Angeles, U.S.
hayden@math.ucla.edu

PP2

MS38: Persistent Path Dirac

Significant progress has been made in advancing the theoretical underpinnings of topological data analysis (TDA) by conceptualizing and refining the Dirac operator, a fundamental mathematical construct used in analyzing topological signals and molecular representations. However, the current methodologies mostly rely on traditional frameworks such as Vietoris-Rips complexes and alpha complexes. This academic pursuit aims to introduce a novel methodological approach by integrating Dirac operators with path homology within the framework of TDA, with a specific emphasis on elucidating molecular structures. Path homology is a theoretical construct that extends the domain of graph homology to accommodate directed graphs and hypergraphs. Thus, it provides a more comprehensive analytical toolkit conducive to interrogating structures endowed with directional attributes. The proposed methodological framework involves extracting features from such complex structures, which are examined by various exemplary instances. This work seeks to contribute to the ongoing discourse within TDA by advancing methodologies that enhance our understanding and analytical capabilities concerning intricate molecular architectures. Thus, it aims to foster broader applicability and robustness in data-driven investigations within computational biology and related disciplines.

Faisal Suwayyid
Michigan State University
suwayyid@msu.edu

PP2

Ecg Classification: Identifying Abnormalies by Leveraging Data for Diagnostic Accuracy

The electrocardiogram (ECG) is a valuable tool for diagnosing heart conditions. However, the standard 12-lead ECG doesn't provide enough detail for certain clinical applications, such as locating the source of an arrhythmia. Current techniques require multiple ECG readings and expensive medical imaging to create detailed anatomical models. We propose an alternative approach that combines the 12-lead ECG with advanced machine learning to first identify different heart conditions, with the ultimate goal of predicting the spatiotemporal activation map of the human heart from the 12-lead ECG data.

Gbocho Masato Terasaki, Zoe Loh
UC Merced
gterasaki@ucmerced.edu, zloh@ucmerced.edu

Gryphon Kumfert, Ryan Nguyen
UC Riverside
gkumf001@ucr.edu, rnguy135@ucr.edu

Ahmadzaki Arjmand
UC Merced
aarjmand@ucmerced.edu

PP2

Neyman-Pearson Multi-Class Classification via Cost-Sensitive Learning

Most existing classification methods aim to minimize the overall misclassification error rate. However, in applications such as loan default prediction, different types of errors can have different consequences. Two popular paradigms have been developed to address this asymmetry issue: the Neyman-Pearson (NP) paradigm and the cost-sensitive (CS) paradigm. Previous works on the NP paradigm have predominantly focused on the binary case, while the multi-class NP problem presents a significantly greater challenge due to its unknown feasibility. In this work, we tackle the multi-class NP problem by establishing a connection with the CS problem through strong duality and propose two algorithms. In the binary NP problem, the key requirement for desirable algorithms is termed NP oracle inequalities. We extend these NP oracle inequalities from the binary case to NP oracle properties in the multi-class scenario, demonstrating that our two algorithms satisfy NP oracle properties under certain conditions. Furthermore, we propose practical algorithms to verify the feasibility and strong duality of a multi-class NP problem, which can offer practitioners the landscape of a multi-class NP problem with various target error levels. Simulations and real data studies confirm the effectiveness of our algorithms. To our knowledge, this is the first study to address the multi-class NP problem with theoretical guarantees.

Ye Tian
Columbia University
Department of Statistics
ye.t@columbia.edu

PP2

MS24: Large Deviation-Informed Sampling for

Rare Events in High Dimensions

Rare and extreme events like hurricanes, energy grid blackouts, dam breaks, earthquakes, and pandemics are infrequent but have severe consequences. Because estimating the probability of such events can inform strategies that mitigate their effects, scientists must develop methods to study the distribution tail of these occurrences. However, calculating small probabilities is hard, particularly when involving complex dynamics and high-dimensional random variables. In this poster, I will discuss our proposed method for the accurate estimation of rare event or failure probabilities for expensive-to-evaluate numerical models in high dimensions. The proposed approach combines ideas from large deviation theory and adaptive importance sampling. The importance sampler uses a cross-entropy method to find an optimal Gaussian biasing distribution, and reuses all samples made throughout the process for both, the target probability estimation and for updating the biasing distributions. Large deviation theory is used to find a good initial biasing distribution through the solution of an optimization problem. Additionally, it is used to identify a low-dimensional subspace that is most informative of the rare event probability. We compare the method with a state-of-the-art cross-entropy-based importance sampling scheme using examples including a tsunami problem.

Shanyin Tong
Columbia University
st3503@columbia.edu

Georg Stadler
Courant Institute for Mathematical Sciences
New York University
stadler@cims.nyu.edu

PP2

Data-Driven Vehicle Design Optimization Through Aerodynamics Informed Dimensionality Reduction

Current methods for optimizing vehicle aerodynamics involve examining a high-dimensional parameter space of different geometric configurations through costly computational studies and wind tunnel experiments. This naturally calls for data-driven approaches to generate lower-cost, efficient surrogate aerodynamic models to expedite the vehicle design cycle. We analyze a dataset of industry automobile geometries with their associated aerodynamic performance obtained from large-eddy simulations. A nonlinear autoencoder is utilized to identify a mapping between the coarsened vehicle representations and a low-dimensional latent space manifold description while a secondary decoder is simultaneously trained to estimate the aerodynamic drag from the latent coordinates. With the current manifold learning procedure, we obtain a differential manifold for which the coordinates effectively describe our vehicle design space in relation to aerodynamic performance. We perform aerodynamic design optimization by making use of the local gradients of drag with respect to the manifold coordinates. We demonstrate that the aerodynamic trends for the decoded geometries produced from the optimization process show agreement with a validation LES analysis. This procedure demonstrates the potential for data-driven methods to accelerate the analysis and optimization of design parameters for ground, air, and marine vehicles.

Jonathan Tran, Kai Fukami
University of California, Los Angeles
jqtranus@g.ucla.edu, kfukami1@g.ucla.edu

Kenta Inada, Daisuke Umehara, Yoshimichi Ono, Kenta Ogawa
Honda Motor Co., Ltd.
kenta_inada@jp.honda, daisuke_umehara@jp.honda, yoshimichi_ono@jp.honda, kenta_ogawa@jp.honda

Kunihiko Taira
University of California, Los Angeles
Mechanical and Aerospace Engineering
ktaira@seas.ucla.edu

PP2

MS34: Identifying Influential Events for Pressure Injury Development: Applying Graph Analysis on Electronic Health Records

We delve into the challenge of identifying influential events that play crucial roles in pressure injury development. First, we construct temporal event graphs (networks) utilizing pressure injury patient data from MIMIC-III Electronic Health Records. As patient events, we include daily average of patient vitals/labs, and daily occurrence of interventions from MIMIC-Extract, a preprocessed dataset extracted from MIMIC-III. Second, we employ inverse percolation centrality, a centrality measure that evaluates the relative importance of patient events within a network by considering both their topological connectivity and their involvement in pressure injury development dynamics. Third, we utilize a control group of non-pressure injury patients to discard common influential events shared by both groups. For the control group, we match pressure injury patients with non-pressure injury patients using propensity score matching based on patient demographic information. Finally, we present our findings regarding pressure injury development, identifying influential events and pathways of pressure injury.

Altansuren Tumurbaatar
Emory University
altaamgl@gmail.com

Wenhui Zhang, Roy Simpson
Emory University
Nell Hodgson Woodruff School of Nursing
wenhui.zhang@emory.edu, roy.l.simpson@emory.edu

Joyce Ho
Emory University
joyce.c.ho@emory.edu

Vicki Hertzberg
Emory University
Nell Hodgson Woodruff School of Nursing Emory
vhertzb@emory.edu

PP2

Broader Engagement (BE): Pinns Based Equation Discovery for Plasma Physics Particle In Cell Simulations As Surrogate Models

Physics Informed Neural Networks (PINNS) over recent years have seen a surge of interest and use due to their seamless and efficient integration of observational data and physical models through the Partial Differential Equation (PDE) constraints. We present a PINNS based method and algorithm to infer or discover a PDE that models given

observational data and can be used as a surrogate model. We demonstrate the capabilities of our method by first learning a series of benchmark modeling PDEs and then show its application within the area of Plasma Physics by learning a governing PDEs for Particle In Cell (PIC) simulation code. The benefits of our equation discovery method are that a PDE can be learned that models a given observational data set; the PINNs model is trained that both models the data and adheres to the discovered PDE; and that the trained PINNS model or discovered equation can be used as surrogate model in place of a finer-scale and more expensive simulation model.

Erik Urzua
University of California, Irvine, U.S.
urzuaerik@gmail.com

PP2

MS27: Extensions of Consensus-Based Methods

One way that is becoming increasingly popular in recent times to solve minimization problems for non-convex problems are multi-particle metaheuristic derivative-free optimization methods. We will focus on the Particle Swarm Optimization (PSO) and the Consensus Based Optimization (CBO) methods, whose basic idea is to iteratively update a population of particles according to a consensus dynamic inspired by social interactions between individuals. We focus on the constrained problem in the scenario of the PSO method. We introduce a micro-macro decomposition of such algorithm, by writing the distribution function as the sum of an equilibrium part and a non-equilibrium part. Subsequently, we address an optimization problem where the cost function is the expectation of a random mapping by virtue of two approaches based on the CBO algorithm. We approximate the true objective function by two suitable samplings and the solve the newly obtained problem with the designated metaheuristics. Several numerical experiments show the validity of the proposed algorithms.

Sara Veneruso, Sabrina Bonandin, Michael Herty
RWTH Aachen University
veneruso@igpm.rwth-aachen.de, bonandin@eddy.rwth-aachen.de, herty@igpm.rwth-aachen.de

PP2

Training Techniques for Learning Neural ODE Models: Accurate Binary Black Hole Dynamical Models for Long-Time Integration and Extrapolation

One important astrophysical application of general relativity is solving the binary black hole problem. These equations predict that two black holes orbit one another and eventually collide, creating disturbances in spacetime that generate gravitational waves (GW). Except in the simplest possible cases, however, the ordinary differential equations that describe the motion of the two black holes are not known. Previous work showed that neural differential equations (NDE) can be used to discover these mechanical models from the GW signal by solving a physics-informed constrained optimization within a space of plausible models. This approach, however, requires solving potentially expensive ODEs multiple times throughout the training procedure. Our modified approach employs a feed-forward neural network (NN) trained in two stages. First, the NN is trained directly to approximate the ODEs right-hand side without considering any physics of the problem. This pre-trained NN can be refined by solving a physics-informed constrained optimization using waveform data. Preliminary results indicate successful training across various orbits, with errors nearing numerical round-off. This enhanced accuracy enables tackling new problem types. For example, we show that the resulting NDE accurately extrapolate to long time durations, be used close to the problem's separatrix, and can be applied to more complex dynamics such as zoom-whirl orbits where multiple distinct timescales appear.

Pranav Vinod
University of Massachusetts, Dartmouth
pvinod@umassd.edu

Scott Field
Mathematics Department, UMASS Dartmouth
Dartmouth, MA , USA
sfield@umassd.edu

Alfa Heryudono
University of Massachusetts Dartmouth
Department of Mathematics
aheryudono@umassd.edu

Zheng Chen
University of Massachusetts, Dartmouth
zchen2@umassd.edu

Brendan Keith
Brown University
brendan_keith@brown.edu

Brian Cornet
University of Massachusetts, Dartmouth
bcornet@umassd.edu

PP2

MS38: Exploring the Power of Persistent Spectral Graphs in Covid-19 Research

The ongoing COVID-19 pandemic has underscored the critical need for computational tools to interpret complex biological data and enhance our understanding of viral spread and treatment efficacy. This study presents a pioneering integration of computational topology and artificial intelligence (AI), showcasing a novel approach to epidemiological research. Specifically, we studied the capacities of persistent spectral graphs (PSGs) for analyzing intricate topological and geometric properties of high-dimensional biological data. Our approach begins with the formulation of persistent Laplacian matrices (PLMs), constructed through applying a dynamic filtration parameter. The harmonic spectra derived from the null spaces of these PLMs encapsulate the underlying topological features, while the non-harmonic spectra elucidate the geometry of high-dimensional datasets. During the COVID-19 pandemic, we integrated PSG, genomics, and deep learning into a Math-AI model to predict the binding free energy (BFE) changes caused by mutations in the interaction between the virus's Spike protein and the human ACE2 receptor or antibodies. Such a Math-AI model has successfully forecasted the predominance of Omicron variants one or two months ahead of their presence, offering a glimpse into a future where mathematical techniques and biology converge to combat viral threats more effectively.

Rui Wang
New York University

wang.rui@nyu.edu

PP2

MS26: Principal Curves in Wasserstein Space

A "principal curve" of a data distribution refers to a class of nonlinear generalizations of the first principal component: namely, we seek a continuous curve which locally "passes through the middle" of the distribution. Originally proposed in the context of statistics by Hastie and Stuetzle, the principal curve problem is closely related to the "average distance problem" in the calculus of variations. Our work introduces the principal curve problem in the Wasserstein space of probability measures. We relate this problem to the problem of finding minimizers to a length-penalized average-distance variational problem in this space; for said problem, we then prove existence of minimizers, stability with respect to the underlying data distribution, and consistency of a discretized variational problem. This latter problem enjoys a "coupled Lloyd's algorithm" type numerical scheme which can be feasibly implemented via existing Wasserstein barycenter solvers. Lastly, we apply this general framework to a concrete problem motivated by recent developments in measurement technologies for gene expression data. Namely, given a stochastic process and batches of samples from different temporal marginals, but without temporal labels on the batches, can we infer the temporal ordering of the batches? This is a version of the so-called seriation problem. We thus show that principal curves in Wasserstein space can be employed as a consistent seriation method for empirical distributions.

Andrew Warren, Anton Afanassiev, Forest Kobayashi
University of British Columbia
awarren@math.ubc.ca, anton.a@math.ubc.ca,
fkobayashi@math.ubc.ca

Geoffrey Schiebinger
The University of British Columbia
geoff@math.ubc.ca

PP2

MS37: Mean-Field Control Barrier Functions: A Framework for Real-Time Swarm Control (Poster)

Control Barrier Functions (CBFs) are an effective methodology to assure safety and performative efficacy in real-time control applications such as power systems, resource allocation, autonomous vehicles, robotics, etc. This approach ensures safety independently of the high-level tasks that may have been pre-planned off-line. For example, CBFs can be used to guarantee that a self-driving car will remain in its lane. However, when the number of agents is large, computation of CBFs can suffer from the curse of dimensionality in the multi-agent setting. In this work, we present Mean-field Control Barrier Functions (MF-CBFs), which extends the CBF framework to the mean-field (or swarm control) setting. The core idea is to model swarms as probability measures in the state space and build control barrier functions in the space of probability measures. Similar to traditional CBFs, we derive safety constraints on the (distributed) controls but now relying on the differential calculus in the space of probability measures. Our numerical experiments show the effectiveness of MF-CBFs applied to swarm tracking and avoidance.

Samy Wu Fung
Colorado School of Mines
swufung@mines.edu

Levon Nurbekyan
Department of Mathematics
Emory University
levon.nurbekyan@emory.edu

PP2

Broader Engagement (BE): Extending Statistics-Informed Neural Network to Multidimensional Stochastic Processes

The statistics-informed neural network (SINN) has been proposed as a machine-learning-based stochastic trajectory generator. SINN is built on a deterministic recurrent neural network that takes an input stream of white noise sequences to generate an ensemble of stochastic trajectories that has statistical properties similar to the original stochastic dynamics. While the capabilities of SINN have been demonstrated using various one-dimensional non-Markovian processes, for this methodology to be a promising tool for real applications (e.g., via surrogate modeling), it is required to extend SINN to reproduce multidimensional stochastic processes. In this presentation, we present our efforts in this regard. First, since many cross terms are to be included in the total loss function to capture the complex statistical properties of a multidimensional process, we employ a self-adaptive loss-balanced technique to effectively balance various loss terms. Second, since the estimation of a multidimensional probability density function (PDF) is computationally inefficient, instead of adding these to the total loss function, we include a set of one-dimensional PDFs of some linear combinations of component processes. Third, we investigate the impact of the dimensionality of the input white noise sequence on the performance of SINN. Lastly, we perform simulation studies using the Langevin dynamics of the Fermi-Pasta-Ulam chain and the kinetic Monte Carlo simulation of surface chemistry.

Zihan Xu, Zihan Xu
University of California Merced
zxu64@ucmerced.edu, zxu64@ucmerced.edu

Changho Kim
University of California, Merced
Applied Mathematics
ckim103@ucmerced.edu

Andy Nonaka
Lawrence Berkeley National Laboratory
Center for Computational Sciences and Engineering
ajnonaka@lbl.gov

Yuanran Zhu
Lawrence Berkeley National Laboratory
yzhu4@lbl.gov

PP2

MS34: Enabling Pre-Shock State Detection Using Electrogram Signals from Implantable Cardioverter-Defibrillators

Early detection of ventricular arrhythmias in patients with implantable cardioverter-defibrillators (ICDs) remains a critical challenge in cardiac care. Our study introduces an innovative framework integrating metric learning, prototype learning, and few-shot learning to enhance the accuracy of predicting ICD shocks. We analyzed electrogram (EGM) data from 326 participants, focusing on EGM seg-

ments recorded before shock events and during normal conditions. Our approach utilizes a Siamese neural network equipped with LSTM units, which utilizes both triplet and pair losses to refine the learning process. This strategy allows the network to generate distinct embeddings that accurately distinguish between pre-shock and normal EGM signals. For classification, we adopted k-means clustering to identify prototypes representing the two states, enhancing our model's ability to predict impending arrhythmias. The evaluation of our framework showed a robust performance, achieving an F1 score of 0.87, a sensitivity of 0.97, and a precision of 0.79. This work not only advances predictive analytics in the cardiac field but also supports improved clinical decision-making in ICD management, potentially enhancing patient safety and outcomes.

Runze Yan
Emory University
Nell Hodgson Woodruff School of Nursing
ryan30@emory.edu

Neal Bhatia
Emory University Hospital
neal.kumar.bhatia@emory.edu

Faisal Merchant
School of Medicine
Emory University
faisal.merchant@emory.edu

Alex Fedorov, Ran Xiao
Nell Hodgson Woodruff School of Nursing
Emory University
aleksandr.vladimirovich.fedorov@emory.edu,
ran.xiao@emory.edu

Cheng Ding
Department of Biomedical Engineering
Georgia Institution of Technology & Emory University
chengding@gatech.edu

Xiao Hu
Emory
Nell Hodgson Woodruff School of Nursing
xiao.hu@emory.edu

PP2

MS33: Optimal Score Estimation Via Empirical Bayes

We study the problem of estimating the score function of an unknown probability distribution ρ^* from n independent and identically distributed observations in d dimensions. Assuming that ρ^* is subgaussian and has a Lipschitz-continuous score function s^*, we establish the optimal rate of $\tilde{\Theta}(n^{-\frac{2}{d+4}})$ for this estimation problem under the loss function $\|\hat{s} - s^*\|^2_{L^2(\rho^*)}$ that is commonly used in the score matching literature, highlighting the curse of dimensionality where sample complexity for accurate score estimation grows exponentially with the dimension d. Leveraging key insights in empirical Bayes theory as well as a new convergence rate of smoothed empirical distribution in Hellinger distance, we show that a regularized score estimator based on a Gaussian kernel attains this rate, shown optimal by a matching minimax lower bound. We also discuss the implication of our theory on the sample complexity of score-based generative models.

Andre Wibisono
Yale University
andre.wibisono@yale.edu

Yihong Wu
Department of Statistics and Data Science
Yale University
yihong.wu@yale.edu

Kaylee Yang
Yale University
yingxi.yang@yale.edu

PP2

MS35: Nonlinear Filtering in Stochastic Dynamical Systems

We present a nonlinear filtering approach for solving inverse problems sequentially in stochastic dynamical systems. By utilizing a Bayesian framework, we construct a generative modeling method to identify the state trajectories from noisy data in a robust manner.

Nicole Yang
Emory University
tianjiao.yang@emory.edu

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

PP2

MS37: Forward-Backward Algorithm for Functions with Locally Lipschitz Gradient: Applications to Mean Field Games.

In this paper, we provide a generalization of the forward-backward splitting algorithm for minimizing the sum of a proper convex lower semicontinuous function and a differentiable convex function whose gradient satisfies a locally Lipschitztype condition. We prove the convergence of our method and derive a linear convergence rate when the differentiable function is locally strongly convex. We recover classical results in the case when the gradient of the differentiable function is globally Lipschitz continuous and an already known linear convergence rate when the function is globally strongly convex. We apply the algorithm to approximate equilibria of variational mean field game systems with local couplings. Compared with some benchmark algorithms to solve these problems, our numerical tests show similar performances in terms of the number of iterations but an important gain in the required computational time.

Xianjin Yang
California Institute of Technology
yxjmath@caltech.edu

Francisco J. Silva
Univesite de Limoges
francisco.silva@unilim.fr

Luis Briceno-Arias
Univ. F. Santa Maria, Santiago, Chile

luis.briceno@usm.cl

PP2

MS22: Mean-Field Variational Inference Via Wasserstein Gradient Flow

Variational inference, such as the mean-field (MF) approximation, requires certain conjugacy structures for efficient computation. These can impose unnecessary restrictions on the viable prior distribution family and further constraints on the variational approximation family. In this work, we introduce a general computational framework to implement MF variational inference for Bayesian models, with or without latent variables, using the Wasserstein gradient flow (WGF), a modern mathematical technique for realizing a gradient flow over the space of probability measures. Theoretically, we analyze the algorithmic convergence of the proposed approaches, providing an explicit expression for the contraction factor. We also strengthen existing results on MF variational posterior concentration from a polynomial to an exponential contraction, by utilizing the fixed point equation of the time-discretized WGF. Computationally, we propose a new constraint-free function approximation method using neural networks to numerically realize our algorithm. This method is shown to be more precise and efficient than traditional particle approximation methods based on Langevin dynamics.

Rentian Yao
University of Illinois at Urbana-Champaign
University of British Columbia
rentian2@illinois.edu

Yun Yang
UIUC
USA
yy84@illinois.edu

PP2

MS31: Toward Foundation Material Model with Nonlocal Attention Operator

Unlike traditional machine learning tasks, physical systems in real-world applications feature scarce measurements and changing hidden states, and therefore calls for an automated self-adaptive mechanism capable of evolving continuously with the current state of system. To this end, we consider the learning of material responses as an exemplar complex physical system modeling problem, and investigate the self-adaptive capability of attention mechanisms toward developing a foundation material model. Specifically, we first show that the attention mechanism is equivalent to a nonlocal neural operator with data-dependent kernels, such that enables the automatic updating of the evolving physical properties via the attention matrix. We then propose Nonlocal Attention Operator (NAO), which provides a forward model in the form of nonlocal constitutive law, ensuring adherence to fundamental physical laws. In a wide variety of scientific applications, including constitutive modeling of material deformation, stress wave propagation, and digital twin modeling, we empirically demonstrate the advantages of NAO over the baseline neural operators and state-of-the-art methods in capturing the change of hidden physical states.

Yue Yu
Department of Mathematics, Lehigh University
yuy214@lehigh.edu

Ning Liu
GEM.inc
ningliu@umich.edu

Tian Gao
IBM Thomas J. Watson Research Center
tgao@us.ibm.com

Siavash Jafarzadeh
Lehigh University
sij222@lehigh.edu

Stewart Silling
Sandia National Laboratories
sasilli@sandia.gov

PP2

MS25: Dynamic Pattern Formation via Distribution Control of Moment Kernelized Population Systems

Targeted coordination of large-scale populations of dynamical systems for tuning patterns of their measurement data is an emerging and essential task in science and engineering. Its application domains span across numerous disciplines, ranging from motion planning of robot swarms in robotics and synchronization of rhythmic networks in network science to excitation of nuclear spin ensembles in quantum science. However, these dynamic pattern formation tasks are challenged by the lack of a principled formulation and the large population size of the systems. In this work, we propose a distribution control formulation for dynamic pattern formation. Specifically, by leveraging the technique of displacement interpolation in optimal transport theory, a dynamic pattern formation task is formulated as an optimal control problem over the space of probability distributions. We then develop the moment kernel transform, which enables the representation of the evolution of the measurement data pattern of a population system in terms of a dynamical system, referred to as a moment system, defined on a reproducing kernel Hilbert space. To control the moment system by using the data, we propose a deep neural network architecture, composed of multiple recurrent neural network layers, which are trained by using the transfer learning technique. The applicability and performance of the developed dynamic pattern formation framework are demonstrated using examples from practical applications.

Wei Zhang, Yi-Hsuan Shih, Jr-Shin Li
Washington University in St. Louis
wei.zhang@wustl.edu, shihy@wustl.edu, jsli@wustl.edu

PP2

MS32: Solving Linear Inverse Problems with Flows by Adaptive Corrupted Trajectory Matching

Flow-based generative models have attracted significant attention for their simplicity and superior performance in the fields of images, video, and speech. By leveraging the instantaneous changes of variables formula, one can compute the probability of an input image under the ODE framework of flow-based models. However, it remains unclear how to use this probability to solve inverse problems given some observed measurements. A major obstacle preventing the application of flow-based models to solve inverse problems is the slow computation of image probability, as the process requires multiple ODE propagations, not to men-

tion the gradient descent process. In our paper, we propose an adaptive corrupted trajectory matching method as an alternative to initial point optimization to expedite the solving process. Theoretically, we demonstrate that solving noisy MAPs at different timesteps is approximately equivalent to solving the global MAP. Extensive results on various inverse problems, including super-resolution, deblurring, and inpainting on datasets such as CelebAHQ, LSUN, and CT, demonstrate the excellence of our algorithm.

Yasi Zhang, Peiyu Yu, Yaxuan Zhu
UCLA
yasminzhang@ucla.edu, yupeiyu98@g.ucla.edu, yaxuanzhu@ucla.edu

Yingshan Chang
CMU
yingshac@andrew.cmu.edu

Feng Gao
Amazon
fenggo@amazon.com

Yingnian Wu
UCLA
ywu@stat.ucla.edu

Oscar F. Leong
University of California, Los Angeles
oleong@stat.ucla.edu

PP2

MS37: A Mean-Field Games Laboratory for Generative Modeling

We demonstrate the versatility of mean-field games (MFGs) as a mathematical framework for explaining, enhancing, and designing generative models. We establish connections between MFGs and major classes of flow and diffusion-based generative models by deriving continuous-time normalizing flows, score-based models, and Wasserstein gradient flows through different choices of particle dynamics and cost functions. Furthermore, we study the mathematical structure and properties of each generative model by examining their associated MFG's optimality condition, which consist of a set of coupled forward-backward nonlinear partial differential equations. The optimality conditions of MFGs also allow us to introduce HJB regularizers for enhanced training of a broad class of generative models. We present this framework as an MFG laboratory which serves as a platform for revealing new avenues of experimentation and invention of generative models.

Benjamin J. Zhang, Markos Katsoulakis
University of Massachusetts Amherst
bjzhang@umass.edu, markos@umass.edu

PP2

MS21: Learning Conditionally Independent Nonparametric Mixture Models Through Implicit Tensor Decomposition

We present an alternating least squares type numerical optimization scheme to estimate conditionally-independent mixture models in Rn, with minimal additional distributional assumptions. Following the method of moments, we tackle a coupled system of low-rank tensor decomposition problems. The steep costs associated with high-dimensional tensors are avoided, through the development of specialized tensor-free operations. Numerical experiments illustrate the performance of the algorithm, and its applicability to various models and applications. In many cases the results exhibit improved reliability over the expectation-maximization algorithm, with similar time and storage costs. We also provide some supporting theory, establishing identifiability and local linear convergence.

Yifan Zhang
University of Texas at Austin
yf.zhang@utexas.edu

Joe Kileel
University of Texas at Austin, U.S.
jkileel@math.utexas.edu

PP2

MS23: Exact Likelihood Score Estimation for Bayesian Sampling with Diffusion SDE under Gaussian Mixture Assumption

Diffusion models have recently gained popularity as potent generative models for Bayesian sampling, which typically involves adding the scaled likelihood score in its generative process. However, due to the analytical intractability of the exact likelihood score, traditional diffusion models often rely on very crude approximations, controlled by some hyperparameters, which can potentially lead to biases in the posterior samples. In this work, we derive an exact expression for the likelihood score under a Gaussian mixture assumption and propose an algorithm for iteratively estimating the likelihood score. Extensive numerical experiments demonstrate that the samples generated using the exact likelihood score from our method converge to the true posterior distribution.

Zezhong Zhang
Oak Ridge National Laboratory
zhangz2@ornl.gov

PP2

MS28: Efficient Two-stage Gaussian Process Regression via Subsampling and Automatic Kernel Searching

Gaussian Process Regression (GPR) is extensively utilized in statistics and machine learning to quantify predictive uncertainty. The efficacy of GPR critically depends on the precise specification of the mean, covariance functions, and associated hyperparameters. We firstly explore the impact of mean function misspecification and introduce a two-stage GPR model to mitigate this issue. Then we propose a novel algorithm, underpinned by a rigorous theoretical bound, to identify the most suitable kernel from a set of candidates. With well specified mean and kernel function, another potential misspecificationthe is from inexact training of GPR. Due to the cubic time complexity, hyperparameters of GPR are usually obtained via training GPR on a randomly sampled dataset. We theoretically and empirically demonstrate that parameters estimated from a subsampled dataset serve as effective initialization for exact GPR training, assuming correct model specification. Finally, we recommend two GPR approachesexact and scalable based on our frameworktailored to align with available computational resources and specific needs for uncertainty quantification. Our comprehensive evaluation across real-world datasets, including UCI benchmarks, along with a case study in a safety-critical medical application, high-

lights the robustness and accuracy of our methodologies.

Shifan Zhao
Emory University
shifan.zhao@emory.edu

PP3

MS47: Topological Neural Network and Deep Learning

Deep neural networks have revolutionized machine learning by leveraging vast amounts of data across various domains. While they excel in processing standard data like images and text, they face challenges when applied to structured scientific data in non-Euclidean domains. Geometric Deep Learning (GDL) extends the capabilities of deep learning to non-Euclidean domains by incorporating geometric principles. However, capturing non-local properties inherent in topological data remains a challenge. In this poster, I will introduce higher-order structures beyond traditional graph-based approaches, known as Topological Neural Networks (TNNs). TNNs offer a deeper understanding of complex data relationships. Moreover, I will discuss potential applications of TNNs, including drug discovery and social media analysis, and propose a novel approach (TopoX) inspired by classical topological concepts and recent advancements in higher-order networks. This framework aims to provide a more sophisticated representation of data in topological domains, pushing TNNs towards new frontiers. Relevant citations include seminal works in machine learning, geometric deep learning, and topological data analysis.

Ibrahem Aljabea
Louisiana State University (Baton Rouge, LA, US)
Louisiana State University (Baton Rouge, LA, US)
ialjab2@lsu.edu

Mustafa Hajij
Santa Clara University, U.S.
mhajij@SCU.edu

PP3

Ellipsoids Methods for Metric Entropy Computations

Historically, metric entropy has played a significant role in various domains of mathematics such as non-linear approximation theory, statistical learning theory, and empirical process theory. Recent advances in data science, and more specifically in deep learning theory, have led to renewed interest in the concept of metric entropy. However, computing the precise value of the metric entropy of a given function class turns out to be notoriously difficult in general; exact expressions are available only in very few simple cases. For this reason, it has become common practice to resort to characterizations of the asymptotic behavior of metric entropy only. Even this more modest endeavor has turned out daunting in most cases and standard results in the literature typically only characterize the leading term in the asymptotic behavior up to a multiplicative constant. We fill this gap by presenting a new general method for the precise characterization of the leading term in the asymptotic expansion of metric entropy for infinite-dimensional ellipsoids. We further argue that our results provide a unified framework for the derivation of the metric entropy of a wide variety of function classes, such as unit balls in Besov spaces, Modulation spaces, Sobolev spaces, and various classes of analytic functions, thereby retrieving and improving, sometimes significantly so, on standard results.

Thomas Allard
ETH Zürich
tallard@ethz.ch

Helmut Bölcskei
ETH Zurich
hboelcskei@ethz.ch

PP3

Sex-Structured Disease Transmission Model and Control Mechanisms for Visceral Leishmaniasis (vl)

Leishmaniasis are a group of diseases caused by more than 20 species of protozoan that are transmitted through the bite of a female sand fly. Several research has been conducted to propose disease control strategies. The mathematical models for the transmission dynamics of the disease studied so far did not consider the sex-biased burden of the disease into consideration. This study introduces a new deterministic sex-structured model for understanding the transmission dynamics of visceral leishmaniasis. Numerical simulations were performed using baseline parameter values, and scenario analysis was performed by changing some of these parameters as appropriate. Our numerical result shows that the implementation of disease-preventive strategies, as well as effectively treating the affected ones can significantly reduce the disease prevalence if applied to more proportion of the male population. The numerical simulation infers that a maximum of 60% of extra preventative measures targeted to only the male population considerably reduces the total prevalence of VL by 80%. It is also possible to decrease the total prevalence of VL by 69.51% when up to 50% of additional infected males receive treatment with full efficacy. Therefore, to reduce the disease burden of visceral leishmaniasis, public health officials and concerned stakeholders need to give more emphasis to the proportion of male humans in their intervention strategies.

Temesgen D. Awoke
Botswana International University of Science and Technology
Botswana International University of Science and Technology
temesgen.y2000@gmail.com

PP3

MS50: Stereographic Spherical Sliced Wasserstein Distance

Comparing spherical probability distributions is of great interest in various fields, including geology, medical domains, computer vision, signal processing, and deep representation learning. The utility of optimal transport-based distances, such as the Wasserstein distance, for comparing probability measures has spurred active research in developing computationally efficient variations of these distances for spherical probability measures. This paper introduces a high-speed and highly parallelizable distance for comparing spherical measures using the stereographic projection and the generalized Radon transform, which we refer to as the Stereographic Spherical Sliced Wasserstein (S3W) distance. We carefully address the distance distortion caused by the stereographic projection and provide an extensive theoretical analysis of our proposed metric and its rotationally invariant variation. Finally, we evalu-

ate the performance of the proposed metrics and compare them with recent baselines in terms of both speed and accuracy through a wide range of numerical studies, including gradient flows and self-supervised learning.

Yikun Bai, Soheil Kolouri
Vanderbilt University
yikun.bai@vanderbilt.edu, soheil.kolouri@vanderbilt.edu

PP3

Touching with Time Series Models to Predict Tourist Traffic

Planning for tourism at all levels of the sector, from the government to a single travel agency, depends on forecasting travel demand. The potential of forecasting to lessen losses resulting from differences in supply and demand is what makes it valuable. To meet the needs of visitors, tourist sites must have accurate projections of the amount of money that will be required in the future for services such as lodging, transportation, and retail establishments. The forecasting methodologies utilized in the study incorporate both trend and seasonal components in the data series, as seen by the trend and seasonal patterns seen in the quarterly visitor arrivals from India and the USA to Nepal. This study considers seasonality-aware four-time series forecasting models. They include the Naive Trend and Seasonal, Time Series Decomposition, Winters Exponential, and SARIMA models.

Dila Ram Bhandari
Tribhuvan University
dilabhandarig@gmail.com

PP3

Bayesian Calibration of the Rosenthal Model

Hot cracking prevents the additive manufacturing of a range of high-strength and temperature alloys. In [Kitt, L., Alexander, et. al: Additive Manufacturing, 73, 103675 (2023)], a hierarchical model that links process parameters, thermal features, and crack formation was proposed. The model uses a multi-source Rosenthal model calibrated against process monitoring to provide thermal features as an input to a simplified crack formation criterion based on the Rappaz-Drezet-Gremaud model. In this work, we apply the Bayesian calibration techniques originally proposed in [Kennedy, Marc C and O'Hagan, Anthony: Journal of the Royal Statistical Society B, 63, 425–464 (2001)] to build a surrogate model for the multi-source Rosenthal model, and also a model deficiency function that accounts for the discrepancy between the model output and the experimentally observed data. We use the Kennedy-O'Hagan model, to perform a probabilistic calibration of the thermal models proposed in Kitt2023. In the original work the analytic thermal models are calibrated using a grid search method which is computationally expensive, and does not quantify calibration parameter uncertainty. The surrogate model improves upon this, requiring only a few minutes to generate a posterior distribution for the calibration parameters and quantifying uncertainty, and achieving superior accuracy.

Jnanajyoti Bhaumik
University at Buffalo, State University of New York
jnanajyo@buffalo.edu

Luke Mohr, Alexander Kitt
Edison Welding Institute, Buffalo
lmohr@ewi.org, akitt@ewi.org

PP3

MS44: Predicting Appendicular Lean Mass, Bone Density and Body Fat Using P-Laplacian Based Semi-Supervised Learning

Nutritionists require accurate data on critical body characteristics such as Appendicular Lean Mass (ALM), Bone Density (BD), and Body Fat Percentage (BF) to conduct their research. Current methods to predict ALM, BD, and BF such as DEXA (Dual X-ray Absorptiometry) scans are not cost-effective. Previous studies have shown biomarkers such as waist circumference, arm length, and head circumference have correlations with ALM, BD, and BF. We applied semi-supervised learning algorithms to predict these three quantities using data on 40+ biomarkers of 700 patients provided by Pennington Biomedical Research Center or PBRC. Semi-supervised learning algorithms have been extensively used for classification and regression, especially when limited training data is available. An effective technique for exploiting unlabeled data in semi-supervised learning is to utilize a graph structure, which may be intrinsic to the data, or constructed based on similarities between data points. p- Laplacian is one of many graph-based semi-supervised learning techniques. This technique has only been used for classification problems, but a recent paper has shown it can also be used for regression. In this poster, we will elucidate the mathematical foundations behind p-Laplacian and demonstrate how we used it to predict ALM, BD, and BF by exploiting the similarity between the patients. We will also compare the results with those of supervised learning algorithms.

Alyssa Blount, Nadejda Drenska, Gyaneshwar Agrahari, Tristan Desoto, Kathryn Simon, Maganizo Kapita, Berendt Grant, Atif Iqbal
Louisiana State University
ablou14@lsu.edu, ndrenska@lsu.edu, gyan.agrahari77@gmail.com, tdesot3@lsu.edu, ksimo53@lsu.edu, jkapit1@lsu.edu, bgrand7@lsu.edu, aiqbal6@lsu.edu

PP3

MS36: An Exact Mathematical Description of Computation with Transient Spatiotemporal Dynamics in a Complex-Valued Neural Network

We study a complex-valued neural network (cv-NN) with linear, delayed interactions. We report the cv-NN displays sophisticated spatiotemporal dynamics, including partially synchronized chimera states. We then use these spatiotemporal dynamics, in combination with a nonlinear readout, for computation. The cv-NN can instantiate dynamics-based logic gates, encode short-term memories, and mediate secure message passing through a combination of interactions and time delays. The computations in this system can be fully described in an exact, closed-form mathematical expression. Finally, using direct intracellular recordings of neurons in slices from neocortex, we demonstrate that computations in the cv-NN are decodable by living biological neurons as the nonlinear readout. These results demonstrate that complex-valued linear systems can perform sophisticated computations, while also being exactly solvable. Taken together, these results open future avenues for design of highly adaptable, bio-hybrid computing systems that can interface seamlessly with other neural net-

works.

Roberto Budzinski, Alexandra Busch
Western University
rbudzins@uwo.ca, abusch5@uwo.ca

Lyle Muller
Department of Mathematics
Western University
lmuller2@uwo.ca

PP3

MS52: Model-Guided Experiment Design to Characterize Spatial Organization of Bacterial Metabolism

Bacteria are known to organize metabolism spatially through the encapsulation of enzymes, substrates, and cofactors in protein-based microcompartments. Microcompartments are of great interest in bioengineering for their ability to control local reaction environments and concentrate toxic intermediates, but key physical properties are unable to be measured directly. To probe key properties including permeability, we built a kinetic model and developed an experimental design method based on principal component analysis to inform sampling times of an experimental assay. Our method successfully constrains the parameter-of-interest, and parameter estimates match previous theoretical predictions. We additionally present preliminary results using experimental design to propose experiments which discriminate between equally parsimonious models in a sparse model selection framework.

Christina Catlett, Brett Palmero, Charlotte Abrahamson, Sasha Shirman, Carolyn Mills, Michael Jewett
Northwestern University
ccatlett@u.northwestern.edu,
brettpalmero2025@u.northwestern.edu,
cabrahamson@u.northwestern.edu,
shirman.sasha@gmail.com, carolynemills@ucsb.edu,
m-jewett@northwestern.edu

Niall M. Mangan
Engineering Sciences and Applied Mathematics
Northwestern University
niall.mangan@northwestern.edu

Danielle Tullman-Ercek
Northwestern University
ercek@northwestern.edu

PP3

MS49: Learning Nonnegative Matrix Factorization from Compressed Data

We study how to find a nonnegative matrix factorization (NMF) from compressed measurements. We consider methods of compression which can be adapted to the data, or can be oblivious. We formulate optimization problems that only depend on the compressed data, but which can recover a nonnegative factorization which closely approximates the original matrix. These optimization problems can be approached with a variety of algorithms, and in particular, we discuss variations of the popular multiplicative updates method for these compressed problems. We test our approaches on examples to validate performance in real-world applications. For the section "Efficient and robust optimization techniques for structured data learning".

Abraar Chaudhry
Department of Mechanical and Aerospace Engineering
Princeton University
azc@princeton.edu

Elizaveta Rebrova
Princeton University
elre@princeton.edu

PP3

Broader Engagement (BE): Optimal Sensor Placements in Gaussian Processes Using Column Subset Selection

In this work, we focus on optimal sensor placement which is then used in a Gaussian Process (GP) to reconstruct a spatially dependent function from a limited set of measurements. We develop an algorithm to select k sensors out of a much larger candidate set of m sensors such that the Bayesian D-optimality of the selected points is maximized. Our approach is to view sensor placement as a column subset selection problem (CSSP) on the kernel matrix of spatial data. We propose an algorithm that uses the Golub-Klema-Stewart approach (GKS) to select sensors which are then combined with a GP regression to reconstruct the function. We also provide an analysis of lower bounds on the D-optimality of these sensor placements using this approach. To reduce the computational cost in the GKS step, we will investigate low rank kernel methods such as Nystrm approximation and Randomly Pivoted Cholesky. We also compare our algorithm against a greedy algorithm in numerical experiments on interpolation problems and spatial temperature data.

Jessie Chen, Arvind Saibaba, Hangjie Ji
North Carolina State University
jchen228@ncsu.edu, asaibab@ncsu.edu,
hangjie_ji@ncsu.edu

PP3

MS51: Clustering Phylogenetic Analysis of Influenza Mutation Data

The impact of virus mutations on public health is profound, leading to increased infectivity, vaccine resistance, and alterations in disease severity. Annually, the CDC characterizes approximately 2,000 virus strains to monitor their evolution and confirm vaccine effectiveness. Grasping the essence of viral mutagenesis and evolution is paramount. Through the integration of genomic analysis, clustering, and dimensionality reduction methods, our study specifically investigates the influence of COVID-19 on influenza virus propagation. Our findings reveal that a robust dimension reduction and clustering approach can yield promising results in deciphering the complex dynamics of virus mutation, providing valuable insights for future research and public health strategies.

Jiahui Chen
Department of Mathematical Sciences
University of Arkansas
jiahuic@uark.edu

PP3

An Improved Analysis of Parallel Diffusion Gener-

ative Model Sampling

Recently, diffusion models have exerted huge impact on many tasks, including but not limited to generative artificial intelligence, scientific machine learning (AI4Science) and optimization. However, generation of the samples can be slow in many cases as diffusion models depend on an iterative sampling procedure. Among many studies focusing on how to accelerate the sampling of diffusion models, one approach is the ParaDiGMS algorithm, which leverages the Picard iteration to parallelize the sampling procedure. Despite the empirical success achieved by the ParaDiGMS algorithm, theoretical understanding of the parallelized sampling technique in the context of diffusion-based generative models is still lacking. Our theoretical analysis in this work reveals that the parallelized sampling technique would eventually bring about the first implementation of the diffusion model with O(polylog d) time complexity, where d denotes the dimension of the underlying data distribution.

Haoxuan Chen
Institute for Computational and Mathematical Engineering
Stanford University
haoxuanc@stanford.edu

Yinuo Ren
Stanford University
yinuoren@stanford.edu

Lexing Ying
Stanford University
Department of Mathematics
lexing@stanford.edu

Grant Rotskoff
Stanford University
rotskoff@stanford.edu

PP3

MS39: Improved Active Learning Via Dependent Leverage Score Sampling

We show how to obtain improved active learning methods in the agnostic (adversarial noise) setting by combining marginal leverage score sampling with non-independent sampling strategies that promote spatial coverage. In particular, we propose an easily implemented method based on the *pivotal sampling algorithm*, which we test on problems motivated by learning-based methods for parametric PDEs and uncertainty quantification. In comparison to independent sampling, our method reduces the number of samples needed to reach a given target accuracy by up to 50%. We support our findings with two theoretical results. First, we show that any non-independent leverage score sampling method that obeys a weak *one-sided ℓ_∞ independence condition* (which includes pivotal sampling) can actively learn d dimensional linear functions with $O(d \log d)$ samples, matching independent sampling. This result extends recent work on matrix Chernoff bounds under ℓ_∞ independence, and may be of interest for analyzing other sampling strategies beyond pivotal sampling. Second, we show that, for the important case of polynomial regression, our pivotal method obtains an improved bound of $O(d)$ samples.

Xiaoou Cheng
Courant Institute, New York University
chengxo@nyu.edu

Atsushi Shimizu, Christopher Musco
New York University
as15106@nyu.edu, cmusco@nyu.edu

Jonathan Weare
Courant Institute
weare@nyu.edu

PP3

Physics-Informed Machine Learning Algorithms As Pde Solvers in Fluid Flow Through Porous Media

The so-called Physics-Informed machine learning (ML) algorithms, which can incorporate the mathematical formulation of the modeled physics phenomena in their training process, have shown outstanding growth in the last years. Two interesting applications are possible for them: building data-free solvers for partial differential equations (PDE) and identifying the mathematical model behind a data set. This kind of ML techniques were tested in many problems, including the flow of fluids through porous media, generally for limited and very simplified scenarios. To evaluate their true capacity, more practical problems must be addressed. In this sense, one of the most useful and practical models in oil and gas reservoir evaluation is expressed as an advection-diffusion PDE called hydraulic diffusivity equation. This model was used to evaluate the limitations of the recently introduced physics-informed ML algorithms. Clear limitations were identified after testing different types of boundary conditions, which, from the physics side, are analogous to different types of reservoirs. The identification of failure scenarios allowed to successfully establish some Galerkin and regularization-based approaches to overcome the intrinsic limitations of such methods.

Luis Constante
LENEP
UENF
luisgamboa@lenep.uenf.br

Adolfo Pires
UENF Universidade Estadual Norte Fluminense Darcy Ribeiro
Laboratório de Engenharia e Exploração de Petróleo
LENEP
puime@lenep.uenf.br

PP3

MS36: Extracting Koopman-Based Features from Eeg Recordings of Brain Activity During Time-Based Decision Making

Discovering dynamical patterns from high fidelity, seconds-to-minutes timeseries is typically a challenging task. Furthermore, the problem increases in complexity when the timeseries themselves only represent partial measurements of some intrinsic dynamics of an unknown underlying dynamical system. In this poster presentation, the timeseries data consist of scalp electroencephalography (EEG) neural recordings obtained from the midfrontal Cz-electrode of healthy participants that performed an interval-timing cognitive task. In the experiment, the participants estimated a 7-second time-period by pressing a button. We implement a data-driven method, which leverages time-delayed coordinates, diffusion maps, and dynamic mode decomposition, to identify neural features in the EEG that correlated with the subjects' reports. The method involves extracting Koopman eigenfunctions from the data and re-

lating them to certain Kolmogorov backward eigenfunctions and to the trajectories of a hypothesized underlying stochastic differential equations system.

Rodica Curtu
University of Iowa
Department of Mathematics
rodica-curtu@uiowa.edu

Pake Melland
Oregon Institute of Technology
pake.melland@oit.edu

PP3

MS48: Exploring Policy Entropy in Reinforcement Learning for Personalization Environments

This talk delves into the intricate behavior of reinforcement learning systems operating within personalized environments, shedding light on the distinct policy entropy characteristics shaped by different learning algorithms. We elucidate that Policy Optimization agents frequently exhibit low-entropy policies during training, leading to a tendency to favor certain actions while neglecting others. In contrast, Q-Learning agents demonstrate a remarkable resilience to such behavior, maintaining consistently high-entropy policies throughout training, a trait highly desirable in practical applications. Through a comprehensive array of numerical experiments complemented by theoretical insights, we unveil the underlying reasons behind these entropy disparities attributed to the choice of learning algorithms. Our findings not only provide empirical evidence but also offer theoretical underpinnings, enriching our understanding of how learning algorithms influence policy entropy dynamics in reinforcement learning contexts. This presentation offers valuable insights for practitioners seeking to optimize policy entropy for enhanced adaptability and performance in personalized environments.

Anton Dereventsov
Lirio, LLC
adereventsov@lirio.com

Andrew Starnes
Lirio AI Research
astarnes@lirio.com

Clayton Webster
Lirio
cwebster@lirio.com

PP3

MS39: CS4ML: A General Framework for Active Learning with Arbitrary Data Based on Christoffel Functions

Active learning is an important concept in machine learning (ML), in which the learning algorithm can choose where to query the ground truth to improve the accuracy of the learned model. As ML tools come to be more commonly used in scientific computing, where data is often expensive to obtain, the use of active learning is expected to be particularly important. In this work, we introduce a general framework for active learning in regression tasks. Our framework extends the standard setup by allowing for general types of data, rather than just pointwise samples of the target function. This generalization covers many cases of practical interest, such as data acquired in transform domains (e.g., Fourier data), vector-valued data (e.g., gradient-augmented data), data acquired along continuous curves, and multimodal data (i.e., combinations of different types of measurements). Our framework considers random sampling according to a finite number of sampling measures and arbitrary nonlinear approximation spaces (model classes). We introduce the concept of generalized Christoffel functions and show how these can be used to optimize the sampling measures. We prove that this leads to near-optimal sample complexity in various important cases. Finally, we demonstrate the efficacy of our framework for gradient-augmented learning with polynomials, Magnetic Resonance Imaging (MRI) using generative models and adaptive sampling for solving PDEs using Physics-Informed Neural Networks (PINNs).

Ben Adcock
Simon Fraser University
ben_adcock@sfu.ca

Juan Cardenas
University of Colorado Boulder
juca7891@colorado.edu

Nick Dexter
Florida State University, U.S.
nick.dexter@fsu.edu

PP3

MS44: Association Between Socioeconomic Status and Cardiovascular Disease in the U.S.: A Deep Learning Approach

This study investigated the relationship between socioeconomic status (SES) and cardiovascular disease (CVD) prevalence in the United States from 2015-2022. Previous research has linked lower SES to higher CVD risk, and machine learning techniques have the potential to uncover novel patterns in this association over the 8-year period. We applied advanced machine learning and deep learning models to data from the Behavioral Risk Factor Surveillance System (BRFSS). The dataset comprised 3,485,529 observations with demographic, socioeconomic, health behavior, and CVD outcome variables. SES indicators included education level, income, employment status, and neighborhood characteristics. Deep neural networks, gradient-boosted trees, and other machine learning models were trained to predict CVD outcomes based on SES and additional variables. The most important SES predictors were education, income, and neighborhood wealth index. After controlling for other factors, the predicted CVD risk was 75% higher for individuals with less than a high school education compared to college graduates. Risk steadily decreased with higher income brackets. This large longitudinal study demonstrates a strong negative relationship between SES and CVD risk. Machine learning methods applied to multi-year survey data revealed potent socioeconomic disparities predictive of cardiovascular outcomes. These disparities persist in the U.S. despite a decline in overall CVD rates over the past decade.

Chandra Dhakal
University of Georgia, Athens
chandrakantdhakal@gmail.com

PP3

D3censent: A New Cnn Ensemble Network for White Blood Cell Classification with Lime Explain-

able

White blood cells (WBC) are crucial components of our immune system, defending our body against infections by destroying viruses, bacteria, parasites, and fungi. Understanding the variety and total count of WBCs can tell us a lot about our health. Convolutional neural networks (CNN), a type of deep learning, are used to identify and recognize blood cells by examining parts of an object. Various CNN models exhibit potential; however, their development often involves ad-hoc processes that neglect unnecessary layers, leading to issues with unbalanced datasets and insufficient data augmentation. To tackle these challenges, we suggest an innovative ensemble strategy that incorporates three CNN architectures, each customized with distinct dropout and maxpooling layers configurations to improve feature-level learning. As a result, our proposed ensemble network that is called D3CENSENT achieves an optimal bias-variance trade-off. Upon evaluation using the well-known Rabbin-WBC dataset, our model surpasses existing state-of-the-art networks by achieving a mean accuracy of 98.53%. It also excels in precision, recall, F1 score, and Area Under the ROC Curve (AUC) across each category. To delve deeper into the interpretability of classifiers, we employ reliable post-hoc explanation techniques, including Local Interpretable Model-Agnostic Explanations (LIME). These methods simplify understanding a black-box model by explaining how feature values are linked to predictions.

Sibasish Dhibar
Indian Institute of Technology Hyderabad
sdhibar@ma.iitr.ac.in

PP3

Vemos: A Gui for Evaluation of Similarity Metrics on Complex Data Sets

Similarity and dissimilarity metrics are a fundamental component of many tasks requiring the analysis and comparison of complex, often visual data. Applications ranging from computer vision to forensics require ways to effectively identify images, find clusters or outliers in data sets, or retrieve data items similar to a query item. However, finding an effective metric for a specific task is challenging due to the complexity of modern data sets and the myriad of possible similarity metrics arising from that complexity. We present VEMOS, a Python package that provides an accessible graphical user interface (GUI) for the evaluation of such comparison metrics. VEMOS provides user-friendly ways to examine individual data items or groups in a data set alongside analyses of metrics performance on the whole data set, such as clustering, multi-dimensional scaling, and retrieval performance analyses. VEMOS aims to help researchers and practitioners evaluate multiple comparison metrics (of similarity or dissimilarity) on rich, diverse data sets.

Gunay Dogan
National Institute of Standards and Technology
gunay.dogan@nist.gov

PP3

MS54: Embrace Rejection! Faster Low-Rank Approximation by Rejection Sampling-Accelerated Randomly Pivoted Cholesky

The randomly pivoted Cholesky (RPCholesky) algorithm is an effective method for producing a low-rank approximation to a positive semidefinite kernel matrix, leading to significant speed-ups for kernel machine learning algorithms. For even faster speed, one can use a blocked version of the RPCholesky algorithm, though the blocked version can produce inaccurate approximations for some examples. This raises the question: is there one algorithm that has the accuracy of ordinary RPCholesky and the speed of the block version? This talk answers this question in the affirmative by introducing accelerated RPCholesky, a rejection sampling-based version of RPCholesky that produces the exact same distribution of outputs as ordinary RPCholesky while taking advantage of the computational efficiencies of blocking. The effectiveness of accelerated RPCholesky is demonstrated using numerical experiments, theoretical results, and applications to scientific machine learning.

Ethan N. Epperly
California Institute of Technology
eepperly@caltech.edu

PP3

Broader Engagement (BE): Analysis of Reporting Trends in Russo-Ukrainian Conflicts Through Topic Modeling Algorithms

National concerns shift with time and can be presented through topics of national discussion. This research analyzes text data with a topic modeling algorithm (Latent Dirichlet Allocation) in order to determine prioritized topics present in news sources surrounding periods of conflict between Russia and Ukraine.

Nya Feinstein
West Virginia University, U.S.
nyafein@gmail.com

PP3

MS39: Agnostic Active Learning of Single Index Models with Linear Sample Complexity

We study active learning methods for single index models of the form $F(\mathbf{x}) = f(\langle \mathbf{w}, \mathbf{x} \rangle)$, where $f : \mathbb{R} \to \mathbb{R}$ and $\mathbf{x}, \mathbf{w} \in \mathbb{R}^d$. In addition to their theoretical interest as simple examples of non-linear neural networks, single index models have received significant recent attention due to applications in scientific machine learning like surrogate modeling for partial differential equations (PDEs). Such applications require sample-efficient active learning methods that are robust to adversarial noise. I.e., that work even in the challenging agnostic learning setting. We provide two main results on agnostic active learning of single index models. First, when f is known and Lipschitz, we show $\tilde{O}(d)$ samples collected via statistical leverage score sampling suffice to learn a near-optimal model. Leverage score sampling is simple, efficient, and widely used for learning linear models. Our result needs no assumptions on data distribution, is optimal up to log factors, and improves quadratically on a recent $O(d^2)$ bound. Second, we show $\tilde{O}(d)$ samples suffice even when f is unknown. Our results use tools from high dimensional probability, including Dudley's inequality and dual Sudakov minoration, and a novel, distribution-aware discretization of Lipschitz functions.

Aarshvi Gajjar, Christopher Musco
New York University
ag7378@nyu.edu, cmusco@nyu.edu

Xingyu Xu

Carnegie Mellon University
xingyuxu@andrew.cmu.edu

Chinmay Hegde
New York University
chinmay.h@nyu.edu

Wai Ming Tai, Yi Li
Nanyang Technological University
waiming.tai@ntu.edu.sg, yili@ntu.edu.sg

PP3

MS42: Discrete Stochastic Interpolant

We introduce the discrete stochastic interpolant, a novel method for interpolating between two probability distributions on a discrete space. This technique facilitates sampling from an unknown target distribution. We demonstrate the application of our method, particularly in the context of natural language generation.

Wei Guo, Yuchen Zhu
Georgia Institute of Technology
wei.guo@gatech.edu, yzhu738@gatech.edu
¡yzhu738@gatech.edu¿

Molei Tao, Molei Tao
School of Mathematics, Georgia Institute of Technology
mtao@gatech.edu, mtao@gatech.edu

Yongxin Chen
Georgia tech
yongchen@gatech.edu

PP3

MS55: Learning from High-Dimensional Stochastic Dynamics

We develop an innovative learning framework that incorporates the noise structure to infer the governing equations from observation of trajectory data generated by stochastic dynamics. Our approach can proficiently capture both the noise and the drift terms. Furthermore, it can also accommodate a wide range of noise types, including correlated and state-dependent variations. Moreover, our method demonstrates scalability to high-dimensional systems when combined with popular deep learning approaches. Through extensive numerical experiments, we showcase the exceptional performance of our learning algorithm in accurately reconstructing the underlying stochastic dynamics.

Ziheng Guo
University of Houston, U.S.
henryguoziheng@gmail.com

Ming Zhong
Illinois Institute of Technology, U.S.
mzhong4@uh.edu

Igor Cialenco
Illinois Institute of Technology
cialenco@iit.edu

PP3

Broader Engagement (BE): Streptomyces Species and Antibiotic Biosynthesis Genes in Plant Root Microbiomes

The soil is an environment inhabited by a large, diverse population of bacteria, many of which can produce antibiotics that can be harmful or beneficial to plants. Most antibiotics derive from actinomycetes, especially Streptomyces, which produce about 80% of known antibiotics. To assess the prevalence of antibiotics and understand their significance in the rhizosphere, soil samples were collected from both Spelman College and Oxford College and the DNA analyzed via shotgun metagenomic sequencing. Soil samples were utilized to grow Arabidopsis thaliana plants to generate metagenomic data from plant roots. Taxonomic diversity of root and soil samples were examined (alpha and beta diversity) and the relative abundance of Streptomyces was visualized. Differential abundance testing using negative binomial across soil and root microbiomes in regards to Actinobacteria, samples indicated higher abundance of several Streptomyces and Actinoplanes species in the roots in comparison to the soil, especially the pathogen S. scabiei. DIAMOND alignments of genes involved in antibiotic production revealed numerous genes conferring production of DAPG, tirandamycin, vicenistatin, nystatin, bottromycin, concanamycin, and kanamycin abundant in roots. Future work includes evaluating bacterial isolates and their respective metabolic pathways to confirm antibiotic production. Overall, Streptomyces species and other pathogenic bacteria significantly affect the soil microbiome and growth of plants.

Tanjanay Hardy
Michigan State University, U.S.
hardytan@msu.edu

Mentewab Ayalew
Spelman College, U.S.
mayalew@spelman.edu

PP3

MS51: Integrating Transformer, Autoencoder, and Circular Fingerprint Techniques with Spectral Graph Methods to Predict Scarcely Labeled Molecular Data

In molecular and biological sciences, due to the expensive and time-consuming nature of experiments, data sets are often small or scarcely labeled. Although transfer learning may help address the challenge of predicting proprties from these data sets, it requires the existence of a related large data set, which may be difficult to ensure. This poster reviews three graph-based models (utilizing Merriman-Bence-Osher (MBO) techniques) that the authors proposed to overcome this challenge. The models consist of graph-based modifications of the MBO scheme integrated with a home-made transformer, an autoencoder, and an extended-connectivity fingerprint algorithm, as well as a consensus method. The authors validated the proposed models on five benchmark data sets and provided thorough comparisons to other competing methods, such as support vector machines, random forests, and gradient boosting decision trees. Residue-similarity (R-S) scores and R-S indices were also used to analyze the performance of the proposed methods. Extensive computational experiments and theoretical analysis demonstrated that the new models perform very well even when as little as 1% of the data set is used as labeled data.

Nicole Hayes
Michigan State University
Department of Mathematics

hayesni4@msu.edu

Ekaterina Merkurjev
Department of Mathematics and CMSE
Michigan State University
merkurje@msu.edu

Guo-Wei Wei
Michigan State University
weig@msu.edu

PP3

Applying Machine Learning to Continuous Monitoring Data for Glucose Level Forecasting in Septic Patients

We present a comparative study to evaluate the accuracy of predictive models for forecasting glucose levels in septic patients with continuous monitoring data acquired from bedside devices. The study compares four state-of-the-art machine learning models based on transformer technologies and one dynamic linear (DLinear) model. Our findings demonstrate the superiority of the DLinear and PatchTST models in forecasting glucose levels for septic patients. These two models exhibit outstanding predictive accuracy for 15-minute forecasts (MMPE≈ 3% in PatchTST), excellent performance for 30-minute forecasts (MMPE ≈ 7.8% in DLinear), and clinically acceptable predictions up to 60 minutes ahead (MMPE ≈ 14.4% in DLinear), based on the prior 30 minutes of continuously acquired glucose data. Between these two competing models, the DLinear model delivers slightly superior performance. These predictive tools have the potential to be further refined and integrated into future digital twins for personalized patient care management. By leveraging the capabilities of machine learning and continuous monitoring data, our work paves the way for improved glycemic control and enhanced clinical outcomes in critically ill patients with diabetes mellitus.

Jianguo Hou
Department of Mathematics, University of South Carolina
Columbia, SC 29208, USA
jianguo@email.sc.edu

Xiang Cao
Beijing Computational Science Research Center
Beijing 100193, China
caoxiang@csrc.ac.cn

Dong Wang
Sichuan Academy of Medical Sciences
University of Electronic Science and Technology of China
wdisolation@163.com

Jue-Xia He
School of Medicine and Life Sciences
Chengdu University of Traditional Chinese Medicine
hejuexia@163.com

Jun Li
School of Mathematical Science, Tianjin Normal University
Tianjin 300387, China
nkjunli@foxmail.com

Hua Jiang
Sichuan Academy of Medical Sciences
University of Electronic Science and Technology of China
hua.jiang@traumabank.org

Qi Wang
University of South Carolina
qwang@math.sc.edu

PP3

MS50: Manifold Learning in the Presence of Symmetries: the G-Invariant Graph Laplacian

Graph-based manifold learning algorithms assume that data lie on or near a d-dimensional manifold M embedded in some high-dimensional Euclidean space; by using a kernel function to measure pairwise affinities of data points, such algorithms construct a graph Laplacian matrix from the data, whose eigenvalues and eigenvectors are then used for tasks such as dimensionality reduction, function representation and approximation, and denoising. In this work, we consider data sets whose data points satisfy an additional symmetry invariance assumption: given a compact Lie group G, for any g in G and data point x, the point gx resulting from the action of g on x is a valid data point (which is not necessarily in the data set, but can be added to it). A possible approach for exploiting this symmetry invariance of data is to construct a G-invariant graph Laplacian (G-GL) by analytically incorporating the pairwise affinities between all the pairs of points generated by the action of G on the data set. The G-GL converges to the Fokker-Planck operator on the data manifold M, with a significantly improved convergence rate compared to the standard graph Laplacian.

Paulina Hoyos
University of Texas Austin
paulinah@utexas.edu

Eitan Rosen
Tel-Aviv University
eitanrose@gmail.com

Joe Kileel
University of Texas at Austin, U.S.
jkileel@math.utexas.edu

Xiuyuan Cheng
Duke University
xiuyuan.cheng@duke.edu

Yoel Shkolnisky
Tel Aviv University
yoelsh@tauex.tau.ac.il

PP3

MS51: Analyzing Single Cell RNA Sequencing with Topological Nonnegative Matrix Factorization

Single-cell RNA sequencing (scRNA-seq) is a relatively new technology that has stimulated enormous interest in statistics, data science, and computational biology due to the high dimensionality, complexity, and downstream analysis. Due to the high dimensionality of the original data, dimensionality reduction is often employed as the first step. Nonnegative matrix factorization (NMF), in particular, offers a unique approach due to its interpretation of resulting low-dimensional components as meta-genes. Extensions, such as manifold regularization, have been utilized for scRNA-seq. However, such regularization only captures the data at a single scale and lacks multiscale consideration. In

this poster, we introduce persistent Laplacian regularized NMF, which captures the geometrical and topological information of the data at multiscale. Similar to persistent homology, a tool from topological data analysis, persistent Laplacian captures the topological invariants, namely the Betti numbers, as well as the homotopic shape of evolution. We validated our approach on real scRNA-seq data.

Yuta Hozumi
Georgia Institute of Technology
hozumiyu@msu.edu

PP3

MS42: Approximating Measures on Function Spaces: Transport and Truncation

We study the problem of computing approximations to probability measures on infinite dimensional spaces in the context of Bayesian inverse problems. We examine truncations of the likelihood and posterior that enable practical computation while preserving the structural properties of the prior. The error analysis for these approaches can be understood from the perspective of the perturbation analysis of Bayesian inverse problems, providing rigorous convergence guarantees and mesh invariance. Further, we examine the interplay between truncation and transport, elucidating interesting and useful structures that appear for the transport map from prior to posterior when the action of the likelihood is restricted to a finite dimensional subspace– these results are analogous to the representer theorem for Hilbert spaces but in a stochastic setting.

Alexander Win Hsu
University of Washington
owlx@uw.edu

Ricardo Baptista
Caltech
rsb@caltech.edu

Bamdad Hosseini
University of Washington
bamdadh@uw.edu

PP3

MS49: Analyzing Bold Signals with Neural Networks and the Koopman Operator for Neurodegenerative Disease Study

This study explores the application of neural networks (NNs) and the Koopman operator in the analysis of Blood Oxygen Level Dependent (BOLD) signals derived from functional Magnetic Resonance Imaging (fMRI). We propose a novel methodology that leverages NNs to learn the eigenmodes of various brain features and behaviors, which are critical in the characterization of neurodegenerative diseases. By integrating the predictive power of deep learning with the dynamical systems perspective offered by the Koopman operator, our approach not only enhances the understanding of BOLD signals but also holds potential for early identification and study of neurodegenerative conditions. Preliminary results demonstrate the efficacy of our method in capturing significant patterns that correlate with the progression of these diseases, promising a new avenue for neuroscientific research and clinical diagnosis.

Mengqi Hu
University of North Carolina at Chapel Hill
humengqi@unc.edu

Guorong Wu
UNC Chapel Hill
guorong_wu@med.unc.edu

Xiu Yang
Lehigh University
xiy518@lehigh.edu

Yifei Lou
University of North Carolina at Chapel Hill
yflou@unc.edu

PP3

MS44: Stroke Rehabilitation with Machine Learning-Based Residual Severity Classification

Stroke therapy is essential to reduce impairments and improve motor movements by engaging autogenous neuroplasticity. This study uses supervised learning methods to address a clinicians autonomous classification of stroke residual severity labeled data towards improving in-home robotics-assisted stroke rehabilitation. Thirty-three stroke patients participate in in-home therapy sessions using the Motus Nova robotics rehabilitation technology to capture upper and lower body motion. The therapy session summary data is based on high-resolution movement and assistance data and clinician-informed discrete stroke residual severity labels. We demonstrate that the light gradient boosting method provides the most reliable autonomous detection of stroke severity. This method achieved an average of 94% accuracy, measured using the F1-score performance measure with 10-fold cross-validation. We show how objectively measured rehabilitation training paired with machine learning methods can be used to identify the residual stroke severity class with efforts to enhance in-home self-guided, individualized stroke rehabilitation. As data from rehabilitation practices are often of comparable size and nature to the data collected in our study, this suggests that the light gradient boosting method should be considered a standard, more efficient tool for this analysis.

Russell Jeter, Raymond Greenfield
Georgia State University
russell.jeter@outlook.com, russell.jeter@outlook.com

Stephen Housley
Georgia Institute of Technology
russell.jeter@outlook.com

Igor Belykh
Department of Mathematics and Statistics
Georgia State University
ibelykh@gsu.edu

PP3

MS45: Neural Network Prediction of Covid-19 Daily Infection Count

It is well known that the number of confirmed infection cases of COVID-19 is much lower than the true cases. The undercount ratio (the ratio of true cases to confirmed cases) depends on the testing effort. In this poster, we use an artificial neural network to uncover the hidden connection between testing data, confirmed caseload, and the true caseload. The output of our training set, i.e., the true case count, is derived from solving a deconvolution problem using both the death count and the infection fatality

rate (IFR). Multiple factors, such as age distribution, variants, and vaccination, are considered in estimating the time series of IFR.

Ning Jiang, Charles Kolozsvary
Department of Mathematics and Statistics
University of Massachusetts Amherst
ningjiang@umass.edu, ckolozsvary@umass.edu

Yao Li
University of Massachusetts Amherst
Department of Mathematics and Statistics
liyao@umass.edu

PP3

Learning Models for Real-World Dynamical Systems from Multifidelity Data

Accurate modeling of dynamical systems is crucial for supporting operations and decision-making in real-world applications. However, obtaining high-fidelity (HF) data for training these models is often challenging. Conversely, more accessible low-fidelity (LF) data may not adequately capture complex nonlinear dynamics. We present a multifidelity (MF) GPR framework, integrating nonlinear autoregressive schemes to exploits correlation across heterogeneous data sources, allowing for the implicit derivation of functional relationships and uncertainty quantification in predictions. We specifically apply and evaluate this framework for predicting the dynamics of a floating structure in maritime environment. The data comprises both LF numerical simulations and HF in-situ sensor measurements. The initial step of our MF framework involves developing a GPR surrogate model trained on LF data, efficiently replacing time-consuming numerical solvers. Given the high dimensionality of this data, we apply PCA for dimension reduction, which may result in information loss. To assess this impact, we conducted a comparative study between the PCA-based GPR model and a GCN model that processes multidimensional data without reduction. Our findings suggest that while PCA introduces errors, our framework effectively corrects these inaccuracies, ensuring reliable predictions. The MF GPR model captures intricate nonlinear behaviors and aligns closely with noisy sensor data.

Eirini Katsidoniotaki
Massachusetts Institute of Technology
Massachusetts Institute of Technology
eirka289@mit.edu

PP3

MS36: Patched Brain Transformer: Flexible Model for Supervised Pre-Trained Eeg Decoding

We propose Patched Brain Transformer, a novel attention-based model designed for decoding EEG recordings. The model exhibits flexibility concerning the number of EEG channels and recording duration, enabling effective pre-training across diverse datasets. We incorporate the inspection of the architecture to illustrate its behavior during training. Through supervised pre-training, regularization, coupled with time shifts as data augmentation, we achieve state-of-the-art accuracy in multi-participant classification of motor imagery datasets.

Timon Klein
Otto von Guericke University Magdeburg
timon.klein@ovgu.de

Piotr Minakowski
Otto-von-Guericke-Universität Magdeburg, Germany
piotr.minakowski@ovgu.de

Sebastian Sager
Otto-von-Guericke Universität Magdeburg
sager@ovgu.de

PP3

Towards Autonomous Laboratories by Integrating IoT, Robotics, Computer Vision, and AI with Container Technology

Integration of IoT edge devices, robotics, computer vision, and AI is vital for the realization of autonomous laboratories that can independently perform experiments, analyze data, and make decisions to enhance efficiency and productivity across various domains. Container technology emerges as a crucial facilitator in achieving this integration. Our current research explored how to use lightweight containers in deploying CPU-based computer vision apps on edge devices while evaluating various container technologies and images on ARM-based edge devices. Accordingly, we developed an optimized OpenCV container, and showed how integrating containers boosts efficiency in AI vision applications requiring effective real-time data stream processing. However, configuring real-time stream processing systems in the edge setting becomes increasingly challenging with the growing number of edge devices and their interactions, compounded by resource constraints and multi-tenant hosting complexities. We are currently conducting performance benchmarking and analysis of Apache Storm, a representative stream processing engine, on a cluster of ARM-based edge devices including Raspberry Pi 4 and Nvidia Jetson Nano. Furthermore, we are developing an LLM agent-based automation tool for experimental setup, debugging configuration errors, and analyzing the root cause of performance issues when running advanced data science algorithms that need to process vast amount of visual and IoT sensory data.

Turgay Korkmaz
The University of Texas at San Antonio
The University of Texas at San Antonio
korkmaz@cs.utsa.edu

Osamah I. Alqaisi, Kang Peng
The University of Texas at San Antonio
osamah.alqaisi@my.utsa.edu, kang.peng@utsa.edu

ALI S. Tosun
The University of North Carolina at Pembroke
ali.tosun@uncp.edu

PALDEN Lama
The University of Texas at San Antonio
palden.lama@utsa.edu

PP3

MS36: An ML Approach to Inferring Network Connectivity for Coupled Stochastic Oscillator Networks

Large-scale brain oscillations may reflect the synchronous behavior of neuron populations, yet the mechanisms underlying collective neuron dynamics are not well understood. One approach is modeling neuron populations as systems of oscillators: ODEs with stable limit-cycle solutions. How-

ever, neural activity is noisy, necessitating the study of coupled stochastic oscillators. Koopman operator spectral methods provide a universal description of stochastic oscillators [Perez et al 2023 PNAS]. However, the structure of the low-lying spectra for coupled stochastic oscillators remains an important open question. It is challenging to solve the related PDE for the Koopman modes; standard methods suffer from the curse of dimensionality and often yield inadequate results in dimensions $n \geq 4$. Recently, machine learning (ML) methods have proven effective for solving high-dimensional PDEs [Zhai et al 2022 PMLR]. Here, we derive a novel ML-based PDE method to compute the Koopman modes of coupled stochastic oscillators. For weakly coupled systems, our approach can generate eigenfunctions for a range of coupling strengths; we study the dependence of eigenfunctions of small oscillator networks on coupling strength and network connectivity. By performing a linear sensitivity analysis, our approach enables us to infer coupling strength from observed data. We argue that our ML-based approach can contribute to the analysis of large-scale electrophysiological recordings.

Max Kreider, Peter J. Thomas
Case Western Reserve University
mbk62@case.edu, pjthomas@case.edu

Yao Li
University of Massachusetts Amherst
Department of Mathematics and Statistics
liyao@umass.edu

PP3

MS49: The Simplicity Biases in Deep Linear Networks: Flatness, Invariance, and Edge of Stability

In this work, we investigate the training dynamics of deep linear networks with a large learning rate η, commonly used in machine learning practice for improved empirical performance. This regime is also known as the edge of stability, where the largest eigenvalue of the Hessian hovers around $2/\eta$, and the training loss oscillates yet decreases over long timescales. Within this regime, we observe an intriguing phenomenon: the oscillations in the training loss are artifacts of the oscillations of only a few leading singular values of the weight matrices within a small invariant subspace. Theoretically, we analyze this behavior based on the deep matrix factorization problem. Our analysis reveals that for η within a specific range, the oscillations occur within a 2-period fixed orbit of the singular values, while the singular vectors remain invariant across all iterations. Lastly, our analysis also reveals that deep linear networks initialized with small scales are implicitly biased towards solutions with a smaller trace of the Hessian, which may be of independent interest.

Soo Min Kwon
University of Michigan, Ann Arbor, U.S.
kwonsm@umich.edu

PP3

MS43: Data-Regularized Operator Learning for Inverse Problems

Regularization is a technique used in machine learning to prevent overfitting and embed prior information, and it is critical for solving ill-posed inverse problems. Deep learning methods have seen success in learning inverse problems, but they often regularize by adding penalty terms in the loss function or adapting the model architecture to better suit the problem. However, these methods of regularization can lead to increased computational costs during training and may be limited to a specific choice of model. We use an innovative approach known as the "data-regularized operator learning" (DaROL) method to regularize PDE inverse problems via transformation of the data. The DaROL method transforms the data in a way that aids the optimization of the neural network, and it can embed prior information without changing the loss function or the model architecture. This makes the DaROL method a simple and flexible framework for regularization. Here we explain the DaROL method and demonstrate how it regularizes ill-posed inverse problems for deep learning methods.

Jasen Lai
Purdue University
lai216@purdue.edu

Ke Chen
University of Maryland
kechen1993@gmail.com

Chunmei Wang
Department of Mathematics
University of Florida
chunmei.wang@ufl.edu

PP3

MS47: LLM Agents in Social Network Dynamics: A Study on Information Flow

Graph-based structures combined with Large Language Models (LLMs), which have demonstrated their capability to simulate human conversations, provide a powerful framework for simulating and analyzing dynamic social behaviors in networked systems. We develop such a framework that simulates conversations between LLM-powered Generative Agents and measures the information flow in the network using CEM (Conversation Evaluation Model). Each agent acts as a node in our network, characterized by distinct personas. These agents engage in simulated conversation based on the graph structure providing insights into how information propagates through various network topologies. Furthermore, it also reveals the inherent biases and the capabilities of LLMs in simulating realistic conversations. We analyze how different network structures and agent characteristics influence the spread and transformation of information. GNNs enhance our framework by enabling predictions of network behavior and the potential impact of various interventions. Our study opens new avenues for both theoretical research and practical applications in network science and social behavior analysis.

Bishal Lamichhane
University of Nevada, Reno
blamichhane314@gmail.com

PP3

MS36: Deep Neural Network Solver for Fokker-Planck Eigenfunctions

Many coupled neuronal oscillators are modeled by high-dimensional stochastic differential equations. It is well known that eigenvalues and eigenfunctions of the infinitesimal generator of the stochastic differential equation play important roles in stochastic stability, sensitivity analysis, and model reductions. However, eigenvalue problems

are difficult to compute, particularly in higher dimensional cases. In this poster, I will explain how to modify the neural network Fokker-Planck solver for the eigenvalue and eigenfunction problem. The key idea is to use a nonlinear regression to find a low-resolution Monte Carlo solution of eigenfunctions. Then the eigenfunction can be solved by a PINN-like neural network.

Yao Li
University of Massachusetts Amherst
Department of Mathematics and Statistics
liyao@umass.edu

PP3

MS49: Convergence and Complexity Guarantee for Inexact First-Order Riemannian Optimization Algorithms

We are interested in analyzing inexact Riemannian gradient descent (RGD) where Riemannian gradients and retractions are inexactly (and cheaply) computed. Our focus is on understanding when inexact RGD converges and what is the complexity in the general nonconvex and constrained setting. We answer these questions in a general framework of tangential Block Majorization-Minimization (tBMM). We establish that tBMM converges to an ϵ-stationary point within $O(\epsilon^{-2})$ iterations. Under a mild assumption, the results still hold when the subproblem is solved inexactly in each iteration provided the total optimality gap is bounded. Our general analysis applies to a wide range of classical algorithms with Riemannian constraints including inexact RGD and proximal gradient method on Stiefel manifolds. We numerically validate that tBMM converges faster than the classical algorithm when applied to low-rank matrix recovery, nonnegative tensor decomposition, and constrained matrix factorization. This poster is associated with the section entitled "Efficient and robust optimization techniques for structured data learning" (80608).

Yuchen Li
University of Wisconsin-Madison
yli966@wisc.edu

Laura Balzano
Electrical Engineering and Computer Science
University of Michigan
girasole@umich.edu

Deanna Needell
University of California, Los Angeles, U.S.
deanna@math.ucla.edu

Hanbaek Lyu
University of Wisconsin Madison
hlyu@math.wisc.edu

PP3

MS42: Measure Transfer Via Stochastic Slicing and Matching

Transporting and estimating probability measures are fundamental tasks in various generative modeling methods like normalizing flows. An equally crucial aspect is having a suitable metric to gauge model performance and guide algorithm design, particularly for scalability. Our focus lies on an iterative slice-matching approach initially introduced by Pitie et al. for transferring color statistics. While these techniques have proven effective in various data science applications, a comprehensive understanding of their convergence properties remains elusive. Our primary aim is to establish an almost sure convergence, shedding light on the connections with stochastic gradient descent in the context of the sliced-Wasserstein distance. We also explore recovery and stability under specific structural assumptions about the measures. This is joint work with Caroline Moosmueller.

Shiying Li
University of North Carolina - Chapel Hill
shiyl@unc.edu

Caroline Moosmueller
University of North Carolina at Chapel Hill
cmoosm@unc.edu

PP3

MS42: Robust and Efficient Differentiation for Regularized Optimal Transport

In this work, we consider the computation of derivatives for a bi-level optimization problem using regularized optimal transport (OT) distance, with applications including shuffled regression. We focus on the entropy-regularized OT distance and present an analytical expression for the second derivatives (Hessian) with respect to samples. Computing the Hessian poses significant challenges in terms of numerical stability and large memory consumption, especially for large-scale computations. To address the numerical instability arising from the ill-conditioned nature of the problem, we propose a theoretically guided stabilization method. We also derived a quantity that can serve as a surrogate to check the accuracy of the numerical approximation. Our numerical algorithm and the chosen hyperparameters are based on rigorous analysis of the properties of problems. We examine the efficiency and accuracy of the computed Hessian for various data sets. Finally, we apply the proposed algorithm to parameter estimation for shuffled regression of mixed Gaussians and 3D Point Cloud Registration of the MobilNet10 dataset.

Xingjie Helen Li
University of North Carolina, Charlotte, U.S.
xli47@uncc.edu

PP3

Major Or Minor

This project aimed to determine whether a song was major or minor with machine learning. To achieve this, we used a large dataset to have training data and testing data. I used about 80% of the data to train and 20% to test. This helps the model to be adequately ready to classify songs into their respective categories. I used data such as energy, loudness, and tempo to help the model determine a possible category. I used several different machine learning algorithms to determine the final classification. These models included a decision tree, knn learning, and neural network. The highest performing algorithm being the decision tree with a classification accuracy of 72%. In conclusion while the results could vary, a machine learning model can classify songs based on a major or minor key fairly accurately.

Hao Lin
Mira Costa

haogotmilk@gmail.com

PP3

MS43: DeepPropNet - A Recursive Deep Neural Network Propagator for Learning Evolutionary Pde Operators

In this poster, we propose a deep neural network ap- proximation to the evolution operator for time dependent PDE sys- tems over long time period by recursively using one single neural network propagator, in the form of POD-DeepONet with built-in causality feature, for a small time interval. The trained DeepPropNet of moderate size is shown to give accurate prediction of wave solutions over the whole time interval.

Lizuo Liu
Southern Methodist University
lizuo.liu@dartmouth.edu

PP3

MS45: Deep Neural Network Solver for the Wkb Approach to the Fokker-Planck Equation

Many biological systems are modeled by stochastic differential equations, with probability density functions governed by the Fokker-Planck equation. Although there are many recent advancements in the numerical solution of the Fokker-Planck equation, solving the weak noise Fokker-Planck equation remains challenging due to its highly concentrated solutions. We propose a deep-WKB framework for solving singular Fokker-Planck equations. Two neural networks are used to solve the Freidlin-Wentzell quasi-potential function and the pre-factor function separately. Then, two neural networks are re-trained for the Fokker-Planck equation itself.

Yicheng Liu
Mount Holyoke College
liu235y@mtholyoke.edu

PP3

MS58: Supervised Learning of Flow Map for Modeling Stochastic Dynamical Systems

We present a supervised learning technique to accurately learn unknown stochastic dynamical systems from their trajectory data. Our diffusion-based generative model is a training-free score estimation method that leverages the Monte Carlo estimator to directly approximate the score function. With the estimated score function, labeled data can be generated by solving an ordinary differential equation (ODE), corresponding to the reverse-time stochastic differential equation (SDE). The generated labeled data allows us to employ a simple fully connected neural network to learn the generative model in the supervised manner. Our proposed method is efficient in terms of training time and demonstrates high accuracy, as proven by a range of numerical experiments on stochastic systems.

Yanfang Liu
Oak ridge national laboratory
yanfang.liu@mtsu.edu

PP3

MS43: One-Shot Learning for Solution Operators of Partial Differential Equations

Learning and solving governing equations of a physical system, represented by partial differential equations (PDEs), from data is a central challenge in a variety of areas of science and engineering. Current methods require either some prior knowledge (e.g., candidate PDE terms) to discover the PDE form, or they need a large dataset to learn a surrogate model of the PDE solution operator. Here, we propose the first solution operator learning method that only requires one PDE solution, i.e., one-shot learning. We first decompose the entire computational domain into small domains, where we learn a local solution operator, and then we find the solution of a new input function via mesh-based fixed-point iteration (FPI), meshfree local-solution-operator informed neural network (LOINN) or local-solution-operator informed neural network with correction (cLOINN). We tested our method on 7 different PDEs, including linear or nonlinear PDEs, PDEs defined on complex geometries, and PDE systems. Our method demonstrates effectiveness and generalization capabilities across these varied scenarios.

Lu Lu
Yale University
lu.lu@yale.edu

PP3

Data-Driven Solutions for Neuroscience Challenges

In this talk, we will explore a few challenges in computational neurology that we addressed with data-driven techniques. We have applied inverse-problem techniques to identify the origins of neurodegeneration and leveraged deep learning approaches in medical imaging to gain better diagnostic insights, particularly for Alzheimers and Parkinson's diseases. Finally, we will demonstrate how we can accurately reconstruct Local Field Potential time series from rodents using the Eigensystem Realization Algorithm (ERA). This technique from the field of system identification allowed us to identify a high-dimensional linear dynamical system governing the observed data. The complexity of the time series, typically reflected in the dimensionality of the inferred system, provides a measure of brain complexity over time. These techniques hold potential for broader application across various electrophysiological recording modalities.

Pedro Maia
University of Texas at Arlington
pedro.maia@uta.edu

PP3

MS46: Membership Inference Attacks and Privacy in Topic Modeling

Recent research shows that large language models are susceptible to privacy attacks that infer aspects of the training data. However, it is unclear if simpler generative models, like topic models, share similar vulnerabilities. In this work, we propose an attack against topic models that can confidently identify members of the training data in Latent Dirichlet Allocation. Our results suggest that the privacy risks associated with generative modeling are not restricted to large neural models. Additionally, to mitigate these vulnerabilities, we explore differentially private (DP) topic modeling. We propose a framework for private topic modeling that incorporates DP vocabulary selection as a pre-processing step, and show that it improves privacy

while having limited effects on practical utility.

Nico Manzonelli
Harvard University
nicomanzonelli@gmail.com

Wanrong Zhang
School Of Engineering And Applied Sciences
Harvard
imwanrongz@gmail.com

Salil Vahdan
Harvard University
salil_vadhan@harvard.edu

PP3

MS55: Predictive Analysis of Opioid-Addiction Behavior Using Morphological Analysis of Astrocytes

Astrocytes, a subtype of glial cells with a complex star-shaped structure, are active players in many aspects of the physiology of the Central Nervous System (CNS). They reflect their diverse abilities and functions on their complex morphology and high heterogeneity. As a result, quantifying their morphological alterations is of great importance for understanding astrocytes role within the CNS in both the healthy and disease states. To provide a flexible computational platform for the analysis of images of astrocytes, we developed a software pipeline that automatically detects astrocytes and then extracts and analyzes single-cell morphological profiles. We applied this pipeline for the analysis of fluorescent images of astrocytes in the Nucleus Accumbens (NAc) a brain region associated with the reward system and, hence, of critical importance for the study of behavioral changes related to substance of abuse disorders. We applied two levels of analysis including (i) single-cell classification based on supervised learning and (ii) computation of morphological distances, defined in terms of Earth Movers Distance, between astrocyte subpopulations. Results of our study show that (i) single-cell morphological characteristics are predictive of astrocyte location within the NAc; (ii) single-cell morphological characteristics are predictive of behavioral responses to heroin induced withdrawal and relapse; (iii) single-cell morphological alterations are highly specific, in the sense that alterations observed in the heroin model are distinctly different from those observed in the sucrose control model. Our pipeline contributes to understanding astrocytes function within the CNS and their critical role in addiction-related behaviors, highlighting the importance of statistical and geometrical modeling in identifying changes in astrocyte structure that may indicate alterations in their functional capacities.

Michela Marini, Yabo Niu, Heng Zhao
University of Houston, U.S.
mmarini2@cougarnet.uh.edu, yniu4@uh.edu,
hzhao15@uh.edu

Anna Kruyer
University of Cincinnati, U.S.
kruyeraa@ucmail.uc.edu

Demetrio Labate
University of Houston
dlabate@math.uh.edu

PP3

Broader Engagement (BE): Less Sleep During the Wake Cycle After Repeated Closed Head Injury

Head injury can disrupt sleep; repeated head injury may produce compounding effects. However, the cumulative effects of repeated closed head injury on sleep remains unclear. In our prior study, repeated closed head injury every other day indicated a possible conditioning effect on subsequent head injuries and sleep. In this follow-up study, we hypothesized that either once or twice daily closed head injury would disrupt sleep patterns, particularly during the mouse wake cycle, and effect sizes would be cumulative with subsequent injuries. Male C57BL/6J mice (n=12) were randomly assigned to either once (ZT=3) or twice daily (ZT=3 & 6) weight drop head injury (Height: 94cm, Weight: 100g) daily for five consecutive days. Weight drop was administered, and mice were returned to non-invasive piezoelectric sensor sleep recording chambers. Sleep was recorded hourly from ZT=12 to ZT=2 and analyzed for sleep density, defined as sleep percentage per hour during the animals wake cycle. No cumulative effects on righting reflex times or sleep were observed across the week for single or repeated injury. The righting reflex times for the second daily injury were longer than the first injury of the day. A decrease in sleep density was seen at three days post-injury ($F(1, 20) = 52.816, p < 0.001$) compared to baseline. Sleep density was unaffected by injury number on day three ($F(1, 20) = 0.375, p = 0.547$). Further studies will explore injury parameters and interactions between repeated head injuries.

Kyli McQueen
University of Arizona College of Medicine
kylimcqueen@arizona.edu

PP3

MS54: Novel Randomized Algorithms for Low and Full-Rank Factorizations and Their Implementations in Randlapack

We present a pair of novel randomized algorithms: a scheme for performing QR with column pivoting and a method for performing a partial SVD. The first algorithm, called ICQRRP, carefully uses randomized sketching to accelerate both pivot decisions for the input matrix and the process of decomposing the pivoted matrix via Cholesky QR. ICQRRP is applicable to matrices of any aspect ratio. The second algorithm, called ABRIK, improves partial SVD runtime by performing a few multiplications with large blocks of random vectors. Furthermore, ABRIK overcomes the limitation of the power iteration-based Randomized SVD, related to matrices with a slowly decaying spectrum. We implement the algorithms in RandLAPACK by calling into RandBLAS and vendor-provided BLAS/LAPACK libraries. Experiments with these implementations were performed on an Intel Xeon Gold 6245 CPU. ICQRRP demonstrates two orders-of-magnitude speedup relative to LAPACKs standard function for QRCP and superior performance to the state-of-the-art alternative randomized QRCP scheme; ABRIK exhibits an order-of-magnitude speedup over the alternative partial SVD solvers.

Max Melnichenko
University of Tennessee
mmelnic1@vols.utk.edu

Riley Murray
Sandia National Laboratories
rjmurr@sandia.gov

Robert Webber
California Institute of Technology
rwebber@ucsd.edu

PP3

MS43: Solving Optimal Execution Problems in Financial Mathematics Via in-Context Operator Networks

Computing optimal order execution strategies is a fundamental problem in financial mathematics. The optimal execution strategy balances the trade-off between the cost of executing the order and market risk. A prominent approach proposed by practitioners is the propagator model, which employs an optimal control framework to derive these strategies. Yet, enhancing the speed of computing these strategies remains a vital area of research. Concurrently, in-context learning has proven effective in addressing large-scale practical problems. Within the domain of scientific computing, In-Context Operator Networks (ICONs) have emerged. These networks facilitate the learning of operators by merging offline pre-training with online inference, where the operator is approximated using a limited dataset provided as context. This approach eliminates the need for retraining when new contexts arise, offering benefits like rapid inference and reduced data demands. In our study, we apply ICONs to tackle optimal execution problems. Utilizing price dynamics as the context, ICONs establish a correlation between market prices and execution strategies. Subsequently, the optimal strategy is determined by solving the optimal control problem using the operator learned through this process.

Tingwei Meng
University of California, Los Angeles, U.S.
tingwei@math.ucla.edu

Moritz Voss
UC Los Angeles
Department of Mathematics
voss@math.ucla.edu

Liu Yang, Giulio Farolfi, Jona Lelmi
University of California, Los Angeles
liuyang@math.ucla.edu, gfarolfi@math.ucla.edu, jona.lelmi@math.ucla.edu

Nils Detering
University of California, Santa Barbara
nils.detering@hhu.de

Georg Menz
University of California, Los Angeles
georg.menz@gmail.com

Stanley J. Osher
UCLA
sjo@math.ucla.edu

PP3

MS54: The Girard-Hutchinson Estimator Is Bad at Kronecker Trace Estimation

We study the problem of estimating the trace of a matrix A that can only be access via Kronecker-matrix-vector products. That is, we can compute Ax for any vector x that has Kronecker structure. In particular, we study how Hutchinson's Estimator performs in this setting, proving tight rates for the number of matrix-vector products this estimator needs to find a relative error approximation to the trace of A. We find an exact expression for the variance of this estimator, show this is unavoidably exponential, and conclude with evidence that a much faster non-Hutchinson algorithm may exist.

Raphael A. Meyer
New York University
ram900@nyu.edu

Haim Avron
Tel Aviv University
haimav@tauex.tau.ac.il

PP3

MS49: Manifold Optimization for Data Driven Reduced-Order Modeling

The focus of this poster is on data-driven reduced-order modeling. We assume a high fidelity, discrete-time model is available for simulation. The simulator allows state and output trajectories to be collected for any specified initial condition and input signal. An optimal reduced-order model (ROM) requires: (i) the selection of a lower dimensional subspace for the state of the ROM, and (ii) an optimal reduced-order state-space model (evolving on the lower dimensional subspace). A common heuristic is to: (i) select the lower-order subspace using proper orthogonal decomposition (POD), and (ii) use least-squares to fit the reduced-order state-space model on the POD subspace. We demonstrate the potential deficiencies of this heuristic via two simple examples. In order to address these deficiencies, we propose a novel method to optimize the choice of subspace using the Grassmann manifold. Finally, we show that our proposed manifold optimization empirically outperforms the POD heuristic on the two motivating examples and a planar wind farm model. This submission is a part of the minisymposium on "Efficient and robust optimization techniques for structured data learning" (80608).

Rachel Newton, Zhe Du, Laura Balzano, Peter Seiler
Electrical Engineering and Computer Science
University of Michigan
rhoadesr@umich.edu, zhedu@umich.edu, girasole@umich.edu, pseiler@umich.edu

PP3

MS52: Machine Learning for Explicitly Solving Stiff, Parameterized Ordinary Differential Equations

Stiff ordinary differential equations present distinct challenges in various engineering disciplines. Specifically, these types of problems cannot be solved with traditional explicit time integration methods. Rather, costly implicit methods must be used to find solutions. Additionally, the timescale separation of various elements of the solution in stiff problems are handled poorly with modern machine learning techniques. Methods, such as Neural ODE (NODE), have proven to be insufficient for learning stiff systems. This work builds upon the NODE framework, to create surrogate models for stiff problems which can be solved using explicit time integration methods. Additionally, the approach can be extended to include additional input pa-

rameters, in the same fashion as the Parameterized Neural Ordinary Differential Equation approach proposed in [Lee, Parish, Parameterized Neural Ordinary Differential Equations: Applications to Computational Physics Problems, 2021]. This poster will demonstrate a new surrogate modelling approach that allows for the use of explicit methods to model stiff ordinary differential equations which have input-parameter-varying dynamics.

William C. Nockolds
University of Texas at Austin
cole.nockolds@utexas.edu

PP3

MS50: No-Collision Transportation Maps

We present a new class of transportation maps inspired by the no-collision properties of optimal transportation maps. In particular, we discuss the basic principles pertaining to such maps and some theoretical properties such as existence and regularity. Furthermore, we examine their relation with optimal transportation maps and discuss some manifold learning applications where no-collision maps perform at the level of optimal transportation maps or better at a cheaper computational cost. Finally, we discuss the drawbacks and challenges pertaining to no-collision maps and possible roads ahead.

Levon Nurbekyan
Department of Mathematics
Emory University
levon.nurbekyan@emory.edu

Elisa Negrini
University of California, Los Angeles
enegrini@math.ucla.edu

Adam M. Oberman
Department of Mathematics and Statistics
McGill University
adam.oberman@mcgill.ca

Alexander Iannantuono
Softree Technical Systems
alexander@iannantuono.org

PP3

Broader Engagement (BE): Compatible Energy Preserving Discretizations for Nonlinear Optical Wave Propagation: The Maxwell-Duffing Approach.

We explore the modeling and numerical discretization of Maxwell's equations in nonlinear optical media, specifically focusing on the Maxwell-Duffing model. We present the constitutive laws governing electromagnetic wave propagation in non-magnetic, non-conductive media, describing the material's response using a nonlinear cubic Duffing model coupled with Maxwell's equations.We introduce a high-order spatial discretization method based on fully discrete leap-frog finite-difference time-domain (FDTD) methods, designed for the accurate and stable simulation of nonlinear wave propagation. Special attention is given to the implementation of the Second order in time and higher order in space leap-frog scheme and its application to traveling wave solutions. We prove Energy Stability of the Higher Order Yee FDTD Schemes for the cubic Maxwell-Duffing Model and demonstrate these results through Numerical Simulations.This work provides critical insights into the mathematical and computational challenges of modeling nonlinear optical materials, offering robust techniques for advancing research in nonlinear photonics.

Victory Obieke
Oregon State University, U.S.
obiekev@oregonstate.edu

PP3

MS45: Applications of Disease Informed Neural Networks That Includes Human Behavior

In this work, we present modeling and simulation of disease dynamics through Physics-Informed Neural Networks (PINNS) and its application to real data modeled using non-linear system of differential equations. Specifically, we apply PINNS to predict the behavior of diseases described by modified compartmental models that include parameters and variables associated with the governing system describing the dynamics of the disease. Through benchmark problems, we will show that our model validates real-data and demonstrate how PINNs can predict optimal parameters for a given dataset.

Alonso G. Ogueda
George Mason University, U.S.
aogueda@gmu.edu

Padmanabhan Seshaiyer
Mathematical Sciences, George Mason University
pseshaiy@gmu.edu

PP3

Real Time Dyspnea Assessment

Breathlessness, medically known as dyspnea, is a multifaceted symptom and a significant predictor of survival rates among patients dealing with traumatic injury or respiratory ailments. Dyspnea typically signals an underlying serious condition and necessitates prompt intervention to mitigate the risk of respiratory failure or fatality. Presently, the assessment of dyspnea relies on patients self-reporting their perceived level of breathing difficulty through the Borg Dyspnea Score. This research has shed light on the correlation between prefrontal cortex activity and subjective experiences of dyspnea such as breathlessness or labored breathing. Leveraging this understanding, we have devised a method to quantify dyspnea by integrating prefrontal cortex processing, offering an objective measure of dyspnea. This approach aims to assist clinicians in the diagnosis and management of respiratory conditions and injuries more effectively. In our study, we collected objective dyspnea scores (ODS) along with subject-reported Borg scores from 25 patients with chronic obstructive pulmonary disease (COPD) and 3 healthy volunteers. The results of this study demonstrates that the ODS system can distinguish categories of clinically relevant dyspnea and breathing effort with a high degree of accuracy.

Negar Orangi-Fard
Georgia Gwinnett College
Georgia Gwinnett College
norangifard@ggc.edu

PP3

MS53: Unsupervised Simulation Based Inference for Amortized Bayesian Inverse Problems Without

Prior Samples

Advances in modern generative models have made Simulation-Based Inference (SBI) a vital tool across numerous scientific disciplines. However, a significant limitation of SBI is its dependency on prior samples, which may not always be available. For instance, in subsurface imaging, it is impractical to obtain prior samples directly from the Earth. To address this challenge, we have developed extensions to SBI that facilitate its application even when prior samples are unavailable. Our primary contribution is a novel reformulation of the non-amortized backward Kullback-Leibler divergence typically used in variational inference for learning posteriors. Specifically, we incorporate an expectation over observations, which enables two key capabilities: (i) effective amortization of the posterior approximation across a family of observations, and (ii) the ability to learn useful features from a training dataset, thereby enabling posterior learning without direct access to prior samples. We validate our approach through a stylized test, demonstrating that our method can accurately sample from an inverse problem with a known posterior. Furthermore, we showcase a practical implementation on a large-scale subsurface imaging problem that requires solving costly partial differential equations (PDEs).

Felix Herrmann, Rafael Orozco, Tuna Erdinc
Georgia Institute of Technology
felix.herrmann@gatech.edu, rorozco@gatech.edu, herdinc3@gatech.edu

PP3

On the Variational Interpretation of Mirror Play in Monotone Games

Mirror play (MP) is a well-accepted primal-dual multiagent learning algorithm where all agents simultaneously implement mirror descent in a distributed fashion. The advantage of MP over vanilla gradient play lies in its usage of mirror maps that better exploit the geometry of decision domains. Despite extensive literature dedicated to the asymptotic convergence of MP to equilibrium, the understanding of the finite-time behavior of MP before reaching equilibrium is still rudimentary. To facilitate the study of MP's non-equilibrium performance, this work establishes an equivalence between MP's finite-time primal-dual path (mirror path) in monotone games and the closed-loop Nash equilibrium path of a finite-horizon differential game, referred to as mirror differential game (MDG). Our construction of MDG rests on the Brezis-Ekeland variational principle, and the stage cost functional for MDG is Fenchel coupling between MP's iterates and associated gradient updates. The variational interpretation of mirror path in static games as the equilibrium path in MDG holds in deterministic and stochastic cases. Such a variational interpretation translates the non-equilibrium studies of learning dynamics into a more tractable equilibrium analysis of dynamic games, as demonstrated in a case study on the Cournot game, where MP dynamics corresponds to a linear quadratic game.

Yunian Pan
New York University
New York University
yp1170@nyu.edu

PP3

MS42: Approximation Rates for Minimum Integral Probability Metric Generative Models

We present a thorough error analysis for generative models trained to sample from arbitrary data distributions. We derive a priori error bounds based on the model complexity and the size of the training dataset, specifically for models optimized using statistical divergences. Our theoretical framework is applicable to losses of the integral probability metric type, such as Maximum Mean Discrepancy and Wasserstein distance, within both neural network and kernel method-based models. Additionally, we provide a comprehensive set of numerical examples that validate our theoretical error bounds, demonstrating their consistency with experimental results.

Biraj Pandey
University of Washington in Seattle
bpandey@uw.edu

Bamdad Hosseini
University of Washington
bamdadh@uw.edu

PP3

A Novel Implicit Neural Representation-Based Metal Artifact Reduction in Dental CBCT

Dental cone-beam computed tomography (CBCT) has gained popularity as a cost-effective and low-radiation alternative in dental clinics. However, its clinical utility is partly limited by artifacts related to metallic objects in the images. Recently, implicit neural representations (INRs), particularly through multi-layer perceptrons, have shown significant potential for dental CBCT reconstructions. However, these studies have not specifically addressed the challenges associated with metal artifact reduction (MAR). Based on a mathematical analysis of the physical interactions between polychromatic X-ray beams and metal objects, we present an INR-based MAR method that generates two distinct and informative tomographic images. The first image represents the monochromatic attenuation distribution at a specific energy level, while the second captures the nonlinear beam-hardening effect due to the polychromatic nature of X-ray beams. We also discuss the advantages of our proposed method compared to conventional model-based and deep learning-based MAR approaches.

Hyoung Suk Park, Kiwan Jeon
National Institute for Mathematical Sciences
hspark@nims.re.kr, Jeonkiwan@nims.re.kr

PP3

MS54: Asymptotically Free Sketching and Ridge Regression Applications

Classical results in sketching for dimensionality reduction in machine learning assert that, provided the sketch size is sufficiently large, the original unsketched solution is recovered at a fraction of the cost. However, in many practical settings, the sketch size may be smaller than needed for these guarantees. We provide a more general asymptotic result for sketched matrix inversion that holds for any sketch size and reveal that sketching is equivalent to adding ridge regularization. We prove our results for a broad class of asymptotically free sketches encompassing the spectral profiles of most sketches used in practice. We then determine the precise effect of sketching on the generalization error of ridge regression and show that the generalized cross-

validation risk estimator is consistent for sketched ensembles, enabling the efficient evaluation of unsketched ridge regression risk using only sketched data.

Daniel LeJeune
Stanford University
daniel@dlej.net

Pratik Patil
Berkeley
pratikpatil@berkeley.edu

PP3

MS53: The Fundamental Subspaces of Ensemble Kalman Inversion

The Ensemble Kalman Inversion (EKI) method re-interprets the ensemble Kalman filter as an adjoint/derivative-free iterative method for solving least squares problems, particularly in the context of inverse problems. The method has a history of use in applications in the geosciences, and was first investigated more mathematically in [Iglesias, Law, & Stuart 2013]. This initial work laid the foundation for a decade of subsequent developments of both variations of the method (e.g., by adapting different Kalman filtering methods to the inversion setting) as well as theoretical analyses proving rates of convergence of EKI in the linear setting (largely using ODE theory and stochastic calculus to analyze continuous-time limits of the iterative method). This poster aims to provide a broadly accessible introduction to EKI and its linear theory. In particular, we present a new analysis approach that analyzes the discrete iteration directly, relying only on basic knowledge of probability and linear algebra. Our analysis naturally gives rise to an elegant interpretation of the method's properties through the lens of "Fundamental Subspaces" of EKI, in the style of the fundamental subspaces of linear algebra. Numerical experiments illustrate our results.

Elizabeth Qian
Georgia Institute of Technology
eqian@gatech.edu

PP3

Low-Cost Robotic Arms and Large Language Models for Autonomous Scientific Laboratories

The autonomous scientific laboratories paradigm envisions autonomous robots performing trial-and-error scientific experiments. The motivation for this paradigm is to remove the uncertainties present due to human experimenters, where a significant portion of the experiments performed by them are not reproducible. The workflow execution interface (WEI) library from Argonne National Laboratory provides a way of integrating multiple devices and sensors for executing complex scientific workflows. However, existing experiments rely on expensive robots and sensors for performing precise experiments and are thus not good educational tools for upcoming engineers and researchers. To solve this issue, we propose to use the low-cost Arduino Braccio robot and a web camera to perform a trial-and-error object-moving experiment. The interface to the experiment is a fine-tuned Llama 3 large language model (LLM) that takes instructions in English and then converts them into Python code for robot movement. To simulate real-world experiments, where a number of them might fail, the sizes of the objects have been carefully chosen so that the robot has some difficulty in picking up and placing the object, and the experiments fail around 75% of the time. The goal of this work is to demonstrate autonomous scientific experiments using low-cost equipment and sensors to provide more accessible training opportunities to students and professionals.

Sunny Raj
Oakland University
raj@oakland.edu

Sumit K. Jha
Florida International University
sumit.jha@fiu.edu

Pranav Sinha
Oakland University
pranavsinha@oakland.edu

Arvind Ramanathan
Argonne National Laboratory
arvindster@gmail.com

PP3

Broader Engagement (BE): Analyzing Changes in Groundwater Level As a Result of Rainfall at An Infiltration Site in Mitchells Plain, Cape Town

Cape Town, South Africa is a water-scarce region, as demonstrated by the 'Day Zero' water crisis in 2018. In 2019, the City of Cape Town developed a new Water Strategy in which the City committed to diversify its water supply. Most of the city's water comes from surface water sources, but the city also has options for groundwater, including the Cape Flats Aquifer (CFA) which lies beneath the city. Rainfall, currently an underutilized water resource, can be leveraged through stormwater harvesting (SWH) and managed aquifer recharge (MAR) into the CFA. A stormwater detention pond in Mitchells Plain, Cape Town was recently retrofitted with an infiltration swale to enhance the pond's infiltration capacity, serving to recharge the CFA below. This research explores the relationship between rainfall observations and groundwater level (GWL) measurements at various locations across the detention pond. It investigates how long it takes for GWL to respond to rainfall, and at what quantity GWL responds to rainfall at points across the pond. Results show that locations close to the infiltration swale experience greater and more immediate GWL changes in response to rainfall events. This indicates that the process of retrofitting the detention pond enhanced its infiltration capacity. An opportunity is presented to retrofit other ponds around the City to develop blue-green infrastructures, and to utilize stormwater to recharge the CFA as a part of the Citys Water Strategy.

Sonja Rebarber
Swarthmore College, U.S.
srebarb1@swarthmore.edu

Rachelle Schneuwly, Nompumelelo Mnisi, Kirsty Carden
University of Cape Town, South Africa
rachelle.schneuwly@uct.ac.za, mnsnom013@myuct.ac.za, kirsty.carden@uct.ac.za

PP3

Generating Singular Distributions Through Neural

Networks

There is a large body of work on neural networks as function approximators, either in the context of regression or classification. In recent years another use for neural networks has emerged in generative AI, namely for generating high-dimensional probability distributions. It was shown in "High-dimensional distribution generation through deep neural networks" by Perekrestenko et al., 2021 that the set of probability distributions supported on the unit cube in $\mathbb{R}^d$ can be approximated by push-forwards of the uniform distribution on the unit interval through a set of neural networks of cardinality $2^{C\varepsilon^{-d}\log^2(\varepsilon)}$ to achieve an approximation error of ε. In this poster, we consider the class of probability distributions that are supported on (countable) m-rectifiable sets, i.e., on (countable unions of) Lipschitz images of compact sets in $\mathbb{R}^m$. We show that such distributions can be approximated by push-forwards of the uniform distribution on the unit interval through a set of neural networks of cardinality $2^{C\varepsilon^{-m}\log^2(\varepsilon)}$ to achieve an approximation error of ε. Notably, the ambient space dimension d can be replaced by the Hausdorff dimension m of the rectifiable set. This result is striking as in practice the dimension of the data set can be much smaller than the dimension of the ambient space.

Erwin Riegler
ETH Zurich
eriegler@mins.ee.ethz.ch

Alex Bühler
ETH Zürich
albuehler@student.ethz.ch

Yang Pan
Chair for Mathematical Information Science
ETH Zürich
yangpan@mins.ee.ethz.ch

Helmut Bölcskei
ETH Zürich
boelcskei@nari.ee.ethz.ch

PP3

MS52: Implicit Symbolic Regression for Data-Driven Discovery of Conservation Laws

Discovery of conservation laws is an important challenge for data-driven learning of dynamical systems. Conservation laws correspond to implicit equations satisfied by dynamic trajectories. Symbolic regression is a key method for extracting concise mathematical expressions to model relationships in data. However, existing symbolic regression methods are primarily designed for explicit functions, and encounter challenges when attempting to learn implicit equations. Methods for recovering non-trivial implicit equations typically depend on numerical gradients, which require larger amounts of densely sampled training data. We develop a new and simple-to-use method for symbolic regression that enables learning of implicit equations, even with noisy training datasets, which to the best of our knowledge has not been widely feasible previously. Our method is based on the formulation of a probabilistic fitness function using the Kullback-Leibler divergence, and is compatible with existing symbolic regression algorithms. We demonstrate our method on a series of benchmark implicit equations, and use it to recover conserved quantities in dynamical systems.

Graham Roberts, Everett Grethel, Qian Yang
University of Connecticut
graham.roberts@uconn.edu, everett.grethel@uconn.edu, qyang@uconn.edu

PP3

MS50: All You Need Is Resistance: On the Equivalence of Effective Resistance and Certain Optimal Transport Problems on Graphs

The fields of effective resistance and optimal transport on graphs are filled with rich connections to combinatorics, geometry, machine learning, and beyond. In this article we put forth a bold claim: that the two fields should be understood as one and the same, up to a choice of p. We make this claim precise by introducing the parameterized family of p-Beckmann distances for probability measures on graphs and relate them sharply to certain Wasserstein distances. Then, we break open a suite of results including explicit connections to optimal stopping times and random walks on graphs, graph Sobolev spaces, and a Benamou-Brenier type formula for 2-Beckmann distance. We further explore empirical implications in the world of unsupervised learning for graph data and propose further study of the usage of these metrics where Wasserstein distance may produce computational bottlenecks.

Sawyer Robertson
UC San Diego
s5robert@ucsd.edu

PP3

Operator Approximations for Inverse Problems

This poster presents a framework for resolving inverse problems through the use of operator approximations over vector valued RKHSs. This generalizes Koopman based methods for data driven methods in dynamical systems by laying out the general requirements to resolve an inverse problem through an operator approximation. We will present three different examples of this method in action.

Joel Rosenfeld
University of South Florida
Dept. of Math and Stats
rosenfeldj@usf.edu

Benjamin Russo
Riverside Research
brusso@riversideresearch.org

Rushikesh Kamalapurkar
Oklahoma State University
rushikesh.kamalapurkar@okstate.edu

PP3

Broader Engagement (BE): Online Boosted Gaussian Learners for In-situ Detection and Characterization of Protein Folding States in Molecular Dynamics Simulations

With modern technology and High Performance Computing (HPC), Molecular Dynamics (MD) simulations can be task and data parallel. That means, they can be decomposed into multiple independent tasks (i.e., trajectories) with their own data, which can be processed in paral-

lel. Analysis of MD simulations includes finding specific molecular events and the conformation changes that a protein undergoes. In this work, we propose a data parallel, lightweight technique to learn the characteristics of protein folding states in MD simulations. Contrary to other methods, ours can differentiate relevant states in a single protein folding trajectory without requiring centralized global knowledge of the protein dynamics. As its processing and memory overheads are negligible (in the order of milliseconds per window of frames, and kilo bytes respectively) this technique can be coupled with the simulation for in-situ analysis. We also developed a framework which a direct evolution of the Online Boosted Gaussian Learner (OBGL) algorithm, expanding its capabilities to operate within a real-time context. In this framework, the producer component continuously generates window of frames from the MD simulation and places them into a shared buffer. The consumer, retrieves these windows from the buffer for concurrent analysis. The core objective of this framework is to minimize latency and maximize throughput in the processing pipeline, thereby ensuring timely and efficient data analysis.

Harshita Sahni
University of New Mexico, U.S.
hsahni@unm.edu

PP3

MS47: A Network Science Approach to Understanding EV Charging Accessibility in the Texas Triangle

A robust electric vehicle (EV) charging infrastructure is essential for the widespread adoption of electric vehicles and the transition towards sustainable energy. Understanding the spatial distribution of existing infrastructures is crucial for the strategic expansion and futureproofing of the burgeoning EV market. This research presents a detailed analysis of EV charging stations within the Texas Triangle, which includes the state's four major metropolitan areasAustin, Dallas, Houston, and San Antonio. The comprehensive dataset utilized for constructing and analyzing the network graph was sourced from the Alternative Fuels Data Center, which provides extensive information on EV charging stations. We employed a graph-based method to model the distribution of charging stations, effectively capturing the complex spatial dynamics intuitively and computationally efficiently. Additionally, this study examines how specific attributes of EV stations, such as operational hours, pricing, and accessibility, influence the network's capacity to meet regional charging demands. Graph neural networks (GNNs) were utilized to accurately model the spatial distribution of charging stations, optimized using the Adam optimizer. We also explored various complementarity metrics to assess the spatial distribution of the charging station network. Finally, we conducted a temporal analysis to examine how the network evolved.

Yogesh Sapkota
Youngstown State University
yspkta16@gmail.com

PP3

MS53: An Innovative and Rapid Physics-Informed Data-Driven Method for Atmospheric Radiative Transfer Inversion

FORUM (Far-infrared Outgoing Radiation Understanding and Monitoring) is the ninth Earth Explorer satellite mission chosen by the European Space Agency in 2019. It will provide interferometric measurements in the Far-InfraRed (FIR) spectrum, constituting 50% of Earth's outgoing longwave flux emitted by our planet into space. Accurate Top Of the Atmosphere measurements in the FIR are crucial for improving climate models. Existing instruments are insufficient, necessitating new computational techniques. In early mission stages, an End-to-End Simulator was developed to demonstrate proof-of-concept and assess the impact of instrument characteristics on reconstructed atmospheric properties. From a mathematical perspective, two challenges arise: the radiative transfer equation (direct problem) and its inversion (retrieval problem). Both problems can be addressed through a full-physics method, particularly applying the Optimal Estimation approach, a specialized Tikhonov regularization scheme based on Bayesian formulation. However, full-physics methods' computational demands hinder Near Real-Time data analysis. Faster models are essential for next-gen satellites measuring hundreds of spectra per minute and climatology models simulating years of global-scale radiative transfer. To expedite solutions for both problems, a hybrid approach is employed, combining data-driven operator learning using the Moore-Penrose pseudoinverse with neural-network-based Tikhonov regularization.

Cristina Sgattoni
Institute for Applied Mathematics Mauro Picone (IAC) CNR
cristina.sgattoni@cnr.it

PP3

Exploring Covid-Related Relationship Extraction: Contrasting Data Sources and Analyzing Misinformation

The COVID-19 pandemic presented an unparalleled challenge to global healthcare systems. A central issue revolves around the urgent need to swiftly amass critical biological and medical knowledge concerning the disease, its treatment, and containment. Remarkably, text data remains an underutilized resource in this context. In this paper, we delve into the extraction of COVID-related relations using transformer-based language models, including Bidirectional Encoder Representations from Transformers (BERT) and DistilBERT. Our analysis scrutinizes the performance of five language models, comparing information from both PubMed and Reddit, and assessing their ability to make novel predictions, including the detection of "misinformation." Key findings reveal that, despite inherent differences, both PubMed and Reddit data contain remarkably similar information, suggesting that Reddit can serve as a valuable resource for rapidly acquiring information during times of crisis. Furthermore, our results demonstrate that language models can unveil previously unseen entities and relations, a crucial aspect in identifying instances of misinformation.

Tanvi Sharma
Tampere university, Finland
tanvi.sharma@tuni.fi

Frank Emmert-Streib, Nadeesha Perera, Amer Farea
Tampere University
frank.emmert-streib@tuni.fi, nadeesha.perera@tuni.fi,

amer.farea@tuni.fi

PP3

MS36: Radial Basis Function Methods for Neural Field Models

Neural field models are non-linear systems of integro-differential equations intended to model large-scale neural activity. There is growing interest in identifying efficient and accurate schemes for simulating neural field models as they can capture activity dynamics that spread across wide swathes of tissues and that reflect highly complex neural architecture. Recently, a framework has been put forth for analyzing neural field solvers (Avitable 2023) that separates the error due to the numerical representation of the solution (projection) and the error due to approximating the integral operator (quadrature). In this poster, we demonstrate using Radial Basis Function (RBF) interpolation and quadrature methods to combine and simplify this error analysis and to create efficient, robust, and high-order-accurate neural field solvers.

Sage B. Shaw
University of Colorado Boulder
sage.shaw@colorado.edu

Zachary P. Kilpatrick
University of Colorado, Boulder
zpkilpat@colorado.edu

Daniele Avitabile
Vrije Universiteit Amsterdam
d.avitabile@vu.nl

PP3

MS36: An Automated Method for Spike Ripples Localization in Human Intracranial Recordings

Spike ripples, the combination of epileptiform spikes and ripples, have been proposed as a reliable biomarker for the epileptogenic zone. However, detection of spike ripples requires a fast and reliable automated method to identify these infrequent, transient events. To that end, we combined two previously developed spike ripple detection strategies for use in noninvasive scalp electroencephalogram (EEG) recordings: a feature-based algorithm applied to time series data, and a convolutional neural network (CNN) applied to spectrogram images. To apply this method to invasive intracranial data, we first retrained the CNN detector on hand-marked events. We then created an intracranial detector that combined both the feature-based and CNN strategies. We validated the resulting detector on the hand-marked events using a leave-one-out cross-validation procedure. We find that, while the intracranial spike ripple detector identified less than half of hand-marked events (sensitivity, 42%), it had high precision (78.5%) and a low false positive rate (6.7%), resulting in balanced performance (F1 = 0.68). Applying the intracranial spike ripple detector to 109 subjects collected as part of a multicenter, international study, we find that most spike ripples were removed in subjects with good surgical outcome for epilepsy ($p < 0.001$), and that automatically detected spike ripples have improved specificity for epileptogenic tissue compared to alternative electrographic biomarkers.

Catherine Chu, Wen Shi
Harvard Medical School
Massachusetts General Hospital
cjchu@mgh.harvard.edu, wshi3@mgh.harvard.edu

Dana Shaw
Boston University
danashaw@bu.edu

Mark Kramer
Department of Mathematics and Statistics
Boston University
mak@bu.edu

PP3

MS50: Fast Alignment of Images in Sliced 2-Wasserstein Distance

We present a fast algorithm for aligning images using optimal transport. Our method is based on the sliced Wasserstein distance and computes the 1-D Wasserstein distance between radial line projections of the input images. Using this approach, we develop an algorithm that can align two $L \times L$ images in $\mathcal{O}(L^2 \log L)$ operations. We show that our method is robust to rotations, translations and deformations in the images.

Yunpeng Shi
UC Davis
ypshi@ucdavis.edu

PP3

Concentration Inequalities for Sums of Markov Dependent Random Matrices

We give Hoeffding and Bernstein-type concentration inequalities for the largest eigenvalue of sums of random matrices arising from a Markov chain. We consider time-dependent matrix-valued functions on a general state space, generalizing previous that had only considered Hoeffding-type inequalities, and only for time-independent functions on a finite state space. In particular, we study a kind of noncommutative moment generating function, give tight bounds on it, and use a method of Garg et al. to turn this into tail bounds. Our proof proceeds spectrally, bounding the norm of a certain perturbed operator. In the process we make an interesting connection to dynamical systems and Banach space theory to prove a crucial result on the limiting behavior of our moment generating function that may be of independent interest.

Bobby Shi
University of Texas at Austin
bhshi@utexas.edu

PP3

MS57: Tlasdi: Thermodynamics-Informed Latent Space Dynamics Identification

We present a thermodynamics-informed latent space dynamics identification (tLaSDI) method for learning an intrinsic invariant manifold that complies with the first and second principles of thermodynamics from data. The approach utilizes an autoencoder to provide a nonlinear latent representation and the reconstruction of the high-dimensional data and the corresponding latent space dynamics are captured using GENERIC formalism-informed neural networks (GFINNs). The GFINNs are designed to exactly satisfy the degeneracy condition of the GENERIC formalism, ensuring that the latent space dynamics adhere to fundamental thermodynamics principles. We develop

a novel loss formulation based on a rigorous error estimate, which significantly improves the generalization performance. Numerical examples are presented to demonstrate the performance of tLaSDI, which exhibits robust generalization ability, even in extrapolation. In addition, an intriguing correlation is empirically observed between the entropy production rates in the latent space and the behaviors of the full-state solution.

Yeonjong Shin
North Carolina State University
yeonjong_shin@ncsu.edu

Jun Sur Park
Korea Institute for Advanced Study
junsurpark@kias.re.kr

Siu Wun Cheung, Youngsoo Choi
Lawrence Livermore National Laboratory
cheung26@llnl.gov, choi15@llnl.gov

PP3

MS49: Noise-Tolerant Randomized Kaczmarz Variant

Solving large-scale linear systems is a problem that commonly arises in medical imaging, signal processing, and a variety of optimization subroutines. However, the measurement data often becomes corrupted with noise. Therefore, developing solvers that are robust to such inconsistencies in the data is of great interest. The case of symmetric noise is frequently considered, and multiple methods have been shown to be robust in that scenario. In contrast, our work focuses on systems with dense irregular noise. We propose a Randomized Kaczmarz variant that takes advantage of the switching step schedule to ensure convergence in the presence of asymmetric noise. For the minisymposium "Efficient and robust optimization techniques for structured data learning" (80608)

Sofiia Shvaiko
Princeton University
ss7950@princeton.edu

PP3

MS53: Scalable Data Assimilation in Latent Space Using Variational Autoencoder

Data assimilation in high dimensions with sparse observations can prove computationally challenging. By leveraging variational autoencoders and LSTMs to train latent dynamics, we can combine it with standard data assimilation approaches in the latent space to achieve fast, scalable results.

Phillip Si
Georgia Institute of Technology
psi6@gatech.edu

PP3

MS52: Characterizing Reducibility by Solving High-Dimensional Partial Differential Equations with Neural Networks

Dynamical systems are widely used in science and engineering. However, these systems often involve thousands of interacting components and exhibit complex dynamics. This either makes their simulation over large time intervals challenging, or their use as a part of a complex computational model, such as in control and optimization, computationally demanding. In these settings, model reduction is desirable. However, it is not always clear whether this is possible: the system may not be reducible or may be reducible only to a degree, the selected reduced models may not extrapolate beyond the time interval used for training, or they may be too sensitive to initial conditions. In this work, we introduce a partial differential equation whose solution, which we call the latent flow, encodes all possible reduced models along with their dependence on the initial conditions to the system. As this PDE is high-dimensional, we use deep neural networks to approximate its solution. We validate our results for low-dimensional systems using standard finite-element methods, and we show numerical results for high-dimensional chemical reaction networks and some evolution equations.

Carlos Sing Long
Institute for Mathematical and Computational Engineering
Pontificia Universidad Catolica de Chile
casinglo@uc.cl

Qian Yang
University of Connecticut
qyang@uconn.edu

Manuel Sanchez
Institute for Mathematical and Computational Engineering
Pontificia Universidad Catolica de Chile
manuel.sanchez@uc.cl

Vicente Hojas
Pontificia Universidad Católica de Chile
vahojas@uc.cl

PP3

MS47: Understanding the Influence of Genre-Specific Music Using Network Analysis and Machine Learning Algorithms

This study analyzes a network of musical influence using machine learning and network analysis techniques. A directed network model is used to represent the influence relations between artists as nodes and edges. Network properties and centrality measures are analyzed to identify influential patterns. In addition, influence within and outside the genre is quantified using in-genre and out-genre weights. Regression analysis is performed to determine the impact of musical attributes on influence. We find that speechiness, acousticness, and valence are the top features of the most influential artists. We also introduce the IRDI, an algorithm that provides an innovative approach to quantify an artists influence by capturing the degree of dominance among their followers. This approach underscores influential artists who drive the evolution of music, setting trends and significantly inspiring a new generation of artists. The independent cascade model is further employed to open up the temporal dynamics of influence propagation across the entire musical network, highlighting how initial seeds of influence can contagiously spread through the network. This multidisciplinary approach provides a nuanced understanding of musical influence that refines existing methods and sheds light on influential trends and dynamics.

Subham Singh

University of Oklahoma
Subham.Singh-1@ou.edu

PP3

Broader Engagement (BE): In Silico Analysis of the Flagella-Propelled Cell Swimming in Viscous Fluids

Flagella are complex cellular structures with crucial roles in cell propulsion, biomixing, and transport. This project aims to create a computational fluid dynamics model of a green algae cell with two flagella moving through a viscous fluid. We utilize the AMReX-based fluctuating hydrodynamics solver, which features custom algorithms to simulatet the flagella as an immersed chain of force particles with specified motions. A key aspect of our approach involves leveraging Berkeley's supercomputer, Perlmutter, by employing both CPU and GPU-compatible code to enhance simulation performance. We have also improved the previously developed yt-centric JupyterLab notebook with pyAMReX, a native Python API for AMReX codes, to streamline data importation, visualization, and analysis. The simulation, executed on Perlmutter, has allowed us to visualize fluid flow fields and forces on the discretized flagellar particles. By the end of the summer internship, we expect to generate preliminary results that will illustrate how the movement of a beating flagellum attached to the cell body affects both the cells motion and the surrounding fluid, advancing our understanding of flagellar functions.

Lauren Sowards, Gang Xu
University of Central Oklahoma, U.S.
lsowards@uco.edu, gxu@uco.edu

Daniel Ladinges
Lawrence Berkeley National Laboratory, U.S.
drladiges@lbl.gov

Quinton Fister
University of Central Oklahoma, U.S.
qfister@uco.edu

PP3

Contextual Feature Selection with Conditional Stochastic Gates

Feature selection is a crucial tool in machine learning and is widely applied across various scientific disciplines. Traditional supervised methods generally identify a universal set of informative features for the entire population of samples. However, feature relevance often varies with context, where the context itself may not directly affect the outcome variable. Here, we propose a novel end-to-end architecture for prediction with contextual feature selection where the subset of selected features is conditioned on the value of the samples context variables. Our new approach, Conditional Stochastic Gates (c-STG), models the importance of features using conditional Bernoulli variables whose parameters are estimated based on contextual variables using a hypernetwork. Our architecture is composed of the hypernetwork to learn the context-dependent gates followed by a prediction model, which can be linear or nonlinear. We further present a theoretical analysis of our model, indicating that it can improve performance and flexibility over population-level methods in complex feature selection settings. Finally, we conduct an extensive benchmark using simulated and real-world datasets across multiple domains, including health and neuroscience, demonstrating that c-STG leads to improved feature selection capabilities while enhancing prediction accuracy and interpretability.

Ram Dyuthi Sristi
PhD Student
University of California San Diego
rsristi@ucsd.edu

Ofir Lindenbaum
Yale University
ofirlin@gmail.com

Shira Lifshitz, Maria Lavzin, Jackie Schiller
Technion - Israel Institute of Technology
shiralif@campus.technion.ac.il, maria.lavzin@gmail.com, jackie@technion.ac.il

Gal Mishne
UC San Diego
gmishne@ucsd.edu

Hadas Benisty
Technion - Israel Institute of Technology
hadasbe@technion.ac.il

PP3

MS48: Distributional Smoothing for Machine Learning

Distribution smoothing is a method for modifying a target function in an optimization problem using a set of surrogate functions that are easier to optimize. When applied to (stochastic) gradient descent, we substitute the gradient of the target function with the gradient from this more favorable class of functions. This poster will have an exploration of the intuition and motivation behind distributional smoothing. It will also delve into the convergence properties of Smooth Gradient Descent, labeled SmoothGD, and its stochastic counterpart, SmoothSGD, demonstrating that their convergence rates are no worse than those of their non-smoothed counterparts. Finally, it will provide numerical evidence from both machine learning and data science problems.

Andrew Starnes
Lirio AI Research
astarnes@lirio.com

Anton Dereventsov
Lirio, LLC
adereventsov@lirio.com

Clayton Webster
Lirio
cwebster@lirio.com

PP3

Breaking the Quadratic Rank Barrier in Non-Convex Matrix Sensing

Low-rank matrix recovery problems are ubiquitous in many areas of science and engineering. Most of the methods that have been studied for these problems can be divided into two categories: Convex optimization approaches based on nuclear norm minimization, and non-convex approaches that use factorized gradient descent. While the latter promises to be computationally much less expensive, basically all existing recovery guarantees for factorized gradient descent are much more pessimistic with respect to the

number of samples required. In particular, they require the number of samples to scale quadratically with the rank of the ground truth matrix. This is in stark contrast to empirical observations which suggest that the non-convex approaches perform as well as the convex ones with respect to the sample complexity. In this talk, we resolve this issue and we present the first theoretical guarantees to the best of our knowledges for matrix sensing that show that factorized gradient descent recovers the ground truth matrix with a sample size that is optimal in the number of degrees of freedom. Our proof is based on new probabilistic decoupling arguments, which we expect to be of independent interest.

Dominik Stöger
Catholic University of Eichstaett-Ingolstadt, Germany
dominik.stoeger@ku.de

Yizhe Zhu, Yizhe Zhu
UC Irvine
yizhe.zhu@uci.edu, yizhe.zhu@uci.edu

PP3

Broader Engagement (BE): Sonoma State University Covid-19 Model

A model for Covid-19 infection within the Sonoma State University (SSU) community is proposed through means of a simplified differential equations-based SEAIRV model which considers a closed population of students, faculty, at the university. Data was collected by analyzing Covid-19 cases at SSU, as well as data obtained by local agencies regarding vaccination, recovery and death rates amongst individuals in the community. A basic reproductive number (R0) of 3.8663 is calculated for the system using the Next Generation Matrix method, and simulations in R were conducted to consider effectiveness of distinct levels in vaccination; suggesting the rapid spread of the virus without intervention. Our results indicated that disparities between communities, namely age, race, gender, do affect the outcomes from COVID-19 infection exposure. Continued use of masks, social distancing, and increasing the proportion of vaccinated individuals on campus to 100% would help reduce the total number of cases in the population.

Dirk Tolson III
University of California, Berkeley, U.S.
dirktolson925@gmail.com

PP3

MS58: Score Estimation for Supervised Generative Modeling

We present a supervised learning framework of generative models based on score-based diffusion models. This technique uses a reverse-time diffusion process in the form of a stochastic differential equation (SDE) to transport the standard Gaussian distribution to a complex target data distribution, thus, the goal of generating new data boils down to solving reverse-time SDEs. A key challenge of score-based diffusion models is to estimate score functions, which store the information of the target distribution and involve high-dimensional integrals. In this work, we investigate multiple different techniques for efficient score estimation, including Monte Carlo, important sampling and Langevin dynamics. The efficiency of the approaches will be demonstrated via benchmark data sampling and uncertainty quantification problems.

Hoang A. Tran
Oak Ridge National Laboratory
Computer Science and Mathematics Division
tranha@ornl.gov

PP3

DyCA and DMD: Differences and Similarities in Theory and Application

Dynamical Component Analysis (DyCA) is a multivariate/spatio-temporal signal decomposition technique that aims to obtain a data-driven model of the signal, which is governed by a special set of differential equations. The optimization leads to a generalized eigenvalue problem of correlation matrices of the signal and time-shifted signals. Therefore, the algorithm is similar to Dynamic Mode Decomposition (DMD) with respect to its goal and underlying data-driven approach. However, the approach differs from that of DMD. We review the algorithm and its application to both simulated and real-world data. New results are presented to demonstrate similarities and differences between DyCA and DMD. These are discussed with respect to mode selection, amplitude, and signal representation under varying measurement setups, including temporal resolution and noise conditions. DyCA outperforms DMD under certain conditions and may represent a useful alternative algorithm for dimension reduction and data-driven dynamical systems research.

Christian Uhl, Annika Stiehl, Nicolas Weeger
University of Applied Sciences Ansbach
christian.uhl@hs-ansbach.de, annika.stiehl@hs-ansbach.de,
nicolas.weeger@hs-ansbach.de

PP3

MS50: Spectral Embedding with Group Invariance: Theory and Applications

Spectral embedding is a popular kernel-based nonlinear dimensionality reduction method. We consider a data model where each point is drawn from a low-dimensional manifold, M, and a random element from a known group G is applied to it. This introduces nuisance degrees of freedom, affecting the methods ability to capture the required representation. To address this, we employ a minimum G-invariant kernel that computes affinities between the orbits of the G action, effectively reducing the complexity of the data representation. We prove that embedding the data points using the minimum kernel is equivalent to working directly in the quotient space M/G of dimension dim(M)-dim(G), thus capturing the intrinsic geometry of the data and leading to improved convergence rates. we demonstrate our approach on molecular dynamics and cryo-electron microscopy, where managing the effects of transformations is critical.

Yeari Vigder
Tel-Aviv University
yearivig@gmail.com

Paulina Hoyos
University of Texas Austin
paulinah@utexas.edu

Amit Moscovich
Tel Aviv University
mosco@tauex.tau.ac.il

Joe Kileel
University of Texas at Austin, U.S.
jkileel@math.utexas.edu

PP3

MS57: CoNFiLD: Conditional Neural Field Latent Diffusion Model Generating Spatiotemporal Turbulence

This study introduces the Conditional Neural Field Latent Diffusion (CoNFiLD) model, a novel generative learning framework designed for rapid simulation of intricate spatiotemporal dynamics in chaotic and turbulent systems within three-dimensional irregular domains. Traditional eddy-resolved numerical simulations, despite offering detailed flow predictions, encounter significant limitations due to their extensive computational demands. In contrast, deep learning-based surrogate models promise efficient, data-driven solutions. However, their effectiveness is often compromised by a reliance on deterministic frameworks, which fall short in accurately capturing the chaotic and stochastic nature of turbulence. The CoNFiLD model addresses these challenges by synergistically integrating conditional neural field encoding with latent diffusion processes, enabling the memory-efficient and robust probabilistic generation of spatiotemporal turbulence under varied conditions. Leveraging Bayesian conditional sampling, the model can seamlessly adapt to a diverse range of turbulence generation scenarios without the necessity for retraining, covering applications from zero-shot full-field flow reconstruction using sparse sensor measurements to super-resolution generation and spatiotemporal flow data restoration.

Jian-Xun Wang
Dept. of Aerospace and Mechanical Engineering
University of Notre Dame
jwang33@nd.edu

Pan Du, Meet Parikh, Xiantao Fan, Xin-Yang Liu
University of Notre Dame
pdu@nd.edu, mparikh2@nd.edu, fanxiantao1@gmail.com, xliu28@nd.edu

PP3

MS58: A Pseudo-Reversible Normalizing Flow for Stochastic Differential Equation

We present a pseudo-reversible normalizing flow method for efficiently generating samples of the state of a stochastic differential equation (SDE) with various initial distributions. The primary objective is to construct an accurate and efficient sampler that can be used as a surrogate model for computationally expensive numerical integration of SDE, such as those employed in particle simulation. After training, the normalizing flow model can directly generate samples of the SDE's final state without simulating trajectories.The main novelty of our normalizing flow model is that it can learn the conditional distribution of the state, i.e., the distribution of the final state conditional on any initial state, such that the model only needs to be trained once and the trained model can be used to handle various initial distributions. Additionally, we propose to use a pseudo-reversible network architecture to define the normalizing flow model, which has sufficient expressive power and training efficiency for a variety of SDEs in science and engineering, e.g., in particle physics. We provide a rigorous convergence analysis of the pseudo-reversible normalizing flow model to the target probability density function in the KullbackLeibler divergence metric. Numerical experiments are provided to demonstrate the effectiveness of the proposed normalizing flow model.

Pengjun Wang
Auburn University
pzw0030@auburn.edu

Minglei Yang, Diego Del-Castillo-Negrete
Oak Ridge National Laboratory
yangm@ornl.gov, delcastillod@ornl.gov

Yanzhao Cao
Department of Mathematics & Statistics
Auburn University
yzc0009@auburn.edu

Guannan Zhang
Oak Ridge National Laboratory
zhangg@ornl.gov

PP3

MS42: Optimal Transport with Cost-Free Transformations

An extension of the optimal transport problem is proposed, which includes a family of transformations incurring no transportation cost. Such extension improves the *co-registration* among datasets, where certain transformations, such as rotations, displacements and changes of perspective, are a natural component of data acquisition. The extended optimal transport problem pairs two distributions with minimal distortion, while identifying the cost-free explainable components of the map. A data-driven formulation is developed, as well as a methodology for its numerical solution. This complements gradient descent with a game-theory inspired approach, favoring collaborative moves between the cost-free and the regular transformations. The methodology is illustrated through its application to the pairing of both synthetic and real images.

Yating Wang
Courant Institute of Mathematical Sciences
New York University
yw3087@nyu.edu

Esteban G. Tabak
Courant Institute
New York University
tabak@cims.nyu.edu

Debra Laefer
Department of Civil and Urban Engineering
New York University
debra.laefer@nyu.edu

PP3

MS46: Statistic Maximal Leakage

We introduce a privacy metric called statistic maximal leakage that quantifies how much a privacy mechanism leaks about a specific secret, relative to the adversarys prior information about that secret. Statistic maximal leakage is an extension of the well-known maximal leakage. Unlike maximal leakage, it protects a single, known secret. We show that statistic maximal leakage satisfies composition and post-processing properties. Additionally, we show how

to efficiently compute it in the special case of deterministic data release mechanisms. We analyze two important mechanisms under statistic maximal leakage: the quantization mechanism and randomized response. We show theoretically and empirically that the quantization mechanism achieves better privacy-utility tradeoffs in the settings we study.

Shuaiqi Wang
CMU
shuaiqiw@andrew.cmu.edu

PP3

MS56: Rate of Convergence of Policy Improvement Algorithm for Exploratory Stochastic Control Problems

In this presentation, we study the Policy Improvement Algorithm for reinforcement learning for continuous-time entropy-regularized stochastic control problems. We prove the uniform convergence both for the iterative value functions and for their derivatives. More importantly, in the finite horizon case and in the infinite horizon case with a large discount factor, we obtain the exponential rate of convergence, which is new in the literature. Our arguments rely on a simple representation formula for the derivatives of a linear PDE and are much easier than those in the existing works for convergence.

Gaozhan Wang, Jin Ma, Jianfeng Zhang
University of Southern California
gaozhanw@usc.edu, jinma@usc.edu, jianfenz@usc.edu

PP3

Nonlinear Spiked Covariance Matrices and Signal Propagation in Deep Neural Networks

Many recent works have studied the eigenvalue spectrum of the Conjugate Kernel (CK) defined by the nonlinear feature map of a feedforward neural network. However, existing results only establish weak convergence of the empirical eigenvalue distribution, and fall short of providing precise quantitative characterizations of the "spike' eigenvalues and eigenvectors that often capture the low-dimensional signal structure of the learning problem. In this work, we characterize these signal eigenvalues and eigenvectors for a nonlinear version of the spiked covariance model, including the CK as a special case. Using this general result, we give a quantitative description of how spiked eigenstructure in the input data propagates through the hidden layers of a neural network with random weights. As a second application, we study a simple regime of representation learning where the weight matrix develops a rank-one signal component over training and characterize the alignment of the target function with the spike eigenvector of the CK on test data.

Zhichao Wang
UCSD
zhw036@ucsd.edu

PP3

MS42: Efficient Neural Network Approaches for Conditional Optimal Transport with Applications in Bayesian Inference

We present two neural network approaches that approximate the solutions of static and dynamic *conditional optimal transport* (COT) problems. Both approaches enable conditional sampling and conditional density estimation, which are core tasks in Bayesian inference—particularly in the simulation-based ("likelihood-free') setting. Our methods represent the target conditional distributions as transformations of a tractable reference distribution and, therefore, fall into the framework of measure transport. Conditional optimal transport (COT) comprises a unique, canonical choice within this framework, but finding COT maps can be computationally challenging, even in moderate dimensions. To improve scalability, our numerical algorithms use neural networks to parameterize COT maps and further exploit the structure of the COT problem. Our static approach approximates the map as the gradient of a partially input-convex neural network and uses a novel numerical implementation to increase computational efficiency compared to state-of-the-art alternatives. Our dynamic approach approximates the conditional optimal transport via the flow map of a regularized neural ODE; compared to the static approach, it is slower to train but faster to generate samples. We demonstrate both algorithms numerically, comparing them with competing state-of-the-art approaches, using benchmark datasets and simulation-based Bayesian inverse problems.

Zheyu Oliver Wang
Massachusetts Institute of Technology
olivrw@mit.edu

Ricardo Baptista
Caltech
rsb@caltech.edu

Youssef M. Marzouk
Massachusetts Institute of Technology
ymarz@mit.edu

Lars Ruthotto
Departments of Mathematics and Computer Science
Emory University
lruthotto@emory.edu

Deepanshu Verma
Department of Mathematics
Emory University
dverma@clemson.edu

PP3

MS48: The Anytime Convergence of Stochastic Gradient Descent with Momentum: From a Continuous-Time Perspective

We study the stochastic optimization problem from a continuous-time perspective. We propose a stochastic first-order algorithm, called Stochastic Gradient Descent with Momentum (SGDM), and show that the trajectory of SGDM, despite its stochastic nature, converges to a deterministic second-order Ordinary Differential Equation (ODE) in L_2-norm, as the stepsize goes to zero. The connection between the ODE and the algorithm results in delightful patterns in the discrete-time convergence analysis. We prove that, for any β, the sequence x_k governed by running SGDM on a smooth convex function f satisfies $\mathbb{P}\left(f(x_k) - f^* \leq \frac{C\sqrt{\log k}\log(2/\beta)}{\sqrt{k}}, ; \forall t \in \mathbb{N}\right) \geq 1 - \beta$, where C is some constant.

Tianyu Wang, Yasong Feng
Fudan University
wangtianyu@fudan.edu.cn, ysfeng20@fudan.edu.cn

Yifan Jiang
University of Oxford
yifan.jiang@maths.ox.ac.uk

Zhiliang Ying
Columbia University
zying@stat.columbia.edu

PP3

MS54: Robust Randomized Preconditioning for Kernel Ridge Regression

Our work introduces two randomized preconditioning techniques for robustly solving kernel ridge regression (KRR) problems with a medium to large number of data points ($10^4 <= N <= 10^7$). The first method, RPCholesky preconditioning, is capable of accurately solving the full-data KRR problem in $O(N^2)$ arithmetic operations, assuming sufficiently rapid polynomial decay of the kernel matrix eigenvalues. The second method, KRILL preconditioning, offers an accurate solution to a restricted version of the KRR problem involving $k << N$ selected data centers at a cost of $O((N+k^2)k\log k)$ operations. The proposed methods efficiently solve a broad range of KRR problems and overcome the failure modes of previous KRR preconditioners, making them ideal for practical applications.

Robert Webber
California Institute of Technology
rwebber@ucsd.edu

PP3

MS48: Gaussian Smoothing (Stochastic) Gradient Descent

This talk presents Gaussian smoothing as a powerful tool to enhance gradient descent algorithms, particularly in navigating complex and non-convex optimization landscapes. Leveraging Gaussian smoothing, we define a nonlocal gradient that effectively reduces high-frequency noise and rapid fluctuations while preserving essential structural features. The resulting Gaussian smoothing gradient descent (GSmoothGD) approach exhibits remarkable efficacy in traversing away from local minima, thereby substantially improving overall optimization performance. This can be applied to data science and machine learning by Gaussian smoothing stochastic gradient descent (GSmoothSGD) as well. We provide theoretical error estimates on the convergence rates of GSmoothGD and GSmoothSGD iterates, considering factors such as function convexity, smoothness, input dimension, and the Gaussian smoothing radius.

Clayton G. Webster
University of Texas and Lirio LLC
claytongwebster@utexas.edu

Andrew Starnes
Lirio AI Research
astarnes@lirio.com

Anton Dereventsov
Lirio, LLC
adereventsov@lirio.com

PP3

MS46: On the Inherent Privacy Properties of Discrete Denoising Diffusion Models

Privacy concerns have led to a surge in the creation of synthetic datasets, with diffusion models emerging as a promising avenue. Although prior studies have performed empirical evaluations on these models, there has been a gap in providing a mathematical characterization of their privacy-preserving capabilities. To address this, we present the pioneering theoretical exploration of the privacy preservation inherent in *discrete diffusion models* (DDMs) for discrete dataset generation. Focusing on per-instance differential privacy (pDP), our framework elucidates the potential privacy leakage for each data point in a given training dataset, offering insights into how the privacy loss of each point correlates with the dataset's distribution. Our bounds also show that training with s-sized data points leads to a surge in privacy leakage from $(\epsilon, \mathcal{O}(\frac{1}{s^2\epsilon}))$-pDP to $(\epsilon, \mathcal{O}(\frac{1}{s\epsilon}))$-pDP of the DDM during the transition from the pure noise to the synthetic clean data phase, and a faster decay in diffusion coefficients amplifies the privacy guarantee. Finally, we empirically verify our theoretical findings on both synthetic and real-world datasets.

Rongzhe Wei
Georgia Institute of Technology
rwei42@gatech.edu

PP3

MS46: Better Membership Inference Privacy Measurement Through Discrepancy

Membership Inference Attacks have emerged as a dominant method for empirically measuring privacy leakage from machine learning models. Here, privacy is measured by the advantage or gap between a score or a function computed on the training and the test data. A major barrier to the practical deployment of these attacks is that they do not scale to large well-generalized models – either the advantage is relatively low, or the attack involves training multiple models which is highly compute-intensive. In this work, inspired by discrepancy theory, we propose a new empirical privacy metric that is an upper bound on the advantage of a family of membership inference attacks. We show that this metric does not involve training multiple models, can be applied to large Imagenet classification models in-the-wild, and has higher advantage than existing metrics on models trained with more recent and sophisticated training recipes. Motivated by our empirical results, we also propose new membership inference attacks tailored to these training losses.

Ruihan Wu, Pengrun Huang, Kamalika Chaudhuri
University of California, San Diego
ruw076@ucsd.edu, peh006@ucsd.edu, kamalika@ucsd.edu

PP3

MS55: Conditional Regression on a Nonlinear Variable Model

We consider the problem of estimating the intrinsic structure of composite functions of the type F(X)= f(G(X)) where G is the closest point projection onto some unknown smooth curve, and f is some unknown link function. This model is a generalization of the single-index model where G is a linear rank-one matrix. We use the Conditional Regression method to extract information of the nonlinear curve, and show that we can relax the curve of dimensionality in this problem: under some assumptions restricting

the complexity of curve, our estimator for function F can achieve one-dimensional optimal minimax rate, with an extra approximation error dependent on external noise. We also perform numerical tests to verify the robustness of our algorithm, in that even if some assumptions are not satisfied, we also achieve small estimation errors.

Yantao Wu, Mauro Maggioni
Johns Hopkins University
ywu212@jhu.edu, mauromaggionijhu@icloud.com

PP3

MS39: Hybrid Least Squares

The efficient computation of conditional expectations is a central task in various data science problems across statistics, machine learning, scientific computing, and stochastic simulation. A common goal in these areas concerns the rapid assessment of conditional expectations across varying conditioning parameters, possibly infinitely many. Motivated by problems in computational finance, we introduce a surrogate-based method to approximate conditional expectation functions using least squares, with a focus on high-dimensional and noisy settings. Our approach integrates recent concepts from randomized function approximation with classical techniques from optimal experimental design, providing both theoretical guarantees and computational efficiency. Additionally, we propose an adaptive random subspace tailored for least-square approximation in the setting of stochastic simulation. Through an example of estimating the price of a basket option and employing it in downstream tasks like model calibration, we demonstrate the effectiveness of the proposed method in addressing real challenges in computational finance.

Yiming Xu
University of Utah
yxu@math.utah.edu

Ben Adcock
Simon Fraser University
ben_adcock@sfu.ca

Bernhard Hientzsch
Wells Fargo Bank
hientzschb@gmail.com

Akil Narayan
University of Utah
akil@sci.utah.edu

PP3

Broader Engagement (BE): Estimating Variance and Covariance in Practical Datasets Using Spectral Density Functions

In the presentation, I will evaluate and apply a newly proposed method for estimating the spectral density function at frequencies near zero, as detailed by McElroy and Politis (2023). Traditional spectral density estimation techniques, such as kernel smoothing of the periodogram, often face challenges at boundary points. The paper introduces a local polynomial regression approach to address these challenges, particularly focusing on the critical frequency which represents the large-sample variance of the sample mean. My presentation will involve applying this new method across various practical datasets to assess its effectiveness and explore potential applications. (The title and abstract are both imperfect and may need further revision.)

Xiaoman Xu
Smith College, U.S.
xxu76@smith.edu

PP3

MS55: Nonlinear Model Reduction for Slow-Fast Stochastic Systems Near Unknown Invariant Manifolds

We introduce a model reduction technique for high-dimensional stochastic systems having a low-dimensional invariant effective manifold with slow dynamics, and high-dimensional, large fast modes. Given only access to short bursts of simulation, we design an on-the-fly consistent algorithm exploring the effective state space efficiently. This construction enables fast, efficient simulation of the effective dynamics that averages out fast modes, plus estimation of crucial features of such dynamics, including the stationary distribution, identification of metastable states, and residence times and transition rates between them. We also implement our algorithm on three model systems: pinched sphere, butane model, and oscillating half-moons, where in particular, the last example has large nonlinear fast modes. The results demonstrate the accuracy, efficiency, and robustness of our method. This work is published in the Journal of Nonlinear Science (JNS): https://link.springer.com/article/10.1007/s00332-023-09998-8

Sichen Yang
Johns Hopkins University
ysc960109@gmail.com

Felix Ye
SUNY Albany
xye2@albany.edu

Mauro Maggioni
Johns Hopkins University
mauromaggionijhu@icloud.com

PP3

MS53: Improved Reconstruction Speed for 5D Free Running Motion Resolved Using a Variable Projection Augmented Lagrangian (vpal) Method

5D free-running cardiac magnetic resonance (CMR) imaging is a free-breathing CMR method that produces CT-quality images with high spatial and temporal resolution and without radiation. 5D free-running CMR uses compressed sensing-based reconstruction to recover the underlying image data from highly under-sampled k-space in the cardiac and respiratory dimensions. However, using the state-of-the-art alternating direction method of multipliers (ADMM), the iterative reconstruction method still requires several hours on a multi-GPU cluster, which is not viable for clinical usage. Therefore, a more efficient reconstruction algorithm is needed. We propose to use the advanced numerical algorithm to reduce the reconstruction time while preserving image quality. A variable projection augmented Lagrangian (VPAL) method for 5D motion-resolved image reconstruction was developed and compared with the state-of-the-art alternating direction method of multipliers (ADMM) on 15 5D free-running raw data sets. When compared to the ADMM method, VPAL reduced the reconstruction time by 60%, preserved image similarity, had equivalent ejection fraction measurements, and had

superior radiologist ratings. This study shows that using an advanced numerical algorithm for highly under-sampled MR reconstruction both reduces computational time and results in better image quality for diagnostics, bringing 5D free-running imaging is closer to clinical usage.

Yitong Yang
Georgia Institute of Technology and Emory University, U.S.
yitong.yang@emory.edu

Matthias Chung
Emory University, U.S.
matthias.chung@emory.edu

Aws Hamid
Emory University Hospital, U.S.
aws.hamid@emory.edu

Jerome Yerly
Lausanne University Hospital, Switzerland
jerome.yerly@chuv.ch

Davide Piccini
Siemens Healthineers International AG, Switzerland
avide.piccini@siemens-healthineers.com

Matthias Stuber
Lausanne University Hospital, Switzerland
mstuber.mri@gmail.com

John N. Oshinski
Emory University, U.S.
jnoshin@emory.edu

PP3

Interpretable Graph Neural Networks for Disease Classification on High-Dimensional Omics Data

Omics data play crucial roles in exploring disease pathways, forecasting clinical outcomes, and gaining insights for disease classification. However, the significant challenge of dealing with a relatively small number of samples and a large number of features complicates the development of predictive models for omics data analysis, with inherent sparsity in biological networks and the presence of unknown feature interactions adding further complexities. The advent of Graph Neural Networks (GNN) helps alleviate these challenges by incorporating known functional relationships over a graph. However, many existing GNN models utilize graphs either from existing networks or the generated ones alone, which limits model effectiveness. To overcome this restriction, we proposed an innovative GNN model that integrates information from both externally and internally generated feature graphs. We extensively tested the model through simulations and real data applications, confirming its superior performance in classification tasks compared to existing state-of-the-art baseline models. Furthermore, our GNN model can select features with meaningful interpretations in the biomedical context.

Tiantian Yang
Department of Mathematics and Statistical Science
University of Idaho
tyang@uidaho.edu

PP3

MS52: Extrapolation in Data-Driven Learning of Complex Reactive Processes

Reactive processes are complicated nonlinear dynamical systems that are difficult to model and interpret. While ab initio molecular dynamics (MD) is a key computational method that can be used to interrogate reactive chemistry, many challenges remain due to its computational limitations in time and length scales. In this work, we first discuss how it is possible to use the wealth of data generated by a few expensive MD simulations to statistically learn fast kinetic Monte Carlo (KMC) models or evolutionary equations describing complex chemistry. Then, we show how learning a family of reaction networks under different conditions such as temperature and pressure can lead to models that provide more accurate predictions of dynamics in extrapolated regimes. Finally, we demonstrate how our method can be used to predict the evolution of dislocation densities in material at different shock velocities, using a model trained on only a few MD simulations.

Yushuo Niu, Roshan Sebastian, Avinash Dongare, Qian Yang
University of Connecticut
yushuo.niu@uconn.edu, roshan.sebastian@uconn.edu, avinash.dongare@uconn.edu, qyang@uconn.edu

PP3

MS56: A Model-Based Approach for Continuous-Time Policy Evaluation with Unknown Lvy Process Dynamics

This research presents a framework for evaluating policies in a continuous-time setting, where the dynamics are unknown and represented by Lvy processes. Initially, we estimate the model using available trajectory data, followed by solving the associated PDE to conduct the policy evaluation. Our approach encompasses not only the conventional Brownian motion but also the non-Gaussian and heavy-tailed Lvy processes. We have developed an algorithm that demonstrates enhanced performance compared to existing techniques tailored for Brownian motion. Furthermore, we provide a theoretical guarantee regarding the error in policy evaluation given the model error. Experimental results involving both light-tailed and heavy-tailed data will be presented. This research provides a first step to continuous-time model-based reinforcement learning, particularly in scenarios characterized by irregular, heavy-tailed dynamics.

Qihao Ye
University of California, San Diego
q8ye@ucsd.edu

Xiaochuan Tian
UC San Diego
xctian@ucsd.edu

Yuhua Zhu
University of California, Los Angeles
yuhuazhu@ucla.edu

PP3

MS43: Discovery of Governing Equations Using Sparse and Low-Rank Tensor Regression

Data-driven methods for identifying governing equations are gaining popularity by enabling discovery without first-principle derivation. In this paper, we present a novel and efficient tensor regression framework designed to identify

unknown governing equations from data. Our method excels in explicitly discovering governing equations and remains robust and efficient, even when utilizing an extensive dictionary. We demonstrate the effectiveness of our approach with numerical examples involving one-dimensional PDEs and high-dimensional systems. Additionally, we provide theoretical convergence guarantees and estimation error for our algorithm.

Chugang Yi
University of Maryland, College Park
chugang@umd.edu

Yuehaw Khoo
University of Chicago
ykhoo@uchicago.edu

Haizhao Yang
University of Maryland, College Park
hzyang@umd.edu

PP3

MS57: Learning the Geometry-Dependent Solution Operators of Partial Differential Equations

Solving partial differential equations (PDEs) using numerical methods is a ubiquitous task in engineering and medicine. However, the computational costs can be prohibitively high when many-query evaluations of PDE solutions on multiple geometries are needed. Although a school of studies in artificial intelligence (AI) focused on learning PDE operators on a fixed geometry, frameworks that can alleviate the computational burdens on multiple geometries are yet to be developed. We aim to address the challenge by introducing a generic AI framework, named DIffeomorphic Mapping Operator learNing (DIMON), which allows AI to learn geometry-dependent solution operators of different types of PDEs on a wide variety of geometries. We present several examples to demonstrate the performance of the framework in learning both static and time-dependent PDEs on parameterized and non-parameterized domains; these include solving the Laplace equations, reaction-diffusion equations, and a system of multiscale PDEs that characterizes the electrical propagation on thousands of personalized heart digital twins. Accurate and efficient, DIMON can reduce the computational costs of solution approximations on multiple geometries from hours to seconds with significantly less computational resources, thus ushering in fast prediction of PDE solutions with AI on multiple geometries, and advancing the application of AI in engineering and precision medicine.

Minglang Yin
Brown University
myin16@jhu.edu

Nicolas Charon
University of Houston
ncharon@central.uh.edu

Ryan Brody
Johns Hopkins University
rbrody2@jhmi.edu

Lu Lu
Yale University
lu.lu@yale.edu

Natalia Trayanova, Mauro Maggioni
Johns Hopkins University
ntrayanova@me.com, mauromaggionijhu@icloud.com

PP3

MS46: Differentially Private Representation Learning Via Image Captioning

Differentially private (DP) machine learning is considered the gold-standard solution for training a model from sensitive data while still preserving privacy. However, a major barrier to achieving this ideal is its sub-optimal privacy-accuracy trade-off, which is particularly visible in DP representation learning. Specifically, it has been shown that under modest privacy budgets, most models learn representations that are not significantly better than hand-crafted features. In this work, we show that effective DP representation learning can be done via image captioning and scaling up to internet-scale multimodal datasets. Through a series of engineering tricks, we successfully train a DP image captioner (DP-Cap) on a 233M subset of LAION-2B from scratch using a reasonable amount of computation, and obtaining unprecedented high-quality image features that can be used in a variety of downstream vision and vision-language tasks. For example, under a privacy budget of e = 8, a linear classifier trained on top of learned DP-Cap features attains 65.8% accuracy on ImageNet-1K, considerably improving the previous SOTA of 56.5%. Our work challenges the prevailing sentiment that high-utility DP representation learning cannot be achieved by training from scratch.

Yaodong Yu
Meta
yaodongyu@meta.com

PP3

MS58: Improving the Ensemble Score Filter for Sparse and Nonlinear Data Assimilation Problem by Compositing State Dynamics to Observation Function

Recently, the Ensemble Score Filter (EnSF) was introduced to tackle high-dimensional nonlinear data assimilation(DA) problems. The primary limitation of EnSF is its tendency to underestimate the correlation structure within the filtering distribution, rendering it less effective for DA problems with sparse observations. This issue arises because the updates to unobserved dimensions rely heavily on their correlation with observed dimensions. To overcome this, we composite the state dynamics with the observation function in the Bayesian sampling step of EnSF. This composition creates non-zero gradients and updates to the unobserved dimensions without the need for a strong prior correlation structure. Moreover, we employ automatic differentiation and likelihood bias correction techniques to enhance the efficiency of Bayesian sampling in EnSF. Extensive numerical experiments demonstrate that our method maintains accuracy and effectiveness even under conditions of sparse observation.

Zezhong Zhang
Oak Ridge National Laboratory
zhangz2@ornl.gov

PP3

MS55: Minimax Rates for Learning Kernels in Ill-

Posed Inverse Problems

Kernels are efficient in representing nonlocal dependencies and are widely used to design operators between function spaces. Thus, learning kernels in operators from data is of broad interest, presenting a new topic at the intersection of statistical learning and inverse problems. A fundamental question in this context is determining the minimax rate. However, two major challenges distinguish this from classical nonparametric regression and functional data analysis: (i) the absence of universal spaces, such as Sobolev spaces, for learning, and (ii) the ill-posedness of the inverse problem. We address these challenges by first introducing adaptive weighted Sobolev spaces based on the normal operator in the large sample limit. Then, we establish the optimal minimax rate for the mean square error through the study of a tamed least squares estimator. A core technical result is a nonasymptotic estimate for the left tail probability of the smallest eigenvalue of the random normal matrix.

Sichong Zhang, Fei Lu, Xiong Wang
Johns Hopkins University
szhan143@jhu.edu, feilu@math.jhu.edu,
xiong_wang@jhu.edu

PP3

Broader Engagement (BE): AI Foundation Models for Surgical Tool Localization

Recent advancements in foundational language and vision models have achieved state-of-the-art accuracy. Yet, there are still gaps in these models ability to understand and respond to tasks in specific areas, particularly scientific and biomedical domains. This study investigates the performance of these models applied to minimally invasive robotic surgery, by validating their ability to localize and identify surgical tools in endoscopic videos. Integrating the language and vision models presents an opportunity to prompt these models in a natural, intuitive way even for those without a computing or data science background. For instance, one can segment all surgical tools within a robotic surgery scene with simple text prompts like metal or tool. While we experimented with AI-generated descriptions from the CLIP-interrogator, and using those as text prompts for SAM, the overlapping keywords presented challenges due to the models' limited surgical domain expertise. Understanding the visual content in surgical scenes, having the trained vocabulary to prompt and infer from these models is vital to adoption of more complex techniques like semantic action recognition, or long term activity prediction. Addressing this, we will further explore fine-tuning models on robotic surgery datasets to increase contextual awareness in the surgical domain, such as segmenting particular surgical instruments or organs based on name-based prompts.

Ruoxi Zhao
University of California, Merced
rzhao1129@gmail.com

Neil Getty
Argonne National Lab
ngetty@anl.gov

PP3

MS56: Solving Time-Continuous Stochastic Optimal Control Problems: Algorithm Design and Convergence Analysis of Actor-Critic Flow

We propose an actor-critic framework to solve the time-continuous stochastic optimal control problem. A least square temporal difference method is applied to compute the value function for the critic. The policy gradient method is implemented as policy improvement for the actor. Our key contribution lies in establishing the global convergence property of our proposed actor-critic flow, demonstrating a linear rate of convergence. Theoretical findings are further validated through numerical examples, showing the efficacy of our approach in practical applications.

Mo Zhou
University of California, Los Angeles
mozhou366@math.ucla.edu

PP3

MS56: Phibe: Physics-Informed Bellman Equation for Continuous-Time Reinforcement Learning

In this poster, we address the problem of continuous-time reinforcement learning in scenarios where the dynamics follow a stochastic differential equation. When the underlying dynamics remain unknown, and we have access only to discrete-time information, how can we effectively find the optimal policy? We first highlight that the commonly used Bellman equation is not always a reliable approximation to the true value function. We then introduce PhiBE, a PDE-based Bellman equation that offers a more accurate approximation to the true value function, especially in scenarios where the underlying dynamics change slowly. Moreover, we extend PhiBE to higher orders, providing increasingly accurate approximations. Additionally, we present a model-free algorithm to solve PhiBE when only discrete-time trajectory data is available. Numerical experiments are provided to validate the theoretical guarantees we propose.

Yuhua Zhu
University of California, Los Angeles
yuhuazhu@ucla.edu

PP4

MS67: Hard Negative Sampling Via Regularized Optimal Transport for Contrastive Representation Learning

We study the problem of designing hard negative sampling distributions for unsupervised contrastive representation learning. We propose and analyze a novel min-max framework that seeks a representation which minimizes the maximum (worst-case) generalized contrastive learning loss over all couplings (joint distributions between positive and negative samples subject to marginal constraints) and prove that the resulting min-max optimum representation will be degenerate. This provides the first theoretical justification for incorporating additional regularization constraints on the couplings. We re-interpret the min-max problem through the lens of Optimal Transport (OT) theory and utilize regularized transport couplings to control the degree of hardness of negative examples. Through experiments we demonstrate that the negative samples generated from our designed negative distribution are more similar to the anchor than those generated from the baseline negative distribution.

Shuchin Aeron
Tufts University

shuchin@ece.tufts.edu

PP4

MS72: Data-Driven Approximation of Stationary Nonlinear Filters with Optimal Transport Maps

The nonlinear filtering problem aims to determine the conditional probability distribution (posterior) of the state of a stochastic dynamical system given a history of partial and noisy observations. This poster introduces a data-driven nonlinear filtering algorithm when the state and observation processes are stationary. We approximate the posterior using an optimal transport (OT) map, which is the push-forward from a given distribution that is easy to sample from to the posterior, conditioned on a truncated observation window. The OT map is obtained as the solution to a stochastic optimization problem that is solved offline using recorded trajectory data from the state and observations. We present an error analysis of the algorithm under the stationarity and filter stability assumptions, which decomposes the error into two parts related to the truncation window during training and the error due to the optimization procedure. The proposed optimal transport data-driven filter (OT-DDF) demonstrates considerable computational efficiency during the online stage and maintains the flexibility and accuracy of OT in nonlinear filtering through various numerical examples.

Mohammad Al-Jarrah
University of Washington
mohd9485@uw.edu

PP4

Heat and Mass Transfer Through Mhd Darcy Forchheimer Casson Hybrid Nanofluid Flow Across An Exponential Stretching Sheet

This work aims to study the energy and mass transition caused by Casson hybrid nanofluid flow across an extended stretching sheet. Fluid flow is subjected to an inclined magnetic field to control the flow stream. Cu and Al_2O_3 NPs are added to the Casson fluid to generate a hybrid nanoliquid (Blood). This model of flow dynamics is an evolving nonlinear system of PDEs, which is then reduced to a system of dimensionless ODEs using similarity proxies. The ODEs is solved using the analytical program HAM for further processing.

Aisha Alqahtani
Heriot Watt University
alqahtani@pnu.edu.sa

PP4

MS72: Critical Points and Convergence Analysis of Generative Deep Linear Networks Trained with Bures-Wasserstein Loss

We consider a deep matrix factorization model of covariance matrices trained with the Bures-Wasserstein distance. While recent works have made advances in the study of the optimization problem for overparametrized low-rank matrix approximation, much emphasis has been placed on discriminative settings and the square loss. In contrast, our model considers another type of loss and connects with the generative setting. We characterize the critical points and minimizers of the Bures-Wasserstein distance over the space of rank-bounded matrices. The Hessian of this loss at low-rank matrices can theoretically blow up, which creates challenges to analyze convergence of gradient optimization methods. We establish convergence results for gradient flow using a smooth perturbative version of the loss as well as convergence results for finite step size gradient descent under certain assumptions on the initial weights.

Jing An
Department of Mathematics
Duke University
jing.an@duke.edu

PP4

MS77: Leveraging Perceiver IO and Relative Position Encodings for Enhanced Node Classification

Graph transformer models have been limited in their application to large-scale graphs due to the quadratic computational complexity inherent to their attention mechanisms. This often results in a trade-off between scalability and model performance. In this work, we introduce a novel graph transformer architecture inspired by the PerceiverIO framework, which utilizes a combination of latent compression and relative position encoding to efficiently scale graph attention to larger graphs. Our approach mitigates the quadratic cost of pairwise communication between all nodes in a graph by learning a set of latent tokens through which nodes exchange messages. We evaluate our model on several node classification benchmarks and demonstrate that it not only surpasses existing graph transformer models in terms of performance but also maintains high efficiency and adaptability across different graph structures. Overall, our architecture offers a scalable solution for efficiently processing large graph datasets while significantly improving model performance and generalization.

Vinam Arora, Divyansha Lachi
Georgia Institute of Technology
vinam@gatech.edu, divyansha.lachi@gatech.edu

PP4

MS76: Understanding Gaussian-Stein Variational Gradient Descent

For sampling from a Gaussian target, the SVGD dynamics with a bilinear kernel will remain Gaussian as long as the initializer is Gaussian. Inspired by this fact, we undertake a detailed theoretical study of the Gaussian-SVGD, ie, SVGD projected to the family of Gaussian distributions via the bilinear kernel, or equivalently Gaussian variational inference (GVI) with SVGD. We present a complete picture by considering both the mean-field PDE and discrete particle systems. When the target is strongly log-concave, the mean-field Gaussian-SVGD dynamics is proven to converge linearly to the Gaussian distribution closest to the target in KL divergence. In the finite-particle setting, there is both uniform in time convergence to the mean-field limit and linear convergence in time to the equilibrium if the target is Gaussian. In the general case, we propose a density-based and a particle-based implementation of the Gaussian-SVGD, and show that several recent algorithms for GVI, proposed from different perspectives, emerge as special cases of our unified framework. Interestingly, one of the new particle-based instance from this framework empirically outperforms existing approaches. Our results make concrete contributions towards obtaining a deeper understanding of both SVGD and GVI.

Krishnakumar Balasubramanian
University of California, Davis, U.S.

kbala@ucdavis.edu

PP4

MS67: An Efficient Streaming Algorithm for Sparse Eigenvector

In the era of big data, handling large datasets efficiently presents formidable challenges. Our research introduces a novel algorithm tailored to approximate sparse eigenvectors of large matrices. Given a positive semi-definite $A \in \mathbb{C}^{N\times N}$, we show that MAM^* where M is a carefully chosen matrix, provides an effective way to approximate the eigenvectors of A. This drastically reduces the need for frequent memory accesses to the original matrix A. Furthermore, our algorithm is accompanied by theoretical guarantees, ensuring reliability and robustness.

Edem Boahen, Mark A. Iwen
Michigan State University
boahened@msu.edu, markiwen@math.msu.edu

Felix Krahmer
Technical University of Munich
Department of Mathematics
felix.krahmer@tum.de

Simone Brugiapaglia
Concordia University
simone.brugiapaglia@concordia.ca

Hung-Hsu Chou
Technical University of Munich
bemyself26@g.ucla.edu

PP4

MS71: Techniques for Reducing Dimensions and Quantifying Uncertainty in Dynamical Systems

Understanding and predicting the behavior of dynamical systems through partial observations is a critical challenge across various disciplines. In this research, we work towards a robust framework that integrates analytical approaches for dimensional reduction with a Bayesian framework for uncertainty quantification, tailored to the unique complexities of partially observed dynamical systems with inherent noise and symmetries. The work centers on a system observed over time, subject to additional noise and incomplete information retrieval through an observation function. This setup commonly results in data that are either sparse or redundant, complicating the process of making accurate inferences. By leveraging dimensional reduction, our approach effectively simplifies the complexity of the data, while preserving essential system dynamics. Besides this, we employ a Bayesian framework specifically designed to address uncertainties in parameter inference. We will show our results for linear cases and some preliminary results in non-linear cases. The flexibility of our approach, agnostic to specific machine learning models, enables its application across a broad spectrum of disciplines, paving the way for more precise and reliable predictions in scenarios where traditional models might falter due to incomplete or noisy data.

Andrey Bryutkin
Mathematics
MIT
bryutkin@mit.edu

Matthew Levine
Broad Institute
mattlevine22@gmail.com

PP4

Finding Multiple Optimal Solutions

Recent papers have shown that, after fixing a loss function and a dataset, models tend to converge to similar probability distributions regardless of initialization and dataset shuffling. It is thus tempting to conclude that multiple models trained on a given dataset with a given loss will converge to a fixed interpretation (that is, the relative importance of features will be roughly the same for each model). In our work, we show empirically that although randomly initialized models do tend to converge to the same interpretation, other interpretations with similar training and test accuracy metrics exist in at least some datasets. We also introduce a method to find models that take on these different interpretations. Exploring these alternate models is important, as some may have interpretations that more closely reflect the underlying patterns than those found through random initialization.

Matthew Chak, Paul Anderson
California Polytechnic State University, San Luis Obispo
mchak@calpoly.edu, pander14@calpoly.edu

PP4

MS75: Euclidean Mirrors and First-Order Changepoints in Network Time Series

We describe a model for a network time series whose evolution is governed by an underlying stochastic process, known as the latent position process, in which network evolution can be represented in Euclidean space by a curve, called the Euclidean mirror. We define the notion of a first-order changepoint for a time series of networks, and construct a family of latent position process networks with underlying first-order changepoints. We prove that a spectral estimate of the associated Euclidean mirror localizes these changepoints, even when the graph distribution evolves continuously, but at a rate that changes. Simulated and real data examples on organoid networks show that this localization captures empirically significant shifts in network evolution.

Tianyi Chen
Johns Hopkins University
tchen94@jh.edu

Zachary Lubberts
University of Virginia
zlubberts@virginia.edu

Avanti Athreya
Johns Hopkins University
Department of Applied Mathematics and Statistics
dathrey1@jhu.edu

Youngser Park, Carey Priebe
Johns Hopkins University
youngser@jhu.edu, cep@jhu.edu

PP4

MS72: When Can Regression-Adjusted Control Variate Help? Rare Events, Sobolev Embedding

and Minimax Optimality

This paper studies the use of a machine learning-based estimator as a control variate for mitigating the variance of Monte Carlo sampling. Specifically, we seek to uncover the key factors that influence the efficiency of control variates in reducing variance. We examine a prototype estimation problem that involves simulating the moments of a Sobolev function based on observations obtained from (random) quadrature nodes. Firstly, we establish an information-theoretic lower bound for the problem. We then study a specific quadrature rule that employs a nonparametric regression-adjusted control variate to reduce the variance of the Monte Carlo simulation. We demonstrate that this kind of quadrature rule can improve the Monte Carlo rate and achieve the minimax optimal rate under a sufficient smoothness assumption. Due to the Sobolev Embedding Theorem, the sufficient smoothness assumption eliminates the existence of rare and extreme events. Finally, we show that, in the presence of rare and extreme events, a truncated version of the Monte Carlo algorithm can achieve the minimax optimal rate while the control variate cannot improve the convergence rate.

Jose H. Blanchet
Stanford University
jose.blanchet@stanford.edu

Haoxuan Chen
Institute for Computational and Mathematical Engineering
Stanford University
haoxuanc@stanford.edu

Yiping Lu
Courant Institute, New York University
Northwestern University
yiping.lu@northwestern.edu

Lexing Ying
Stanford University
Department of Mathematics
lexing@stanford.edu

PP4

MS73: Data-Driven Aerodynamic Shape Design with Distributionally Robust Optimization

We formulate and solve data-driven aerodynamic shape design problems with distributionally robust optimization (DRO) approaches. DRO aims to minimize the worst-case expected performance in a set of distributions that is informed by observed data with uncertainties. Building on the findings of the work [Gotoh el at.], we study the connections between a class of DRO and robust design optimization, which is classically based on the mean-variance (standard deviation) optimization formulation pioneered by Taguchi. Our results provide a new perspective to the understanding and formulation of robust design optimization problems. It enables data-driven and statistically principled approaches to quantify the trade-offs between robustness and performance, in contrast to the classical robust design formulation that captures uncertainty only qualitatively.

Long Chen
RPTU Kaiserslautern, Germany
long.chen@scicomp.uni-kl.de

Jan Rottmayer
RPTU Kaiserslautern
jan.rottmayer@scicomp.uni-kl.de

Lisa Kusch
TU Eindhoven
l.kusch@tue.nl

Nicolas R. Gauger
TU Kaiserslautern
nicolas.gauger@scicomp.uni-kl.de

Yinyu Ye
Stanford University
yinyu-ye@stanford.edu

PP4

MS66: Conditional Sampling with Stochastic Interpolants and Follmer Processes

In this work, we address probabilistic forecasting via generative modeling. To do so, we construct novel SDEs that map a point mass of the current state to a conditional distribution of future states, based on the framework of stochastic interplants. The SDEs can be learned from data via simulation-free training, and the drifts can be tuned a posteriori to optimize the estimation accuracy in KL divergence of path measures. We show that the optimal drift corresponds to Follmer processes. Experiments on stochastic Navier Stokes equations and video forecasting demonstrate the effectiveness and scope of this work.

Yifan Chen
NYU
yifan.chen@nyu.edu

PP4

MS70: libROM: a Free, Lightweight, Scalable C++ Library for Data-Driven Physical Simulation Methods

In decision-making applications where multiple forward simulations are needed, such as parameter study, design optimization, optimal control, uncertainty quantification, and inverse problems, we need to repeatedly solve forward problems. However, subject to the model complexity and the fineness of the discretization, the computational cost of forward simulations can be high. It may take a long time to complete a single forward simulation with the available computing resource. In this presentation, we will introduce various reduced order modeling techniques, which aim to lower the computational complexity and maintain a good accuracy, including projection-based intrusive nonlinear model reduction and non-intrusive model reduction approaches. We will demonstrate the implementation of these reduced order modeling techniques in libROM (www.librom.net) and its application to numerical solvers for solving various physics problems.

Siu Wun Cheung
Lawrence Livermore National Laboratory
cheung26@llnl.gov

PP4

Reassessing Relationality for Bipolar Data

Methods to cluster people across relationality, a concept developed to describe how similarly individuals respond to

survey items measuring the intensity of agreement or disagreement towards specific question statements, have recently gained traction. These methods cluster people into social affinity groups, sharing similarities in their outlooks on social life. Relational Class Analysis (RCA) is currently the most commonly used method for relationality clustering. RCA has been applied to identify affinity groups in social spheres as varied as politics, musical preferences, and attitudes towards science. In this study, we highlight limitations in RCAs ability to accurately identify the number and underlying structure of these groups. These limitations stem from RCAs mathematical underpinnings and its insensitivity to the orientational quality of the survey items they use: they demand them to place themselves in a support or rejection zone and then express how intensely their support or rejection is. We develop a method, called Bipolar Class Analysis (BCA), that aims to address this foundational limitation. BCA conceptualizes peoples attitudinal positions as moving along agree/disagree subspaces and assesses relationality by taking into account position switches across these subspaces. We run extensive simulation analyses of data organized around different relationality structures to show that BCA is more accurate than RCA and other available alternatives.

Manuel Cuerno
CUNEF University
manuel.mellado@cunef.edu

Sergio Galaz-García
Universidad Carlos III of Madrid
sgalaz@uc3m.es

Fernando Galaz-García
Durham University
fernando.galaz-garcia@durham.ac.uk

Telmo Pérez-Izquierdo
BCAM / UPV
tperezizquierdo@bcamath.org

PP4

Image-Based Goal Conditioned Reinforcement Learning with a Gflownet Planner and Transformer Policy

Most tasks faced by autonomous agents in practical settings are long-horizon tasks. The complex nature of long-horizon tasks arises from the delay in obtaining a corresponding reward for a long sequence of actions. Goal-conditioned reinforcement learning is an approach to tackle long-horizon tasks, which uses a planner followed by a policy to achieve the final goal. The key challenges in this approach are two-fold: finding the optimal sequence of subgoals and finding a policy with enough generalizability to attain the final goal. Previous approaches with conventional policies require pre-training for each distinct long-horizon task. However, it is known that pre-trained transformers generalize well across different tasks, which would imply a few-shot and zero-shot learning approach for the policy. Thus, the proposed model explores the use of such a transformer-based policy, which can successfully generalize to new tasks without pre-training for each distinct long-horizon task. Additionally, a novel planner is proposed based on Generative Flow Networks which samples objects from a distribution in a sequence of constructive steps. This planner would sample the optimal sequence of sub-goals not by minimizing a loss function, but rather by sampling such that the generated sample probabilities are proportional to a reward function. Further, the use of large language models to add text descriptions to the corresponding image at each state is explored.

Sathish Kumar, Swaroop Damodaran
Cleveland State University
s.kumar13@csuohio.edu,
s.damodaran19@vikes.csuohio.edu

Sumit K. Jha
Florida International University
sumit.jha@fiu.edu

Arvind Ramanathan
Data Science and Learning Division
Argonne National Laboratory
ramanathana@anl.gov

PP4

Algorithmic Stability of Heavy-Tailed Stochastic Gradient Descent with Momentum

Recent studies have shown that heavy tails can appear in stochastic optimization and that the heaviness of the tails is related to the generalization error. In this work, we consider heavy-tailed stochastic gradient descent but in the case where a momentum variable is added, and are able to obtain the associated generalization bound under appropriate assumptions which allow for both convex and non-convex loss functions. Then by focusing on the case of a quadratic loss function, we are able to compare the generalization bounds of SGDs with and without momentum. Our theoretical findings are supported by numerical experiments. This is a joint work with Lingjiong Zhu.

Thanh L. Dang
Florida State University
Department of Mathematics
td22v@fsu.edu

PP4

MS63: Kernel Methods for Operator Learning

We introduce a kernel-based framework for learning operators between Banach spaces. We show that even with simple kernels, our approach is competitive in terms of cost-accuracy trade-off and either matches or beats the performance of Neural Network methods on a majority of PDE-based benchmarks. Additionally, our framework offers several advantages inherited from kernel methods: simplicity, convergence guarantees, a priori error estimates, and Bayesian UQ. It is, therefore, a natural benchmark for operator learning problems.

Matthieu Darcy
Caltech
mdarcy@caltech.edu

PP4

Learning Salient Semantic Subsequences to Explain Deep Time Series Classifiers

Explainable AI helps end users trust deep learning solutions for time series classification. However, existing explainability methods for time series classifiers focus on explaining individual time points to generate a local view or they generate shapelets based explanations to obtain a global view. Time-points based explanations are too hard

for end users to understand without a higher-level abstraction. None of these methods produce semantically meaningful explanations. Semantic based methods produce explanations using standard time series characteristics such as burst, pulses, trends, and seasonality. End users can easily understand the explanation in terms of time series characteristics. In this paper, we formalize the notion of semantic based explanations, studying the open problem of semantic-based explainability for deep time series classifiers, a challenging, and meaningful problem setting. We design a novel explainability method, SMUX, which learns semantically meaningful subsequences for explaining deep time series classifiers by adaptively ensuring that its explanation spotlights the regions in an input time series that a model uses specifically to its predicted class. SMUX adopts a gradient-based approach composed of three interdependent modules that combine to generate consistent, class-specific semantic subsequences that remain faithful to the classifier's behavior yet are intuitively understood by end users.

Aurongzeb Deeder
Johns Hopkins University
daurong1@jh.edu

Ramesh Doddaiah
Worcester Polytechnic Institute
rdoddaiah@wpi.edu

PP4

Neural Operator Simulation of Turbulent Multiphase Flow with Applications in Ocean Wave Modeling

Global warming has increased wave heights and the frequency of extreme wave events. As a result, ocean waves have been contributing to more coastal erosion, flooding, and fatigue damage of ocean-based structures. Researchers interested in modeling the effects of ocean waves usually rely on slow iterative methods such as particle-in-cell (PIC). Therefore, researchers have good occasion to find a faster alternative for modeling ocean waves. To this end, we have created a Neural Operator (NO) that can predict ocean wave behavior on grids with random obstructions. To create data to train our NO, we used the particle-in-cell method to create 290 examples of an ocean wave crashing against a random obstruction. Each simulation modeled 38,000 water particles for 600 time steps, which required 30 minutes. From these simulations, we created training and testing datasets of 274 and 16 image pairs, respectively, where the input image is the grid with the random obstruction, and the output image is the shape of the ocean wave. After training the NO, we validated it against our 16 test points. At a resolution of 256 x 256, our NO had an inference time of only 0.073 second. The NO results were quite accurate, especially given that only 274 training points were used. Our approach can be expanded to create fast 3D time-varying simulations. These NOs can be used to predict levee overtopping, flooding, blast waves, and even the effect of water drops over wildfires.

Joshua Deng
Orange Lutheran High School
Joshd2026@outlook.com

PP4

Bayesian D-Optimal Experimental Design for Weak-Constraint 4D-Var

In data assimilation, the weak-constraint approach provides a method to incorporate model errors in the dynamics of the governing equations. We propose a new framework for optimal sensor placement in the weak-constrained setting, by first deriving the D-optimal criterion assuming that the model error is Gaussian and the dynamics are linear. We discuss algorithmic approaches to efficiently evaluate this criterion and provide an algorithm to find near-optimal experimental designs using column subset selection. We show the effectiveness of our approach and the computational benefits of our algorithms on model problems.

Hugo Diaz
North Carolina State University
hsdiazno@ncsu.edu

PP4

Machine Learning Vs Copula-Based Quantile Regression Methods for Prediction of Petrophysical Properties with Seismic Attributes: A Comparative Study

In recent years, machine learning methods have become popular to solve many data-driven problems in earth sciences. But this does not necessarily imply that they are always the only and the best option from the point of view of precision and efficiency. This work presents a comparison of several standard machine learning methods with respect to the copula-based quantile regression method for the prediction of petrophysical properties from seismic attributes in the framework of geological-petrophysical modeling of oil fields. From the methodological point of view, a systematic methodology of Data Science is applied, showing each stage of the application of the methods, which includes the statistical analysis of the data, the modeling of the joint dependency relationship, training, validation and prediction. The results of petrophysical property predictions from both approaches are compared with reference data at the well log scale. Finally, both approaches are applied to a real case study both at the well log scale and at the seismic scale in a 2D inline section.

Martin A. Diaz-Viera
INSTITUTO MEXICANO DEL PETROLEO
mdiazv@imp.mx

Van Huong Le
University of Delaware
levanhuong15011989@gmail.com

Rodrigo Lopez-Aburto
Posgrado del Instituto Mexicano del Petroleo
lopezaburtorodrigo@gmail.com

PP4

Lower Bounds for Private Online Learning

We provide lower bounds for Differentially Private (DP) Online Learning algorithms. Our result shows that, for a broad class of (ε, δ)-DP online algorithms, for number of steps T such that $\log T \leq O(1/\delta)$, the expected number of mistakes incurred by the algorithm grows as $\Omega(\log T/\delta)$. This matches the upper bound obtained by Golowich and Livni (2021) and is in contrast to non-private online learning where the number of mistakes is independent of T. Our work partially addresses the open question in Sanyal and

Ramponi (2022).

Daniil Dmitriev
ETH Zurich
dmitrievdaniil97@gmail.com

Amartya Sanyal
Department of Computer Science, University of Copenhagen
amartya18x@gmail.com

Kristóf Szabó
ETH Zurich
krist.sz13@gmail.com

PP4

MS66: Robust Blockwise Random Pivoting

The interpolative decomposition (ID) aims to construct a low-rank approximation formed by a basis consisting of row/column skeletons in the original matrix and a corresponding interpolation matrix. This work explores fast and accurate ID algorithms from five essential perspectives for empirical performance: (a) skeleton complexity that measures the minimum possible ID rank for a given low-rank approximation error, (b) asymptotic complexity in FLOPs, (c) parallelizability of the computational bottleneck as matrix-matrix multiplications, (d) error-revealing property that enables automatic rank detection for given error tolerances without prior knowledge of target ranks, (e) ID-revealing property that ensures efficient construction of the optimal interpolation matrix after selecting the skeletons. While a broad spectrum of algorithms have been developed to optimize parts of the aforementioned perspectives, practical ID algorithms proficient in all perspectives remain absent. To fill in the gap, we introduce robust blockwise random pivoting (RBRP) that is parallelizable, error-revealing, and exact-ID-revealing, with comparable skeleton and asymptotic complexities to the best existing ID algorithms in practice. Through extensive numerical experiments on various synthetic and natural datasets, we empirically demonstrate the appealing performance of RBRP from the five perspectives above, as well as the robustness of RBRP to adversarial inputs.

Yijun Dong
New York University
yd1319@nyu.edu

PP4

Surrogate Modeling for Particle Accelerator Simulations Using Point Clouds and Graph Networks

Simulations are crucial tools in particle physics, enabling the design and optimization of modern particle accelerator systems. These simulations face a significant challenge in balancing speed and accuracy, especially when dealing with complex beam behaviors and interactions among numerous accelerator components. The computational intensity of these simulations presents a substantial hurdle for comprehensive studies. In this work, we leverage data-driven machine learning techniques, with a focus on point cloud networks and graph networks, to build surrogate models that help address the computational challenges and draw physical insights about the underlying system. Point cloud networks efficiently extract symmetry-invariant global features from large-scale datasets; graph networks facilitate modeling physical interactions between particles via various message-passing algorithms. We combine the two ideas in a framework to enable efficient surrogate modeling that captures particle interactions in a reduced dimensional space. We test the proposed framework in the context of particle accelerator simulations, including predicting downstream beam parameters and image-based model diagnostics.

Tiffany Fan, Murray C. Cutforth
Stanford University
tiffan@stanford.edu, mcc4@stanford.edu

Auralee Edelen
SLAC National Accelerator Laboratory
edelen@slac.stanford.edu

Marta D'Elia
Pasteur Labs, Stanford University
marta.delia@simulation.science

Eric F. Darve
Institute for Computational and Mathematical Engineering
Stanford University
darve@stanford.edu

PP4

Convolutional Framelets Using a Tensor Representation

Patch-based representations of images have shown success as an efficient signal representation in image analysis. Convolutional framelets is a novel representation of one-dimensional signals which integrates local and nonlocal properties of the signal. The potential of this approach led us to investigate generalizing the framework to two-dimensional signals. Our approach incorporates techniques from tensor analysis with the aim of preserving the structure of the original signal. We then show numerical results which suggests our approach is better than the convolutional framelets approach where the respective representations are derived from a higher order singular value decomposition and the singular value decomposition. Our numerical comparison adapts a method, which to our knowledge is one of the few, that quantitatively shows that tensor representations, particularly the t-svd, are quantitatively better than vectorized representations of multi-dimensional data.

Nickolas Fularczyk
Harbor Branch Oceanographic Institute
Florida Atlantic University
nfularczyk@fau.edu

Demetrio Labate
University of Houston
dlabate@math.uh.edu

PP4

MS70: Data-Driven Approaches for Solving the Allen-Cahn Equation

The Allen-Cahn equation describes the process of phase separation and transition in phase field modeling of multi-component physical systems. High-fidelity numerical simulations of the Allen-Cahn equation can be computationally expensive due to its stiffness, compounded by the necessity for multiple simulations in parametric Allen-Cahn scenarios within a multi-query setting. In this presenta-

tion, we will introduce two approaches to accelerate the numerical simulations of the Allen-Cahn equation. One approach is based on model order reduction, wherein we develop a structure-preserving, gradient-preserving operator inference technique for learning the reduced operators. The other approach is deep learning-based, involving the design of specialized convolutional neural network models to learn fully-discrete operators. We will present numerical experiments demonstrating the efficiency of the proposed methods.

Yuwei Geng
University of South Carolina
ygeng@email.sc.edu

Zhu Wang
Department of Mathematics
University of South Carolina
wangzhu@math.sc.edu

Lili Ju
University of South Carolina
Department of Mathematics
ju@math.sc.edu

PP4

MS75: Solving Euclidean Distance Geometry Problems from Few Samples Via Iteratively Reweighted Least Squares

Euclidean Distance Geometry (EDG) problems arise in molecular conformation, sensor network localization and in manifold learning, where the provided data includes pairwise distance information, and the goal is to find an Euclidean point embedding from a minimal number of sampled pairwise distances. We propose an iterative algorithm based on iteratively reweighted least squares (IRLS) which minimizes quadratic models derived from a series of continuously differentiable, non-convex relaxations of the rank function. We establish a local convergence analysis of this IRLS method which applies if a minimal random set of observed distance is provided. As a technical tool, we establish a Restricted Isometry Property for EDG measurement operators on the rank-r manifold tangent space which might be of independent interest for the analysis of other non-convex approaches. Furthermore, we assess data efficiency, scalability and generalizability through numerical experiments with simulated data as well as real-world data, demonstrating the proposed algorithm's ability to identify the underlying geometry from fewer distance samples compared to the state-of-the-art.

Ipsita Ghosh
University of North Carolina at Charlotte
ighosh2@charlotte.edu

Christian Kümmerle
Johns Hopkins University
kuemmerle@uncc.edu

Abiy F. Tasissa
Tufts University
Abiy.Tasissa@tufts.edu

PP4

Deconstructing Reconstruction: Structural Biases in Networks Reconstructed from Time Series Dynamics

Many real-world systemsfrom biology to financial marketsare governed by complex networks of interactions where the network structures are either directly unobservable or must be inferred from data. Often, researchers perform this network reconstruction task by using time series recordings of node-level dynamics in the system. A variety of network reconstruction techniques exist, ranging from causal methods for time series analysis to statistical methods. Despite the multitude of reconstruction techniques, many of the techniques underlying biases remain unexamined. Our study explores these biases using network models which control for specific network statistics and a pipeline of network reconstruction which allows us to test diverse time series analysis and reconstruction approaches. Across many different pipeline variations (i.e., dynamical process, time series analysis, reconstruction technique, sparsification procedure), we find that reconstructed networks consistently show characteristic values for a variety of summary statistics, regardless of the original networks characteristics. Our study offers a new perspective on the structural biases in network reconstruction and hints at potential limitations in reconstructing networks with dynamically redundant edges. Additionally, we scrutinize the preserved network statistics in these reconstructions, offering insights into the network properties targeted by reconstruction methods.

Cory Glover, Erik Weis, Sagar Kumar, Joey Ehlert, Brennan Klein
Northeastern University
glover.co@northeastern.edu, erik.weis@gmail.com, kumar.sag@northeastern.edu, ehlert.j@northeastern.edu, b.klein@northeastern.edu

PP4

MS68: A Gaussian Process Language Model

We present a Gaussian process language model (GPLM). GPLM is a two-level hierarchical Bayesian model, with the base level representing categorical token probabilities as a function of sequence location, and the top level establishing correlations among those token probabilities by means of a multi-output Gaussian process over a latent space that maps to the simplex of categorical probabilities. The latent-space model is stationary with respect to sequence location, which permits covariances to be estimated directly from the data without resort to parameter optimization. We show results obtained using genomic sequences to predict (infill) sequence gaps and predict likely regions of variation.

Carlo Graziani
Argonne National Laboratory
cgraziani@anl.gov

PP4

Utilizing Dynamical Systems and Multipopulation Models to Track Disease Spread in Cuba: A Focus on Dengue, Covid-19, and HIV

The management and prediction of infectious diseases such as dengue, COVID-19, and HIV are critical challenges in public health, particularly in regions like Cuba, where these diseases significantly impact the population. This work introduces innovative approaches using dynamical systems and multipopulation models to understand and predict the spread of these diseases. By integrating real-time data and leveraging the versatility of these models, we aim to

provide robust analytical tools that can assist in making informed public health decisions. We specifically explore the effectiveness of these models in capturing the dynamics of disease transmission among different population groups and geographical distributions in Cuba. The results showcase the potential of mathematical modeling in enhancing disease surveillance and response strategies, offering insights crucial for health policy planning and implementation. This approach not only aids in managing the current health issues but also in preparing for future outbreaks, thereby contributing to a resilient health infrastructure.

Raul Guinovart-Díaz
Universidad de La Habana
guino@matcom.uh.cu

David Guinovart, Mriganka Shekha Chaki, Mohammed Qaraad
University of Minnesota
The Hormel Institute
guino001@umn.edu, chaki014@umn.edu, qaraa001@umn.edu

PP4

MS65: Control-Deeponet: Deep Operator Networks for PDE-Constrained Optimal Control under Uncertainty

In this presentation, we address the challenges of applying the DeepONet architecture to PDE-constrained optimization under uncertainty. While DeepONet has been effective in providing fast surrogates in solving time-dependent PDE-constrained optimization (PDECO) problems [Wang, Sifan, et al. 'Fast PDE-Constrained Optimization via Self-Supervised Operator Learning,' arXiv, 25 Oct. 2021], it encounters limitations in cases where the operator depends on multiple inputs. In PDECO under uncertainty, the solution operator maps the product of two Banach spaces the control space and the uncertain parameter space (e.g., heat conduction field for the heat equation, permeability coefficient for Darcy flow diffusion equation, material property in elasticity equation) to the PDE solution Banach space. To address this challenge, we employ the Multiple Input Operator Network (MIONet) [Jin, Pengzhan, et al. 'MIONet: Learning Multiple-Input Operators via Tensor Product,' SIAM Journal on Scientific Computing, vol. 44, no. 6, Dec. 2022]. MIONet provides a versatile framework for learning surrogate models that can handle operators dependent on multiple inputs. We demonstrate how this surrogate model can be effectively used to optimize risk-averse measures and obtain robust controls under uncertainty.

Sebastian Gutierrez Hernandez
Georgia Tech
shern3@gatech.edu

Haomin Zhou
Georgia Institute of Technology
School of Mathematics
hmzhou@math.gatech.edu

Chen Peng
Georgia Institute of Technology
pchen402@gatech.edu

PP4

MS65: Harnessing Data Efficiency with a Fixed-Point Iteration Analogous to Adaptive Mesh Refinement

Representing complex physical systems through statistical models necessitates extensive data collection, which can be both resource-intensive and computationally demanding. Experimentation often requires significant labor and resources, and high-fidelity simulations impose heavy computational burdens. This challenge necessitates more data-efficient regimes. Our approach draws inspiration from two concepts: mesh refinement and deep equilibrium models. Mesh refinement discretizes the computational domain into varying resolutions based on the specific complexity required by different regions of the domain, and deep equilibrium models are a type of deep learning model that map inputs to steady-state solutions. We propose a novel framework that integrates these two concepts to enhance the training of surrogate models. We apply this framework to training Physics-Enhanced Deep Surrogate (PEDs), a type of surrogate model that ensembles a neural network with an approximate physics solver to predict the target properties of physical systems governed by partial differential equations (PDEs). By employing a fixed-point iteration solver, our framework recursively feeds the output or target property back into the surrogate model. This feedback loop provides the neural network with continuous information on the dynamics of the physical system, enriching the models understanding without utilizing additional data. This thereby increases the models accuracy while being data efficient.

Xian Hadia
Georgia Tech
School of Computational Science and Engineering
xhadia3@gatech.edu

Raphaël Pestourie
MIT Mathematics
rpestourie3@gatech.edu

PP4

MS66: Tensor Train Based Sampling Algorithms for Approximating Regularized Wasserstein Proximal Operators

In this poster, we will introduce an algorithm based on tensor train approximations for sampling from a target distribution. We employ a semi-backward Euler discretization on the score dynamic and a kernel formula that comes from approximating the regularized Wasserstein operator. Moreover, we will utilize tensor train approximation to capture the evolution of high-dimensional probability density in Langevin dynamics. A rigorous analysis of the algorithm on Gaussian distributions will be presented. To show the effectiveness of our methods, we will provide numerical examples including sampling from various distributions and solving Bayesian inverse problems.

Fuqun Han
University of California, Los Angeles
fqhan@math.ucla.edu

Stanley J. Osher
UCLA
sjo@math.ucla.edu

Wuchen Li
University of South Carolina

wuchen@mailbox.sc.edu

PP4

MS69: Paired Autoencoders for Inference and Regularization

In this work, we describe a reduced approach that exploits technologies from machine learning (e.g., neural networks and auto-encoder networks) and dimensionality reduction models (e.g., low-rank and latent representations) to advance various technologies for inverse problems. We consider a decoupled approach for surrogate modeling, where unsupervised learning approaches are used to efficiently represent the input and target spaces separately, and a supervised learning approach is used to represent the mapping from one latent space to another. We demonstrate that our approach can outperform others in scenarios where training data for unsupervised learning is easily available, but the number of input/target pairs for supervised learning is small, and we introduce how the approach can be used for defining regularization or prior knowledge, and/or as a surrogate model for inversion, forward propagation, and adjoint free methods.

Emma Hart
Emory University
ehart5@emory.edu

Julianne Chung
Department of Mathematics
Emory University
julianne.mei-lynn.chung@emory.edu

Matthias Chung
Emory University, U.S.
matthias.chung@emory.edu

PP4

Application of Neural Operators to the Phase Field Modeling of Brittle Fractures

Brittle fracture is a common yet volatile phenomenon that occurs in hard substances. Its unpredictability poses challenges in numerous domains, including medicine, engineering, and environmental science. Typically, brittle fracture is modeled using the finite element method (FEM), which is extremely slow. Recent breakthroughs in machine learning have led to the use of convolutional neural networks (CNNs) to model brittle fracture. Although a CNN is faster than FEM or other conventional methods, its accuracy is limited by its focus on local relationships rather than the global relationships that determine fracture paths. In this work, we use a Fourier Neural Operator (FNO) to model brittle fracture. This method performs spectral convolutions in the frequency domain, thereby allowing the FNO to capture the global influence of all cracks and cavities. We trained the FNO on 2080 input-output image pairs of size 512 x 512, where the input image is a plate containing arbitrarily-shaped cavities, and the output image is the fracture path if the plate were to be loaded under tension. After training, the FNO predicted fracture path of a previously unseen input in 2 seconds. Compared to FEM, which required 25 minutes to solve the same problem, our FNO is 750 times faster. Also, we found that the FNO is one-third more accurate than a CNN in predicting the correct fracture path. This work shows the potential of FNOs in revolutionizing brittle fracture analysis across diverse fields.

Zhiheng He
Amador Valley High School
zhiheng.he@gmail.com

PP4

Protecting Data Privacy in Social Network Studies Using Synthetic Data

As network data becomes increasingly integral in research, synthetic data are essential aids in understanding social networks and preserving data confidentiality. Synthetic data allow for researchers to uncover patterns on the changes of networks while maintaining confidentiality. Much of synthetic data generation in categorical and mixed data settings is specific to healthcare contexts, where data privacy is of utmost importance. Synthetic data generators developed for healthcare data offer versatile solutions for preserving privacy in longitudinal social network studies. Confidentiality in social network data is important, especially when social networks involve sensitive topics, such as risky health behavior or illegal activities. Focusing on understanding networks over regularly tested intervals, we employ various methodologies, including approximating the joint probability distribution as a series of conditional distributions, GAN modeling approaches, and a combination of Bayesian networks with variational autoencoders. We will also use novel methods to compare the original to the synthetically generated data to assess quality of the generated data. With synthetic data, researchers can mimic real-world scenarios, enabling the exploration of social dynamics while ensuring the protection of individual privacy. This study underscores the transformative potential of synthetic data methodologies, emphasizing their role as ethical imperatives in social network studies.

Hortencia J. Hernandez, Amy Wagler
University of Texas at El Paso
hjhernandez4@miners.utep.edu, awagler2@utep.edu

PP4

MS62: Federated Learning with Differential Privacy and More

In the burgeoning field of machine learning, federated learning (FL) has emerged as a pivotal paradigm for training models directly on distributed local datasets without centralizing them, thus mitigating privacy leakage. However, challenges remain in ensuring that the machine learning process itself does not leak sensitive information. This talk explores the integration of differential privacy within FL to address these challenges. We will discuss the principles of differential privacy and its implementation techniques that help protect participants local datasets during the FL process. Further, we will introduce recent advancements in this approach, highlighting their effect on the trade-off between model accuracy and privacy loss.

Rui Hu
University of Nevada, Reno
ruihu@unr.edu

PP4

MS70: On the Rotational Invariance and Hyperbolicity of Shallow Water Moment Equations for Free Surface Flows

In this poster, we introduce our work on the incompressible

Navier-Stokes equations with free surfaces. We explore a novel class of models that describe this system, traditionally modeled using shallow water equations. We demonstrate the rotational invariance and hyperbolicity of these newly-derived shallow water moment models.

Juntao Huang
Texas Tech Univeresity
juntao.huang@ttu.edu

PP4

MS65: Towards Optimization of Biological Systems Leveraging PDE Model Discovery

Deep learning has transformed computational physics. Physics-enhanced deep surrogate models and physics-informed neural networks (PINNs) offer new solutions for complex physical problems. These models use machine learning to approximate physical systems, incorporating known physical laws to enhance accuracy and generalization. They can outperform conventional neural networks in accuracy, require less training data, and are faster than high-fidelity numerical solvers. Single-cell and spatial omics technologies are revolutionizing our understanding of biological systems. They allow for high-resolution analysis of individual cells and the spatial organization of tissues. These technologies illuminate the intricate dynamics of biological systems, from organ development to disease treatment. There are emerging synergies between these fields. Biological systems are complex and data-scarce but well-studied. The methodologies developed in computational physics are well-suited for such situations. The cross-pollination between these fields promises to advance our understanding of both machine learning and biological systems.

Xiangming Huang
Georgia Tech
School of Computational Science and Engineering
hxm19990719@gmail.com

PP4

MS68: Learning While Stuck: Reconstructing Energy-Based Models from Metastable States

In this work we rigorously show that energy based models can be learned efficiently from samples drawn from a class of metastable states of reversible Markov chain samplers. This holds even when these states are far away from the true model in terms of global metrics like KL divergence, energy, or other order parameters. We establish that metastable states that satisfy a strong metastability condition can approximate the true model well in single variable conditionals. This allows us to learn the true model using methods like Pseudo-Likelihood, even when the samples come from a metastable state concentrated in a small region of the state space. Explicit examples of such meta-stable states can be constructed from regions that effectively bottleneck the probability flow and cause poor mixing of the Markov chain. We show several numerical examples where the correct energy function is learned from samples drawn from a Markov chains metastable state, well before the chain mixes to the desired model. This gives an explicit example of an ML problem where learning succeeds despite having the "wrong' data but the right hypothesis. Our work also opens the possibility of learning or verifying energy based deep-learning models in cases where the true distribution is hard to sample from.

Abhijith Jayakumar, Andrey Lokhov, Sidhant Misra
Los Alamos National Laboratory
abhijithj@lanl.gov, lokhov@lanl.gov, sidhant@lanl.gov

Marc Vuffray
Theoretical Division and CNLS
Los Alamos National Laboratory
vuffray@lanl.gov

PP4

MS62: Gradient-Based Methods for Explanations and Confidence Metrics in Distributed Learning

Gradient-based methods provide a practical computational approach for obtaining explanations of AI models. Algorithms based on path integrals have been shown to satisfy commonsense axiomatic properties similar to those expected from Shapley values in collaborative game theory. In this talk, we present a new gradient-based method for explaining AI decisions using integrated decision gradients an adaptive approach to computing path integrals that leads to robust explanation. In particular, the approach resolves the saturation problem of classical integrated gradients by employing an importance factor to focus on the area of the path integral where the model makes its decision. This approach naturally leads to an efficient attribution-based confidence metric for AI models. Finally, we demonstrate how these approaches naturally extend to distributed learning models. The talk will build upon our recent results published at NeurIPS19, IJCAI22, AAAI24, and DAC24.

Sumit K. Jha
Florida International University
sumit.jha@fiu.edu

PP4

Distributed Gradient Descent for High Dimension Sparse Regression

We study linear regression from data distributed over a network of agents (with no master node) by means of LASSO estimation, in high-dimension, which allows the ambient dimension to grow faster than the sample size. While there is a vast literature of distributed algorithms applicable to the problem, statistical and computational guarantees of most of them remain unclear in high dimension. This paper provides a first statistical study of the Distributed Gradient Descent (DGD) in the Adapt-Then-Combine (ATC) form. Our theory shows that, under standard notions of restricted strong convexity and smoothness of the loss functions–which hold with high probability for standard data generation models–suitable conditions on the network connectivity and algorithm tuning, DGD-ATC converges globally at a linear rate to an estimate that is within the centralized statistical precision of the model. In the worst-case scenario, the total number of communications to statistical optimality grows logarithmically with the ambient dimension, which improves on the communication complexity of DGD in the Combine-Then-Adapt (CTA) form, scaling linearly with the dimension. This reveals that mixing gradient information among agents, as DGD-ATC does, is critical in high-dimensions to obtain favorable rate scalings.

Yao Ji
Purdue

ji151@purdue.edu

Gesualdo Scutari, Harsha Honnappa
Purdue University
gscutari@purdue.edu, honnappa@purdue.edu

PP4

MS59: Universal Sharpness Dynamics in Neural Network Training: Fixed Point Analysis, Edge of Stability, and Route to Chaos

In gradient descent dynamics of neural networks, the top eigenvalue of the Hessian of the loss (sharpness) displays a variety of robust phenomena throughout training. This includes early time regimes where the sharpness may decrease during early periods of training (sharpness reduction), and later time behavior such as progressive sharpening and edge of stability. We demonstrate that a simple 2-layer linear network (UV model) trained on a single training example exhibits all of the essential sharpness phenomenology observed in real-world scenarios. By analyzing the structure of dynamical fixed points in function space and the vector field of function updates, we uncover the underlying mechanisms behind these sharpness trends. Our analysis reveals (i) the mechanism behind early sharpness reduction and progressive sharpening, (ii) the required conditions for edge of stability, and (iii) a period-doubling route to chaos on the edge of stability manifold as learning rate is increased. Finally, we demonstrate that various predictions from this simplified model generalize to real-world scenarios and discuss its limitations.

Dayal Singh Kalra, Maissam Barkeshli, Tianyu He
University of Maryland, College Park
dayal@umd.edu, maissam@umd.edu, tianyuh@umd.edu

PP4

MS65: Multifidelity Linear Regression for Scientific Machine Learning from Scarce Data (Poster)

Machine learning (ML) methods have garnered significant interest as potential methods for learning surrogate models for complex engineering systems for which traditional simulation is expensive. However, in many scientific and engineering settings, training data are scarce due to the cost of generating data from traditional high-fidelity simulations. ML models trained on scarce data have high variance and are sensitive to vagaries of the training data set. We propose a new multifidelity training approach for scientific machine learning that exploits the scientific context where data of varying fidelities and costs are available; for example high-fidelity data may be generated by an expensive fully resolved physics simulation whereas lower-fidelity data may arise from a cheaper model based on simplifying assumptions. We use the multifidelity data to define new multifidelity Monte Carlo estimators for the unknown parameters of linear regression models, and provide theoretical analyses that guarantee accuracy and improved robustness to small training budgets. Numerical results show that multifidelity learned models achieve order-of-magnitude lower expected error than standard training approaches when high-fidelity data are scarce.

Elizabeth Qian
Georgia Institute of Technology
eqian@gatech.edu

Anirban Chaudhuri
MIT
anirbanc@oden.utexas.edu

Dayoung Kang
Georgia Tech
School of Aerospace Engineering
dkang339@gatech.edu

Vignesh Sella
University of Texas Austin
vsella@utexas.edu

PP4

Quantum-Enhanced Neural Network for Efficient Financial Market Prediction

In this study, we explore the application of quantum machine learning for predicting financial market trends, aiming to surpass the performance of traditional machine learning models in both accuracy and computational efficiency. Utilizing historical financial data, including stock prices and market indicators, we implement a Quantum Neural Network to develop predictive models on quantum hardware. These quantum models are benchmarked against their classical counterparts, such as Support Vector Machines and Recurrent Neural Networks, to assess their predictive accuracy and computational speed. Our findings reveal that quantum-enhanced machine learning models offer an improvement in processing speed, leveraging quantum parallelism and entanglement to achieve faster data processing and model training times. Moreover, the quantum models exhibit enhanced predictive accuracy, showcasing their potential to offer a quantum advantage in the analysis of complex financial datasets. This research not only highlights the advantages of applying quantum computing to financial market predictions but also opens new avenues for future investigations into the scalability of quantum machine learning algorithms and their applicability across various domains.

Yousaf A. Khaliq
Middle Tennessee State University
yak2a@mtmail.mtsu.edu

PP4

Behavior Based Entity Resolution for Big-Data Networks

Entity resolution techniques that rely on unreliable metadata, such as self-reported name, address, and phone number as features for entity-linking are vulnerable to manipulation. Blocking strategies facilitate computation but reduce the scope of the solution. Both blocking and self-reported data reduce the quality of the final entity linkage. Our approach combines entity behavior graphs with existing entity resolution techniques and distributed similarity computation using Apache Spark to deliver a process capable of rapid execution on large datasets without the need for blocking or reliance on self-reported data. We begin with a network where each unresolved entity is a node and each interaction between entities is an edge. Aggregating entity interactions provides a feature vector for each entity. Similarity is computed in parallel between feature vectors using Apache Spark. The result is used to create a new entity network with similarities as edges. Pruning this network by a similarity threshold and applying connected components produces independent subgraphs of like entities. The advantages of this method are that it is: unsupervised, massively scalable, not reliant on self-reported data, and not limited by blocking. This method has suc-

cessfully identified networks of fraud-actors. The compute time for a driver and 4 nodes each with 64 GB memory and 8 cores to calculate and store similarities was: 40M nodes, 1B edges, 15 minutes; 40M nodes, 100B edges, 18 hours

Byron F. Kilbourne, Jason Stajduhar
Elder Research Inc
byron.kilbourne@elderresearch.com, jason.stajduhar@elderresearch.com

PP4

MS64: Bayesian Optimization with Noise-Free Observations: Improved Regret Bounds Via Random Exploration

In this talk, we study Bayesian optimization with noise-free observations. We introduce new algorithms rooted in scattered data approximation that rely on a random exploration step to enhance the accuracy of surrogate models for the objective function. Our algorithms retain the ease of implementation of the classical GP-UCB algorithm, but the additional random exploration step accelerates their convergence, nearly achieving the optimal simple regret bound established by Bull. Furthermore, the new algorithms outperform GP-UCB and other popular Bayesian optimization strategies in several examples.

Hwanwoo Kim
University of Chicago
hwanwoo.kim@duke.edu

PP4

Learned Large-Scale Robust Matrix Completion Via Deep Unfolding

Robust matrix completion (RMC) is a technique for recovering a low-rank matrix from a subset of its entries, where some of the entries are corrupted. In this study, we introduce a deep-learning-augmented approach to RMC, called Learned Robust Matrix Completion (LRMC). Our approach utilizes Deep Unfolding, which converts each iteration of the RMC algorithm into a Deep Neural Network (DNN) layer and leverages DNN training to optimize the performance of the algorithm. Through extensive empirical experiments on synthetic datasets and real-world applications, we demonstrate that LRMC outperforms state-of-the-art methods, suggesting it as an attractive choice for addressing RMC problems.

Chandra S. Kundu
University of Central Florida
UNIVERSITY OF CENTRAL FLORIDA
chandra.kundu@ucf.edu

HanQin Cai
University of Central Florida
hqcai@ucf.edu

PP4

MS63: Complexity Bounds for Operator Learning

Operator learning frameworks generalize neural networks and define a methodology for the data-driven approximation of operators, such as the solution operator of a PDE. How much data is necessary to learn operators in a purely data-driven manner? How many tunable weights and biases are needed? This poster summarizes recent work providing partial answers to these questions. Upper and lower bounds on the complexity of operator learning are discussed.

Samuel Lanthaler
California Institute of Technology
slanth@caltech.edu

PP4

Zero-Truncated Poisson Regression for Sparse Multiway Count Data Data Corrupted by False Zeros

We propose a novel statistical inference paradigm for zero-inflated multiway count data that dispenses with the need to distinguish between true and false zero counts. Our approach ignores all zero entries and applies zero-truncated Poisson regression on the positive counts. Inference is accomplished via tensor completion that imposes low-rank structure on the Poisson parameter space. Our main result shows that an N-way rank-R parametric tensor $\boldsymbol{\mathcal{M}} \in (0,\infty)^{I\times\cdots\times I}$ generating Poisson observations can be accurately estimated from approximately $IR^2\log_2^2(I)$ non-zero counts for a nonnegative canonical polyadic decomposition. Several numerical experiments are presented demonstrating that our zero-truncated paradigm is comparable to the ideal scenario where the locations of false zero counts are known a-priori.

Richard B. Lehoucq
Sandia National Laboratories
rblehou@sandia.gov

PP4

MS66: Sketchy Moment Matching: Toward Fast and Provable Data Selection for Finetuning

We revisit data selection in a modern context of finetuning from a fundamental perspective. Extending the classical wisdom of variance minimization in low dimensions to high-dimensional finetuning, our generalization analysis unveils the importance of additionally reducing bias induced by low-rank approximation. Inspired by the variance-bias tradeoff in high dimensions from the theory, we introduce Sketchy Moment Matching (SkMM), a scalable data selection scheme with two stages. (i) First, the bias is controlled using gradient sketching that explores the finetuning parameter space for an informative low-dimensional subspace $\mathcal{S}$; (ii) then the variance is reduced over $\mathcal{S}$ via moment matching between the original and selected datasets. Theoretically, we show that gradient sketching is fast and provably accurate: selecting n samples by reducing variance over $\mathcal{S}$ preserves the fast-rate generalization $O(\dim(\mathcal{S})/n)$, independent of the parameter dimension. Empirically, we concretize the variance-bias balance via synthetic experiments and demonstrate the effectiveness of SkMM for finetuning in real vision tasks.

Qi Lei, Yijun Dong
New York University
ql518@nyu.edu, yd1319@nyu.edu

Hoang Phan, Xiang Pan
New York University, U.S.
hvp2011@nyu.edu, xiangpan@nyu.edu

PP4

MS63: Any-Dimensional Equivariant Neural Net-

works

Traditional supervised learning aims to learn an unknown mapping by fitting a function to a set of input-output pairs with a fixed dimension. The fitted function is then defined on inputs of the same dimension. However, in many settings, the unknown mapping takes inputs in any dimension; examples include graph parameters defined on graphs of any size and physics quantities defined on an arbitrary number of particles. We leverage a newly-discovered phenomenon in algebraic topology called representation stability to define equivariant neural networks that can be trained with data in a fixed dimension and then extended to accept inputs in any dimension. Our approach is black-box and user-friendly, requiring only the network architecture and the groups for equivariance, and can be combined with any training procedure.

Eitan Levin
Caltech
eitanl@caltech.edu

PP4

MS71: A Jax Toolbox for Learning Stochastic Differential Equations with Uncertainties from Real-World Data Streams

The development of data-informed predictive models for dynamical systems is of widespread interest in many disciplines. We present a unifying framework for blending mechanistic and machine-learning approaches to identify dynamical systems from noisily and partially observed data. This approach leverages Bayesian statistics and auto-differentiable data assimilation, and is instantiated in an open source Jax software package. We compare pure data-driven learning with hybrid models which incorporate imperfect domain knowledge. Our formulation is agnostic to the chosen machine learning model, is presented in both continuous and discrete-time settings, and is compatible both with model errors that exhibit substantial memory and errors that are memoryless. We will present formulations and experiments to examine data-driven point-wise and distributional estimates of differential equations in these settings.

Matthew Levine
Broad Institute
mattlevine22@gmail.com

PP4

Deep Operator Learning for Modeling Li-Ion Batteries

The widespread adoption of Li-ion batteries highlights their importance in modern energy storage systems. However, their complex multi-physics and multi-scale nature poses significant challenges for accurate modeling and computation. One of the primary challenges in Li-ion battery modeling is finding the balance between accuracy and efficiency. Detailed models encompassing all physics require high computational resources, while simplified models compromise accuracy. Deep operator learning offers a promising solution by integrating machine learning with scientific computing. In this poster, we demonstrate the efficacy of this approach with two examples. Firstly, we propose a physics-informed deep operator neural network (Phase-Field DeepONet) for fast simulations of pattern formation governed by gradient flows of free-energy functionals. This is achieved by employing the DeepONet as a time-stepper to approximate the dynamic response of a phase field and introducing the free energy into the training loss. Secondly, we develop a DeepONet-based computation framework for a homogenized model for all-solid-state batteries. This can accelerate the simulation by replacing the conventional time-consuming numerical methods for solving the mass transport equation and the reaction kinetics equation with DeepONets.

Wei Li
Northeastern University
we.li@northeastern.edu

Martin Z. Bazant
Departments of Chemical Engineering and Mathematics
Massachusetts Institute of Technology
bazant@mit.edu

Juner Zhu
Northeastern University
j.zhu@northeastern.edu

PP4

Automated Security Response Through Conjectural Online Learning under Information Asymmetry

Stochastic games arise in many complex socio-technical systems, particularly in security contexts, where the defender and the attacker interact under asymmetric information feedback. Existing computational methods for asymmetric information stochastic games (AISGs) are primarily offline and can not adapt to equilibrium deviations. The resulting defense strategies are inadequate when facing online nonstationary attacks. To address these limitations, we propose conjectural online learning (COL), an online method for generic AISGs. COL uses a forecaster-actor-critic (FAC) architecture, where subjective forecasts are used to conjecture the opponents' strategies within a lookahead horizon and Bayesian learning is used to calibrate the conjectures. To adapt strategies to nonstationary environments, COL uses online rollout with cost function approximation (actor-critic). We prove that the conjectures produced by COL are asymptotically consistent with the information feedback in the sense of a relaxed Bayesian consistency. We also prove that the empirical strategy profile induced by COL converges to the Berk-Nash equilibrium. We evaluate our method in a simulated IT infrastructure through an advanced persistent threat use case. COL produces effective security strategies adapting to a changing environment and enjoys faster convergence than current reinforcement learning techniques.

Tao Li
New York University
Department of Electrical and Computer Engineering
tl2636@nyu.edu

Kim Hammar, Rolf Stadler
Division of Network and Systems Engineering
KTH Royal Institute of Technology
kimham@kth.se, stadler@kth.se

Quanyan Zhu
New York University
Department of Electrical and Computer Engineering

qz494@nyu.edu

PP4

MS59: How Do Transformers Learn Topic Structure: Towards a Mechanistic Understanding

While the successes of transformers across many domains are indisputable, accurate understanding of the learning mechanics is still largely lacking. Their capabilities have been probed on benchmarks which include a variety of structured and reasoning tasksbut mathematical understanding is lagging substantially behind. Recent lines of work have begun studying representational aspects of this question: that is, the size/depth/complexity of attention-based networks to perform certain tasks. However, there is no guarantee the learning dynamics will converge to the constructions proposed. In our paper, we provide fine-grained mechanistic understanding of how transformers learn semantic structure, understood as capturing co-occurrence structure of words. Precisely, we show, through a combination of mathematical analysis and experiments on Wikipedia data and synthetic data modeled by Latent Dirichlet Allocation (LDA), that the embedding layer and the self-attention layer encode the topical structure. In the former case, this manifests as higher average inner product of embeddings between same-topic words. In the latter, it manifests as higher average pairwise attention between same-topic words. The mathematical results involve several assumptions to make the analysis tractable, which we verify on data, and might be of independent interest as well.

Andrej Risteski
Carnegie Mellon University
Machine Learning Department
aristesk@andrew.cmu.edu

Yuchen Li
Carnegie Mellon University
yuchenl4@cs.cmu.edu

Yuanzhi Li
CMU
yuanzhil@andrew.cmu.edu

PP4

MS70: Data-Assisted Algorithms for Inverse Random Source Problems

Inverse source scattering problems are essential in various fields, including antenna synthesis, medical imaging, and earthquake monitoring. In many applications, it is necessary to consider uncertainties in the model, and such problems are known as stochastic inverse problems. Traditional methods require a large number of realizations and information on medium coefficients to achieve accurate reconstruction for inverse random source problems. To address this issue, we propose a data-assisted approach that uses boundary measurement data to reconstruct the statistical properties of the random source with fewer realizations. We compare the performance of different data-driven algorithms under this framework to enhance the initial approximation obtained from integral equations. Our numerical experiments demonstrate that the data-assisted approach achieves better reconstruction with only 1/10 of the realizations required by traditional methods.Among the various Image-to-Image translation algorithms that we tested, the pix2pix method outperforms others in reconstructing well-separated inclusions with accurate positions. Our proposed approach results in stable reconstruction with respect to the observation data noise.

Ying Liang
Duke University
ying.liang@duke.edu

Peijun Li
Purdue University
lipeijun@purdue.edu

Yuliang Wang
Beijing Normal University
BNU-HKBU United International College
yuliangwang@uic.edu.cn

PP4

Maximizing Marginal Variance in Martingales for Unsupervised Learning

We present an unsupervised statistical learning approach that investigates all martingale couplings with one variable and one fixed marginal. In this framework, we propose that the variable marginal distribution, which maximizes variance, serves as a solution to the learning objective. We illustrate the applicability of this approach in a variety of unsupervised learning contexts, such as data clustering and principal curve and surface inference. We establish the existence of solutions to our optimization scheme and provide consistency result. Additionally, we show that a specific instance of our method is equivalent to classical principal component analysis (PCA), implying that our approach generalizes PCA.

Tongseok Lim
Purdue University
lim336@purdue.edu

Brendan Pass
U. Alberta
pass@ualberta.ca

Marcelo Souza
Purdue University
souzam@purdue.edu

PP4

MS77: Pairwise Alignment Improves Graph Domain Adaptation

Graph-based methods, pivotal for label inference over interconnected objects in many real-world applications, often encounter generalization challenges, if the graph used for model training differs significantly from the graph used for testing. This work delves into Graph Domain Adaptation (GDA) to address the unique complexities of distribution shifts over graph data, where interconnected data points experience shifts in features, labels, and in particular, connecting patterns. We propose a novel, theoretically principled method, Pairwise Alignment (Pair-Align) to counter graph structure shift by mitigating conditional structure shift (CSS) and label shift (LS). Pair-Align uses edge weights to recalibrate the influence among neighboring nodes to handle CSS and adjusts the classification loss with label weights to handle LS. Our method demonstrates superior performance in real-world applications, including node classification with region shift in social networks, and the pileup mitigation task in particle colliding experiments.

For the first application, we also curate the largest dataset by far for GDA studies. Our method shows strong performance in synthetic and other existing benchmark datasets.

Shikun Liu
Georgia Institute of Technology
shikun.liu@gatech.edu

Deyu Zou
University of Science and Technology of China
zoudeyu2020@mail.ustc.edu.cn

Han Zhao
University of Illinois Urbana Champaign
hanzhao@illinois.edu

Pan Li
Georgia Institute of Technology
panli@gatech.edu

PP4

MS61: Parametrized Wasserstein Hamiltonian Flow

In this work, we propose a numerical method to compute the Wasserstein Hamiltonian flow (WHF), which is a Hamiltonian system on the probability density manifold. Many well-known PDE systems can be reformulated as WHFs such as the Schrdinger equation. We use parameterized functions as push-forward maps to characterize the solution of WHF, and convert the PDE to a finite-dimensional ODE system, which is a Hamiltonian system in the phase space of the parameter manifold. We establish theoretical error bounds for the continuous time approximation in Wasserstein metric. For the numerical implementation, neural networks are used as push-forward maps. We design an effective symplectic scheme to solve the derived Hamiltonian ODE system so that the method preserves desirable quantities such as Hamiltonian. The computation is done by a fully deterministic symplectic integrator without any neural network training. The proposed algorithm is a sampling-based approach that scales well to higher dimensional problems. In addition, the method also provides an alternative connection between the Lagrangian and Eulerian perspectives of the original WHF through the parameterized ODE dynamics.

Shu Liu
University of California, Los Angeles, U.S.
shuliu@math.ucla.edu

Hao Wu
Wells Fargo
hwu406@gmail.com

Xiaojing Ye
Department of Mathematics & Statistics
Georgia State University
xye@gsu.edu

Haomin Zhou
School of Mathematics, Georgia Institute of Technology
GA, USA
hmzhou@gatech.edu

PP4

MS59: Non-Linear Activation Soothes NTK Conditioning for Wide Neural Networks: a Study in the Relu Case

Non-linear activation functions are well known to improve the expressivity of neural networks, which is the main reason of their wide implementation in neural networks. In this work, we showcase a new and interesting property of certain non-linear activations, focusing on the most popular example of its kind – Rectified Linear Unit (ReLU). By comparing the cases with and without this non-linear activation, we show that the ReLU has the following effects: (a) better data separation, i.e., a larger angle separation for similar data in the feature space of model gradient, and (b) better NTK conditioning, i.e., a smaller condition number of neural tangent kernel (NTK). Furthermore, we show that the ReLU network depth (i.e., with more ReLU activation operations) further magnifies these effects. Note that, without the non-linear activation, i.e., in a linear neural network, the data separation and NTK condition number always remain the same as in the case of a linear model, regardless of the network depth. Our results imply that ReLU activation, as well as the depth of ReLU network, helps improve the worst-case convergence rate of GD, which is closely related to the NTK condition number.

Chaoyue Liu, Like Hui
Halicioglu Data Science Institute
University of California, San Diego
chl212@ucsd.edu, lhui@ucsd.edu

Xiao Liu
N/A
xiaohydro@gmail.com

PP4

MS75: Assessing and Improving Accountability of Stochastic Neighbor Embedding Methods: A Map-Continuity Perspective

Visualizing high-dimensional data is an important routine for understanding biomedical data and interpreting deep learning models. Neighbor embedding methods such as t-SNE and UMAP are popular visualization methods, but recent studies suggest that they are likely to produce visual artifacts and lead to incorrect scientific conclusions. We find that, perhaps disturbingly, the embedding map displays discontinuity across many datasets. We identify two roots of observed discontinuity—namely overconfidence-inducing discontinuity and fracture-inducing discontinuity—that lead to overconfident clusters and spurious structures. Based on a statistical technique known as leave-one-out, we propose two metrics to remedy these issues: perturbation score and singularity score. Experiments on datasets across several domains show that the two metrics identify dubious embedding points, aid hyperparameter selection, and improve visualization diagnostics.

Zhexuan Liu
University of Wisconsin, Madison
zliu999@wisc.edu

Rong Ma
Harvard University
rongma@hsph.harvard.edu

Yiqiao Zhong
University of WisconsinMadison

yiqiao.zhong@wisc.edu

PP4

MS67: A Subspace Constrained Randomized Kaczmarz Method for Structure or External Knowledge Exploitation

We study a version of the randomized Kaczmarz algorithm for solving linear systems where the iterates are confined to the solution space of a selected subsystem. We show that the subspace constraint leads to an accelerated convergence rate, especially when the system has approximately low-rank structure. On Gaussian-like random data, we show that it results in a form of dimension reduction that effectively increases the aspect ratio of the system. Furthermore, this method serves as a building block for a second, quantile-based algorithm for solving linear systems with arbitrary sparse corruptions, which is able to efficiently utilize external knowledge about corruption-free equations and achieve convergence in difficult settings, such as in almost-square systems. Numerical experiments on synthetic and real-world data support our theoretical results and demonstrate the validity of the proposed methods for even more general data models than guaranteed by the theory.

Jackie Lok, Elizaveta Rebrova
Princeton University
jackie.lok@princeton.edu, elre@princeton.edu

PP4

Optimal Experimental Design for Control Problems Governed by PDE-Constrained Bayesian Inverse Problems

We propose a framework for Control-Oriented Optimal Experimental Design (OED) of linear PDE-constrained Bayesian inverse problems. In particular, we focus on optimal control problems that are parameterized by the solution of an inverse problem relying on experimental data gathered from sensors. While classical Bayesian OED techniques provide sensor placements that minimize the posterior uncertainty in the inversion parameter, these designs may not be ideal at minimizing uncertainty in the optimal control and the resulting final state. As experimental resources are often limited by cost or feasibility, one needs to prioritize the designs that reduce the uncertainty in the final goal. Hence, we propose design criteria and fast computational methods for finding sensor placements that minimize the end-goal uncertainty in optimal control. We present illustrative numerical experiments in the context of a convective heat transfer application, in which we reconstruct an ambient heat source term based on initial state measurements and, with the reconstructed dynamics, control a stationary source to attain a target state within a given period of time.

Madhusudan Madhavan
North Carolina State University
mmadhav@ncsu.edu

Alen Alexanderian
North Carolina State University, U.S.
alexanderian@ncsu.edu

Arvind Saibaba
North Carolina State University
asaibab@ncsu.edu

Bart G. Van Bloemen Waanders, Rebekah White
Sandia National Laboratories
bartv@sandia.gov, rebwhit@sandia.gov

PP4

Prediction of Natural Disaster Data Using Fractional Count Processes

Modelling wildfire events has been studied in the literature using the Poisson process, which essentially assumes the independence of wildfire events. In this talk, we use the fractional Poisson process to model the wildfire occurrences in California between June 2019 - April 2023 and predict the wildfire events that explains the underlying memory between these events. We introduce method of moments and maximum likelihood estimate approaches to estimate the parameters of the fractional Poisson process, which is an alternative to the method proposed by Cahoy (2010). We obtain the estimates of the fractional parameter as 0.8, proving that the wildfire events are dependent. The proposed model has reduced prediction error by 90% compared to the classical Poisson process model.

Sudeep R Bapat
Indian Institute of Technology Bombay
sudeepb@iitb.ac.in

Aditya Maheshwari
Indian Institute of Management Indore
adityam@iimidr.ac.in

PP4

MS77: Principle-Guided Graph Foundation Model Design

Graph Foundation Model outperforms vanilla GNNs which benefits from larger training scale with positive transferability. Nonetheless, the principle for how to achieve such positive transfer remains unclear. In this poster, I will present a graph vocabulary perspective to find the basic units in the graph domain, transferable across different tasks and datasets. Such vocabulary design should be guided by existing principles from network analysis, graph expressiveness, and network stability perspectives.

Haitao Mao
Michigan State University
Michigan State University
haitaoma@msu.edu

PP4

MS68: Classifying Imbalanced Data

Imbalanced describes any dataset in which there is one class of data represented by significantly less samples than another called the minority class. A common type of data found in many fields including toxicology/drug discovery, medical diagnostics, fraud detection, and nuclear physics, to name a few, imbalanced data poses problems of bias to machine learning models. Traditional machine learning models were built with the assumption of balanced data; therefore, they learn in favor of the majority class well represented in the dataset. Common approaches rebalance the data or rebalance the model. Of the former category, we present a novel algorithm, MUBO: Majority Undersampling with Bilevel Optimization, that leverages a bilevel optimization formulation to optimize training data and overcome model bias and, ultimately, improve the classi-

fication of imbalanced data. Experiments with benchmark imbalanced datasets show that MUBO results in F1 score increases of up to 11% over the classic SMOTE method and up to 8% over the state-of-the-art SMOTified-GAN method.

Karen Medlin
University of North Carolina at Chapel Hill
kmedlin@unc.edu

Sven Leyffer
Argonne National Laboratory
leyffer@anl.gov

M. Gregory Forest
University of North Carolina at Chapel Hill
Departments of Mathematics & Biomedical Engineering
forest@unc.edu

Krishnan Raghavan
Mathematics and Computer Science
Argonne National Laboratory
kraghavan@anl.gov

PP4

MS76: On Independent Samples Along the Langevin Diffusion and Algorithm

Markov chains are widely used to sample from probability distributions, and their mixing time tells us how long we need to wait before the law of the iterates is close to the stationary distribution. This poster discusses the "independence time" of a Markov chain, which tells us how long we need to wait before the iterates have small mutual information with the initial random variable. It shows the independence time for the Langevin diffusion along with discrete time implementations of the diffusion, and shows that for strongly-log concave targets, the mutual information goes to 0 exponentially fast. These convergence rates are tight and the results are proven using strong data processing inequalities and regularity properties of these Markov chains. It's based on joint work with Jiaming Liang and Andre Wibisono.

Siddharth Mitra
Yale University
siddharth.mitra@yale.edu

PP4

MS77: Graphmad: Graph Mixup for Data Augmentation Using Data-Driven Convex Clustering

We develop a novel data-driven nonlinear mixup mechanism for graph data augmentation and present different mixup functions for sample pairs and their labels. Mixup is a data augmentation method to create new training data by linearly interpolating between pairs of data samples and their labels. Mixup of graph data is challenging since the interpolation between graphs of potentially different sizes is an ill-posed operation. Hence, a promising approach for graph mixup is to first project the graphs onto a common latent feature space and then explore linear and nonlinear mixup strategies in this latent space. In this context, we propose to (i) project graphs onto the latent space of continuous random graph models known as graphons, (ii) leverage convex clustering in this latent space to generate nonlinear data-driven mixup functions, and (iii) investigate the use of different mixup functions for labels and data samples. We evaluate our graph data augmentation performance on benchmark datasets and demonstrate that nonlinear data-driven mixup functions can significantly improve graph classification.

Madeline Navarro, Santiago Segarra
Rice University
nav@rice.edu, segarra@rice.edu

PP4

Towards Tuning-Free Minimum-Volume Nonnegative Matrix Factorization

Nonnegative Matrix Factorization (NMF) has emerged as a versatile and powerful tool in diverse domains, including machine learning, data mining, and signal processing. Among various NMF approaches, minimum-volume NMF stands out as a prominent method to extract indentifiable solutions. In 2019, Leplat et al. proposed a version of minimum-volume NMF that can handle the rank-deficiency case. However, this work did not address an important issue of selecting an appropriate λ parameter to adapt to varying noise levels, resulting in suboptimal performance. Furthermore, the attained errors failed to achieve machine precision, which limits the method's practical applicability. To overcome these limitations, our research aims to investigate the relationship between λ and the noise level through a series of well-designed experiments. Additionally, we present a novel Majorization-Minimization (MM) variant for minimum-volume NMF. By employing the MM philosophy, we optimize a relaxed surrogate function, enhancing the optimization process and yielding improved results compared to the traditional hard function. Our experimental findings demonstrate the efficacy of the proposed method, showcasing errors that consistently fall below machine precision.

Duc Toan Nguyen
Texas Christian University
duc.toan.nguyen@tcu.edu

Eric Chi
Rice University
echi@rice.edu

PP4

MS61: Wasserstein Convergence Guarantees for a General Class of Score-Based Generative Models

Score-based generative models (SGMs) is a recent class of deep generative models with state-of-the-art performance in many applications. In this presentation, we establish convergence guarantees for a general class of SGMs in 2-Wasserstein distance, assuming accurate score estimates and smooth log-concave data distribution. We specialize our result to several concrete SGMs with specific choices of forward processes modeled by stochastic differential equations, and obtain an upper bound on the iteration complexity for each model, which demonstrates the impacts of different choices of the forward processes. We also provide a lower bound when the data distribution is Gaussian. Numerically, we experiment SGMs with different forward processes, some of which are our newly proposed methods, for unconditional image generation on CIFAR-10. We find that the experimental results are in good agreement with our theoretical predictions on the iteration complexity, and the models with our newly proposed forward processes can outperform existing models.

Hoang M. Nguyen

Department of Mathematics
Florida State University
hmn19d@fsu.edu

PP4

An Optimal Multivariate Functional Classifier

The task of classification using multiple functional covariates has become increasingly relevant over the past years with the growing sizes of data and technological advancements allowing to collect data over different domains, from one-dimensional signals to two-dimensional images. Despite the growing importance of this task, there are not many approaches available to make use of functional variables from different domains to perform this classification task. We therefore contribute to this line of research by (i) extending the functional neural network (FNN) approach (initially developed for regression problems) to the classification task; (ii) proving the minimax optimality of the proposed approach; (iii) investigating the behavior of the approach in presence of functional variables from different domains (i.e. a combination of one- and two-dimensional functional variables); and (iv) highlighting the good comparative performance of this approach through simulation and real-world data studies. The proposed method, that we call Multivariate Functional Classifier (MFC), complements the few existing methods for the classification of multivariate functional data and does not require the functional Gaussian assumption.

Ukamaka Nnyaba, Nedret Billor
Auburn University
uvn0001@auburn.edu, billone@auburn.edu

Roberto Molinari
Auburn University
Department of Mathematics and Statistics
robmolinari@auburn.edu

PP4

MS69: Dynamic Inverse Problems: Efficient Methods for Dynamic Image Reconstruction with Motion Estimation

Large-scale dynamic inverse problems are typically ill-posed and suffer from complexity of the model constraints and large dimensionality of the parameters. A common approach to overcome ill-posedness is through regularization that aims to add constraints on the desired parameters, typically on space by sparsity of the coefficients or the gradient and with constraints on the temporal dimension. In this work, we propose a new method that incorporates a model for the temporal dimension by estimating the motion of the objects alongside solving the regularized problems. In particular, we consider the optical flow model as part of the regularization that simultaneously estimates the motion and provides an approximation for the desired image. To overcome the large dimensionality of the parameters, we combine our approach with a generalized Krylov subspace method for computational efficiency. To illustrate the effectiveness of the prescribed approach, we present numerical experiments from computerized tomography and image deblurring applications.

Toluwani Okunola
TU Berlin
toluwaokunola@gmail.com

PP4

Covering Numbers of Relu Networks

Covering numbers of ReLU network families are important as they allow to quantify the fundamental limits in function approximation and generalization in the context of non-parametric regression. Previous studies have reported upper bounds on covering numbers for specific network architectures and covering ball radii, which in turn allow to lower-bound the function approximation error and upper-bound the generalization error in regression problems. The results presented in this poster complete this picture by deriving comprehensive upper bounds in general settings, i.e., for various types of network architectures and all covering ball radii. More importantly, we also establish lower bounds that are tight w.r.t. the upper bounds we obtain. The resulting complete characterization of covering numbers allows us to characterize the behavior of network compression, network quantization, network approximation, and non-parametric regression with ReLU networks. In particular, we show in a constructive manner that fully connected ReLU networks with base-2 quantized weights and independent choices of width and depth, can approximate 1-Lipschitz functions in a memory-optimal manner. We also establish that the generalization error in non-parametric regression using ReLU networks achieves the optimal rate of convergence.

Weigutian Ou
Chair for Mathematical Information Science
ETH Zürich
wou@ethz.ch

Philipp Schenkel
FZI Research Center for Information Technology
schenkel@fzi.de

Helmut Bölcskei
ETH Zurich
hboelcskei@ethz.ch

PP4

Thoracic Aortic Geometry in Scale Space for Enhanced Clinical Decision-Making

The scale and resolution of anatomy extracted from medical CT images is paramount in the study and advancement of clinical decision-making tools. Furthermore, the quality and signal that can be computed from an image goes beyond trivial 1D size metrics and can introduce compounding error in higher dimensional features such as anatomical shape. Recent computational methods have been presented which quantify the shape of the aorta and successfully use a tandem size-shape feature space to increase the accuracy in classifying the impending success of surgery from pre-operative imaging data. The most promising of the metrics is the fluctuation in total integrated Gaussian curvature (dK) over a surface mesh model of the aortic external surface. However, a significant tradeoff has been identified between noise reduction and shape signal preservation within the scale space parameters; namely smoothing intensity, meshing density, and the partitioning performed in the computation of (dK). The presented is an analysis into identifying the regions of high signal strength within a 300+ aortic dissection patient dataset implying a critical resolution scale at which quantified shape variations are linked to aortic intervention outcomes. Additionally, these

results allow the benefiting clinical prediction models to contain inherent confidence intervals sourced from the image properties rather than the statistics of a potentially biased or skewed patient training population.

Joseph Pugar
Postdoc
jpugar@uchicago.edu

Luka Pocivavsek, Julius Kim, Kelly Twohig, David Jiang
University of Chicago
lpocivavsek@bsd.uchicago.edu,
juliuskim1012@uchicago.edu,
kelly.twohig@uchicagomedicine.org,
david.jiang@uchospitals.edu

PP4

A Variant Equilibrium Optimizer Approach for Dimensionality Reduction in Cancer Microarray Data

In cancer research, identifying relevant gene subsets from high-dimensional microarray data is crucial for accurate classification and understanding of underlying biological mechanisms. This work proposes a novel variant of the Equilibrium Optimizer (EO) algorithm tailored for gene subset selection in cancer microarray data analysis. Our approach integrates a rank-based mechanism for parameter updates and incorporates an adaptive exploration factor, enhancing the algorithm's exploration-exploitation balance. Furthermore, we strategically deploy the logistic map to augment the performance of the Enhanced Local Escaping Operator (eLEO) within the original EO framework. The efficacy of our proposed algorithm is evaluated by conducting experiments on standardized test functions from the IEEE Congress on Evolutionary Computation (CEC) 2021 competition, which serves as a benchmark for assessing optimization algorithms. The proposed algorithm is also applied to real-world cancer microarray datasets to demonstrate its practical utility in improving classification accuracy by selecting optimal gene subsets and reducing dimensionality. The experimental results showcase the competitiveness of the proposed variant EO against cutting-edge metaheuristics, affirming its efficacy in tackling high-dimensional optimization problems and its potential for advancing cancer research through improved data analysis techniques.

Mohammed Qaraad, Mriganka Shekha Chaki, David Guinovart
University of Minnesota
The Hormel Institute
qaraa001@umn.edu, chaki014@umn.edu,
guino001@umn.edu

PP4

Explosive Polarization-Depolarization in Opinion Dynamics with Nonlinear Incidence

Polarization significantly influences societal divisions across economic, political, religious, or ideological lines. Understanding these mechanisms is key to devising strategies to mitigate such divisions and promote depolarization. Our study examines how nonlinear incidence in opinion dynamics affects polarization within structured communities. The model:

$$\dot{x}_i = -\gamma x_i + \beta(1 - x_i)\sum_{k=1}^{N} A_{ik}x_k(1 + \alpha x_k^{p-1}), \qquad (1)$$

where x_i measures individual opinion intensity, γ the resilience rate, β the change rate, α quantifies the nonlinear incidence influence, with p as the exponent, and A representing the adjacency matrix. Setting $\alpha = 0$ initially simplifies to a linear model, but at $\alpha = 1$, the system undergoes a first-order phase transition , i.e., explosive polarization, respectively depolarization,the latter resilient to γ changes. We developed a degree-based model, $\langle\dot{x}\rangle = (\tilde{\beta} - \gamma)\langle x\rangle + \tilde{\beta}(\alpha - 1)\langle x\rangle^2 - \alpha\tilde{\beta}\langle x\rangle^3$, representing average opinion. Continuity at critical points is crucial for predicting polarization patterns, resembling the critical eigenvector in continuous transitions.

Haoyang Qian
Department of Mathematics
Florida State University
hq22@fsu.edu

PP4

Learning Effective Memory Kernels For Coarse-Grained Simulations

Accurately capturing the dynamical behavior of real physical systems usually require Non-Markovian models to account for memory effects. The Generalized Langevin Equation (GLE) provides a robust framework for building such models. However, the main bottleneck in using the GLE is determining the form and parameters of the memory kernel. Here we propose a new approach to circumvent this by using a Bayesian framework to extract these parameters from experimental or simulation data. To introduce our method, we first focus on a simple case of a one-dimensional free particle with an exponentially decaying memory kernel. Synthetic data is generated through numerical integration of the GLE by mapping the system onto an equivalent, higher dimensional Markovian system. Using the numerical update rule and by eliminating all unphysical variables introduced in the mapping, we derive an expression for the likelihood. With a suitable prior, we show that we can successfully sample the decay rate of the corresponding memory kernel for the synthetic data. To establish the practical applicability of our method, we also generalize our likelihood derivation for the case of a coarse-grained model of a protein and demonstrate its efficacy in predicting the memory kernel decay rates of different residues in the protein.

Nikhil Ramesh, Pedro Pessoa, Banu Ozkan, Steve Presse
Arizona State University
nrames19@asu.edu, ppessoa@asu.edu,
banu.ozkan@asu.edu, spresse@asu.edu

PP4

MS63: A Spectral Analysis of Graph Neural Networks on Dense and Sparse Graphs

In this work we propose a random graph model that can produce graphs at different levels of sparsity. We analyze how sparsity affects the graph spectra, and thus the performance of graph neural networks (GNNs) in node classification on dense and sparse graphs. We compare GNNs with spectral methods known to provide consistent estimators for community detection on dense graphs, a closely related task. We show that GNNs can outperform spectral methods on sparse graphs, and illustrate these results with numerical examples on both synthetic and real graphs.

Luana Ruiz
Department of Applied Mathematics and Statistics

Johns Hopkins University
lrubini1@jh.edu

PP4

Reducing the Cost of Posterior Sampling in Linear Inverse Problems via Task-dependent Score Learning

Score-based diffusion models (SDMs) offer a powerful way to sample from the posterior distribution in various Bayesian inverse problems. Traditional methods rely on multiple evaluations of the forward mapping to generate each posterior sample, which can be computationally demanding. Our research focuses on linear inverse problems, particularly in fields like medical imaging, where the forward mapping is expensive and frequent posterior sampling is necessary. We introduce a novel approach that eliminates the need for forward mapping evaluations during posterior sampling. By shifting the computational effort to an offline task, we train the score of a specific diffusion-like random process. This training is task-dependent, requiring knowledge of the forward mapping but not the measurement data. The conditional score for the posterior is then derived from the auxiliary score through affine transformations, ensuring no introduction of error. Our findings generalize to infinite-dimensional diffusion models and are supported by rigorous numerical analysis and experiments. This method can significantly reduce computational costs, making it highly beneficial for applications demanding frequent and efficient posterior sampling.

Fabian Schneider
LUT University
Lappeenranta, Finland
fabian.schneider@lut.fi

Duc-Lam Duong
LUT University
Duc-Lam.Duong@lut.fi

Matti Lassas
University of Helsinki
matti.lassas@helsinki.fi

Maarten de Hoop
Rice University
mvd2@rice.edu

Tapio Helin
LUT University
Tapio.Helin@lut.fi

PP4

MS62: Dealing Doubt: Unveiling Threat Models in Gradient Inversion Attacks under Federated Learning

Federated Learning (FL) has emerged as a leading paradigm for decentralized, privacy preserving machine learning training. However, recent research on gradient inversion attacks (GIAs) have shown that gradient updates in FL can leak information on private training samples. While existing surveys on GIAs have focused on the honest-but-curious server threat model, there is a dearth of research categorizing attacks under the realistic and far more privacy-infringing cases of malicious servers and clients. Here we present a survey and novel taxonomy of GIAs that emphasize FL threat models, particularly that of malicious servers and clients. We first formally define GIAs and contrast conventional attacks with the malicious attacker. We then summarize existing honest-but-curious attack strategies, corresponding defenses, and evaluation metrics. Critically, we dive into attacks with malicious servers and clients to highlight how they break existing FL defenses, focusing specifically on reconstruction methods, target model architectures, target data, and evaluation metrics. Lastly, we discuss open problems and future research directions.

Yichuan Shi
Massachusetts Institute of Technology
yshi23@mit.edu

Olivera Kotevska, Viktor Reshniak
Oak Ridge National Laboratory
kotevskao@ornl.gov, reshniakv@ornl.gov

Abhishek Singh
Massachusetts Institute of Technology
abhi24@media.mit.edu

PP4

A Chaotic Niching Sparrow Search Algorithm for Solving Clustering Optimization Problems

Sparrow Search Algorithm (SSA) is one of the most recent optimization algorithms which is known for its good optimal ability along with fast convergence, Although, the SSA has a lot of merits, it is still facing numerous drawbacks namely falling into the local optima, steady convergence, etc. Therefore, a chaotic version of the Sparrow Search Algorithm (SSA) using the niching technique is proposed with different chaotic mechanisms such as bernoulli map, logistic map, chebyshev map, circle map, cubic map, iterative chaotic map, piecewise map, signer map, sinusoidal map, and tent map to address the previously mentioned drawbacks of SSA. Our proposed algorithm is named a Chaotic Niching Sparrow Search Algorithm (CNSSA) where the niching technique is used for updating the position of followers and scouters in SSA. The suggested CNSSA 's performance is assessed using a few benchmark test functions that were obtained from the 2014 Congress on Evolutionary Computation (CEC14). The outcomes showed that the CNSSA solves benchmark test functions more effectively and with better results. Furthermore, clustering problems are also solved using the CNSSA. When compared to other algorithms namely the original SSA, GWO, PSO, TLBO, JAYA, ACO, BA, and GSA then it is discovered that the proposed CNSSA produces superior clustering results.

Gagandeep Kaur Sidhu, Jatinder Kaur
Chandigarh University, Gharuan, Punjab
Gagan.kw777@gmail.com, jatindere1486@cumail.in

PP4

Difussion over Ramified Domains: Solvability and Fine Regularity

We consider a domain $\Omega \subseteq \mathbb{R}^2$ with branched fractal boundary Γ^∞ and parameter $\tau \in [1/2, \tau^*]$ introduced by Achdou and Tchou, for $\tau^* \simeq 0.593465$, which acts as an idealization of the bronchial trees in the lungs systems. For each $\tau \in [1/2, \tau^*]$, the corresponding region Ω is a non-Lipschitz domain, which attains its roughest structure at the critical value $\tau = \tau^*$ in such way that in this endpoint parameter the region Ω fails to be an extension domain, and its ramified boundary Γ^∞ is not post-critically finite. Then, we investigate a model equation related to the diffu-

sion of oxygen through the bronchial trees by considering the realization of a generalized diffusion equation

$$\frac{\partial u}{\partial t} - \mathcal{A}u + \mathcal{B}u = f(t,x) \quad \text{in } (0,\infty)\times\Omega$$

with inhomogeneous mixed-type boundary conditions

$$\frac{\partial u}{\partial \nu_{\mathcal{A}}} + \beta u = g(x,t) \quad \text{on } (0,\infty)\times\Gamma^{\infty}, \qquad u = 0 \quad \text{in } (0,\infty)\times(\partial\Omega\setminus\Gamma^{\infty})$$

and $u(x,0) = u_0 \in C(\overline{\Omega})$. These domains are projections of bronchial trees in 2D. Calculating their measure involves an optimization problem that requires computational calculations and geometric concepts.

Kevin D. Silva Perez
University of Puerto Rico at Rio Piedras
kevin.silva1@upr.edu

PP4

MS75: A Riemannian Distance Geometry Algorithm with Adaptive Sampling

The Euclidean distance geometry (EDG) problem is an important computational task that appears in many applications, ranging from molecular conformation in computational chemistry to dimensionality reduction in machine learning and sensor network localization. Using the partial pairwise distance information of a point set as an input, EDG aims to obtain the configuration of the point set. When partial distances are all that are available, previous work shows that a Riemannian matrix completion approach adapted to the specifics of the EDG problem provides local convergence guarantees along with strong numerical results. However, this approach relies on uniform sampling of the distance matrix, which may not always be practical in certain applications. In this work, we seek to extend this method beyond the uniform sampling models and provide adaptive sampling schemes that improve the robustness of this algorithm.

Chandler Smith, Abiy F. Tasissa
Tufts University
chandler.smith@tufts.edu, Abiy.Tasissa@tufts.edu

PP4

MS71: Discovering Nonlinear Oscillatory and Chaotic Systems from Partial Observations

Complex multi-component systems from cells and tissues to biochemical reactors often exhibit oscillatory and chaotic nonlinear dynamics that are essential to their signaling properties and functions. Despite the rapid advancement of sensor and imaging technology, many physical and biological systems can only be partially observed with practitioners in need of model-fitting tools that can account for this missing information. Here we develop an automated inference method that discovers predictive differential equation models from a few noisy partial observations of a systems state. We illustrate our method on a combination of both simulation and experimental data from a variety of physical, chemical and biological systems showing that in many cases noisy partial observations are sufficient to infer predictive multivariate dynamical systems.

George Stepaniants
MIT Mathematics & IDSS
gstepan@caltech.edu

PP4

Minimum-Risk Recalibration of Classifiers

Recalibrating probabilistic classifiers is vital for enhancing the reliability and accuracy of predictive models. Despite the development of numerous recalibration algorithms, there is still a lack of a comprehensive theory that integrates calibration and sharpness (which is essential for maintaining predictive power). In this paper, we introduce the concept of minimum-risk recalibration within the framework of mean-squared-error (MSE) decomposition, offering a principled approach for evaluating and recalibrating probabilistic classifiers. Using this framework, we analyze the uniform-mass binning (UMB) recalibration method and establish a finite-sample risk upper bound of order $\tilde{O}(B/n + 1/B^2)$ where B is the number of bins and n is the sample size. By balancing calibration and sharpness, we further determine that the optimal number of bins for UMB scales with $n^{1/3}$, resulting in a risk bound of approximately $O(n^{-2/3})$. Additionally, we tackle the challenge of label shift by proposing a two-stage approach that adjusts the recalibration function using limited labeled data from the target domain. Our results show that transferring a calibrated classifier requires significantly fewer target samples compared to recalibrating from scratch. We validate our theoretical findings through numerical simulations, which confirm the tightness of the proposed bounds, the optimal number of bins, and the effectiveness of label shift adaptation.

Zeyu Sun, Dogyoon Song
Electrical and Computer Engineering
University of Michigan
zeyusun@umich.edu, dogyoons@umich.edu

Alfred O. Hero
The University of Michigan
Dept of Elect Eng/Comp Science
hero@eecs.umich.edu

PP4

MS67: Kaczmarz Based Iterative Hard Thresholding Techniques for Low-Rank Tensor Recovery

Hybrid methods that leverage the Kaczmarz and Iterative Hard Thresholding (IHT) algorithms have been shown to be particularly effective in tackling sparse vector recovery problems. In this poster, I will discuss our recent work, in which we developed Kaczmarz based IHT techniques to recover low-rank tensors from a few linear measurements. I will also present some theoretical convergence guarantees followed by empirical results that highlight the effectiveness of these methods against different classes of measurement operators. Associated poster with minsymposium on Structure in Data: Theory, Learning, and Algorithms

Shambhavi Suryanarayanan
Princeton University
shambhavi.suryanarayanan@gmail.com

PP4

MS64: Lookahead-Free Bayesian Optimization Via Gittins Indices

Bayesian optimization is a technique for efficiently optimizing unknown functions in a black-box manner. To

ensure their decisions take into account more than just the short-term potential benefit of the current query, it is desirable for Bayesian optimization policies to be non-myopic. However, current non-myopic approaches rely on lookahead techniques, which are complex and computationally intensive. We introduce a new lookahead-free policy for Bayesian optimization, using an unexplored connection with the Pandora's box problem, a well-studied problem in the economics literature. The Pandora's box problem admits a Bayesian-optimal solution based on an acquisition function called the Gittins index. We demonstrate empirically that the same acquisition function performs well for Bayesian optimization, particularly in higher dimensions. The acquisition function is computationally efficient, and it naturally handles problems with heterogeneous query costs. Our work constitutes a first step towards integrating ideas from Gittins index theory into policies for Bayesian optimization.

Alexander Terenin
Cornell University
avt28@cornell.edu

PP4

Exploring Complex Systems with Markov Bridge and RWR Models

Understanding the structural and functional complexities of networked systems is pivotal for advancing both theoretical and applied network science. These complex systems, whether they arise in nature, society, or engineered settings, often consist of numerous interacting components that follow nonlinear dynamics. This work employs stochastic processes, specifically Markov Bridge processes and Random Walk with Restart (RWR) models, to explore and reconstruct the topology of networked systems. Our research focuses on real-world transportation networks and offers theoretical insights and practical methodologies. Our first application utilizes GPS data from San Francisco taxis to construct a Markov Bridge model, enabling us to examine the stochastic nature of route choices influenced by vehicle occupancy. We introduce a novel entropy measure based on the Markov Bridge, demonstrating enhanced sensitivity to changes in network topology compared to traditional measures such as spectral entropy. In our second application, we implement an RWR model to reconstruct the topology of multi-city, multi-layer multiplex transportation networks. This approach allows us to uncover and quantify the varied topological and dynamic characteristics inherent in these complex network systems. Through these applications, our work contributes to the broader understanding of network dynamics, offering valuable insights and tools for the mathematical modeling of data-rich, interconnected systems.

Chenxiao Tian
Princeton University
Princeton University
ct3471@princeton.edu

Jürgen Hackl
Princeton University
hackl@princeton.edu

PP4

MS62: Privacy for Activation Release in Collaborative Training and Inference

There has been a recent advent of generative AI methods powered by transformers and large foundation models. This has been overshadowed by serious concerns about privacy-sensitivities of the raw datasets during training, fine-tuning and prompt tuning while maintaining computational efficiency in client-server settings. This necessitates new mechanisms of training with formal privacy guarantees for releasing intermediate activations (embeddings), by clients that hold the sensitive data to a computationally more powerful server. The privacy concern for NLP modalities, have arisen due to lack of privacy guarantees at the token level, word embedding level and sentence embedding level. Current differential privacy methods are applied upon extracting the pre-trained embeddings from a standard NLP architecture, while treating these embeddings as tabular data. As differential privacy is not immune to pre-processing, the privacy guarantees applied at this attack surface do not translate to privacy guarantees at the token, word or sentence level. Moreover, preventing model identifiability in this setting of inference serving systems where a server holds a model zoo of multiple models via latency and accuracy fingerprinting attacks is yet another important adjacent problem. This talk provides formal methods to cater to such relevant real-world problems in collaborative training and inference.

Praneeth Vepakomma
MIT
MBZUAI
vepakom@mit.edu

PP4

MS61: Efficient Computation of Mean-Field Control Based Barycenters from Reaction-Diffusion Systems

We develop a class of barycenter problems based onmean field control problems in three dimensions with associated reactive-diffusion systems of unnormalized multi-species densities. This problem is the generalization of the Wasserstein barycenter problem for single probability density functions. The primary objective is to present a comprehensive framework for efficiently computing the proposed variational problem: generalized Benamou-Brenier formulas with multiple input density vectors as boundary conditions. Our approach involves the utilization of high-order finite element discretizations of the spacetime domain to achieve improved accuracy. The discrete optimization problem is then solved using the primal-dual hybrid gradient (PDHG) algorithm, a first-order optimization method for effectively addressing a wide range of constrained optimization problems. The efficacy and robustness of our proposed framework are illustrated through several numerical examples in three dimensions, such as the computation of the barycenter of multi-density systems consisting of Gaussian distributions and reactive-diffusive multi-density systems involving 3D voxel densities. Additional examples highlighting computations on 2D embedded surfaces are also provided.

Arjun Vijaywargiya, Guosheng Fu
University of Notre Dame
arjunveejay@nd.edu, gfu@nd.edu

Stanley J. Osher
UCLA
sjo@math.ucla.edu

Wuchen Li
University of South Carolina

wuchen@mailbox.sc.edu

PP4

Machine Learning-Assisted Agent-Based Model Predicts Outcomes of Cancer Immunotherapy and Highlights Spatial Complexity of the Tumor Microenvironment

Immune checkpoint inhibitors (ICIs) have changed the landscape of molecular therapeutics for cancers. However, ICIs do not work equally well for all patients. There has been a growing interest in using computational models to optimize clinical responses. Ordinary differential equations (ODEs) have been widely used for mechanistic modeling in immunotherapy because they allow rapid simulations of temporal changes in the cellular and molecular populations involved. Agent-based models (ABMs) have gained popularity because they can model more detailed spatial heterogeneity that better reflect the complexity seen in vivo. In the context of anti-PD-1 immune checkpoint inhibitors, we compare treatment outcomes simulated from an ODE model and an ABM to show the importance of including spatial components in computational models of cancer immunotherapy. We consider tumor cells of high and low antigenicity and two distinct cytotoxic T lymphocyte (CTL) killing mechanisms, which differ based on the antigenicity of tumor cells. Our ABM reveals different spatial distributions of tumor and immune cells even when the temporal trajectories look similar. We also use machine learning (ML) to predict tumor status in the intermediate and long term based on immediate treatment responses in the ABM. Our ML-assisted method eliminates the need to simulate the ABM until equilibrium and suggests an effective way to reduce computational time of ABMs.

Yixuan Wang
University of Michigan
shirlynw@umich.edu

Daniel Bergman
University of Michigan
Department of Mathematics
bergmand@umich.edu

Erica Trujillo, Anthony Fernald, Lie Li, Alexander Pearson, Randy Sweis
University of Chicago
Department of Medicine, Section of Hematology/Oncology
erica.trujillo815@gmail.com, afernald@bsd.uchicago.edu, lli2@uchicagomedicine.org, alexander.pearson@uchicagomedicine.org, rsweis@bsd.uchicago.edu

Trachette Jackson
University of Michigan - Ann Arbor
tjacks@umich.edu

PP4

MS73: Aonn-2: A Deep Learning Based Shape Optimization Method

Shape optimization has been playing an important role in a large variety of engineering applications. Existing shape optimization methods are generally mesh-dependent and therefore encounter challenges due to mesh deformation. To overcome this limitation, we present a new adjoint-oriented neural network method, AONN-2, for PDE-constrained shape optimization problems. This method extends the capabilities of the original AONN method, which is developed for efficiently solving parametric optimal control problems. AONN-2 inherits the direct-adjoint looping (DAL) framework for computing the extremum of an objective functional and the neural network methods for solving complicated PDEs from AONN. Furthermore, AONN-2 expands the application scope to shape optimization by taking advantage of the shape derivatives to optimize the shape represented by discrete boundary points. AONN-2 is a fully mesh-free shape optimization approach, naturally sidestepping issues related to mesh deformation, with no needs for maintaining mesh quality and additional mesh corrections. A series of experimental results are presented, highlighting the flexibility, robustness, and accuracy of AONN-2.

Xili Wang
Peking University, Beijing, China
xiliwang@stu.pku.edu.cn

PP4

MS69: Hierarchical Bayesian Inverse Problems: a High-Dimensional Statistics Viewpoint

We analyze hierarchical Bayesian inverse problems using techniques from high dimensional statistics. Our analysis leverages a property of hierarchical Bayesian regularizers that we call approximate decomposability to obtain non-asymptotic bounds on the reconstruction error attained by maximum a posteriori estimators. The new theory explains how hierarchical Bayesian models that exploit sparsity, group sparsity, and sparse representations of the unknown parameter can achieve accurate reconstructions in high-dimensional settings.

Nathan Waniorek, Daniel Sanz-Alonso
University of Chicago
waniorek@uchicago.edu, sanzalonso@uchicago.edu

PP4

MS71: Using Identifiability Analysis to Evaluate Wastewater Surveillance and Public Health Data As Indirect Observables to Inform Predictive Disease Models

Disease modeling facilitates model-based predictions for public health purposes using limited and indirect observations of unknown fidelity and noise, like cases and hospital admission rates. Wastewater-based surveillance is an increasingly available data stream which may improve calibration of disease models. Unlike traditional public health measures, wastewater samples reflect the entire population in a sewershed community since individuals infected with SARS-CoV-2 shed viral RNA in their stool regardless of symptomology. But the utility of these measurements to inform models is unknown and depends on both functional characteristics of the chosen disease model and quality of measurements. In this study, we evaluate whether wastewater surveillance data improves the calibration of parameters in compartmental disease models using identifiability analysis. Structural identifiability methods determine whether the parameters of a model can be uniquely identified locally or globally, given specified observables and assuming that data is perfect, using differential algebra approaches. Practical identifiability methods assess the influence of data quality on parameter estimation given an actual data set. Identifiability analysis informs the trust one gives parameter values, which can have important mean-

ing for public health decision-making. We characterize the identifiability of disease transmission models under cases for varying data availability and quality during a pandemic.

Maria M. Warns, Katelyn Plaisier Leisman, Shuotao Diao, Guyi Chen, David Morton, Jaline Gerardin, Aaron Packman
Northwestern University
mariawarns2025@u.northwestern.edu,
katelyn.leisman@northwestern.edu,
shuotao.diao@northwestern.edu,
guyichen2024@u.northwestern.edu,
david.morton@northwestern.edu,
jgerardin@northwestern.edu,
a-packman@northwestern.edu

Niall M. Mangan
Engineering Sciences and Applied Mathematics
Northwestern University
niall.mangan@northwestern.edu

PP4

MS68: Performance Analysis of Sequential Experimental Design for Calibration in Parallel Computing Environments

Calibration of computationally expensive simulations is usually carried out with a data-driven emulator. The effectiveness of the calibration process can be significantly improved by using a sequential/online construction of such an emulator. The expansion of parallel computing environments can lead to further calibration efficiency gains by allowing for the evaluation of the simulation model at a batch of parameters in parallel in a sequential design. However, understanding the performance implications of different sequential approaches in parallel computing environments introduces new complexities since the rate of the speed-up is affected by many factors, such as the run time of a simulation model and the variability in the run time. This work proposes a new way to understand and benchmark the performance of different sequential procedures for the calibration of simulation models in parallel environments. We provide metrics and a suite of techniques for visualizing the numerical experiment results and demonstrate these with a novel sequential procedure. The proposed performance model and the sequential procedure along with the state-of-art techniques are implemented in the Python software package Parallel Uncertainty Quantification (PUQ), which allows users to run a simulation model in parallel. PUQ is an open-source software package at https://github.com/parallelUQ/PUQ

Stefan Wild
Lawrence Berkeley National Laboratory
Applied Mathematics and Computational Research
wild@lbl.gov

Ozge Surer
Department of Information Systems & Analytics
Miami University
surero@miamioh.edu

PP4

MS59: Large Stepsize Gradient Descent for Logistic Loss: Non-Monotonicity of the Loss Improves Optimization Efficiency

We consider gradient descent (GD) with a constant stepsize applied to logistic regression with linearly separable data, where the constant stepsize η is so large that the loss initially oscillates. We show that GD exits this initial oscillatory phase rapidly — in $\mathcal{O}(\eta)$ steps, and subsequently achieves an $\tilde{\mathcal{O}}(1/(\eta t))$ convergence rate after t additional steps. Our results imply that, given a budget of T steps, GD can achieve an *accelerated* loss of $\tilde{\mathcal{O}}(1/T^2)$ with an aggressive stepsize $\eta := \Theta(T)$, without any use of momentum or variable stepsize schedulers. Our proof technique is versatile and also handles general classification loss functions (where exponential tails are needed for the $\tilde{\mathcal{O}}(1/T^2)$ acceleration), nonlinear predictors in the neural tangent kernel regime, and online *stochastic gradient descent* (SGD) with a large stepsize, under suitable separability conditions.

Jingfeng Wu, Peter Bartlett
University of California, Berkeley
uuujf@berkeley.edu, peter@berkeley.edu

Matus Telgarsky
New York University
mjt10041@nyu.edu

Bin Yu
University of California, Berkeley
binyu@berkeley.edu

PP4

MS73: A Probabilistic Approach to Shape Derivatives

We introduce a novel mesh-free and direct method for computing the shape derivative in PDE-constrained shape optimization problems. Our approach is based on a probabilistic representation of the shape derivative and is applicable for second-order elliptic PDEs with Dirichlet boundary conditions and a general class of target functions. The probabilistic representation derives from a boundary sensitivity result for diffusion processes due to Constantini, Gobet and El Karoui. We provide a Taylor test to verify the accuracy of our methodology.

Maximilian Wuerschmidt
Trier University, Germany
wuerschmidt@uni-trier.de

Frank Seifried
University of Trier
seifried@uni-trier.de

Volker H. Schulz
University of Trier
Department of Mathematics
Volker.Schulz@uni-trier.de

Schlegel Luka
Trier University
schlegel@uni-trier.de

PP4

MS59: Provable Acceleration of Nesterov's Accelerated Gradient for Asymmetric Matrix Factorization and Linear Neural Networks

We study the convergence rate of first-order methods for the asymmetric matrix factorization (MF) problem, which is a prototype of a broad class of nonconvex optimization problems. Specifically, for a rank-r matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, we show that gradient descent (GD) can find $\mathbf{X}_T \in \mathbb{R}^{m\times d}$

and $\mathbf{Y}_T \in \mathbb{R}^{n\times d}$ satisfying $\|\mathbf{X}_T\mathbf{Y}_T^\top - \mathbf{A}\|_{\mathrm{F}} \le \epsilon\|\mathbf{A}\|_{\mathrm{F}}$ in $T = O(\kappa^2 \log\frac{1}{\epsilon})$ iterations with high probability, where $d \ge r$ and κ denotes the condition number of $\mathbf{A}$. Furthermore, we show that an $O(\kappa \log\frac{1}{\epsilon})$ accelerated rate can be attained by Nesterov's accelerated gradient (NAG) method. Different from standard balanced initialization, we adopt an unbalanced initialization. Our initialization and analysis can be extended to linear neural networks, where we show that NAG can attain a linear convergence rate. In particular, our width requirement only scales with the rank of the output matrix, while previous results achieving the same rate require excessive widths that depend on the condition number of the data matrix.

Zhenghao Xu
Georgia Institute of Technology
zhenghaoxu@gatech.edu

Molei Tao, Molei Tao
School of Mathematics, Georgia Institute of Technology
mtao@gatech.edu, mtao@gatech.edu

Yuqing Wang
Georgia Institute of Technology
ywang3398@gatech.edu

Rachel Ward
Department of Mathematics
University of Texas at Austin
rward@math.utexas.edu

Tuo Zhao
Georgia Institute of Technology
tourzhao@gatech.edu

PP4

MS59: Over-Parameterization Exponentially Slows Down Gradient Descent for Learning a Single Neuron

We revisit the problem of learning a single neuron with ReLU activation under Gaussian input with square loss. We particularly focus on the over-parameterization setting where the student network has $n \ge 2$ neurons. We prove the global convergence of randomly initialized gradient descent with a $O\left(T^{-3}\right)$ rate. This is the first global convergence result for this problem beyond the exact-parameterization setting ($n = 1$) in which the gradient descent enjoys an $\exp(-\Omega(T))$ rate. Perhaps surprisingly, we further present an $\Omega\left(T^{-3}\right)$ lower bound for randomly initialized gradient flow in the over-parameterization setting. These two bounds jointly give an exact characterization of the convergence rate and imply, for the first time, that *over-parameterization can exponentially slow down the convergence rate*. To prove the global convergence, we use a three-phase structure to analyze GD dynamics. Along the way, we prove gradient descent automatically balances student neurons, and use this property to deal with the non-smoothness of the objective function. To prove the convergence rate lower bound, we construct a novel potential function that characterizes the pairwise distances between the student neurons. We show this potential function converges slowly, which implies the slow convergence rate of the loss function.

Weihang Xu, Simon Du
University of Washington
xuwh01@gmail.com, ssdu@cs.washington.edu

PP4

MS64: Gaussian Process Regression under Computational and Epistemic Misspecification

Gaussian process regression is a classical kernel method for function estimation and data interpolation. In large data applications, computational costs can be reduced using low-rank or sparse approximations of the kernel. This poster investigates the effect of such kernel approximations on the interpolation error. We introduce a unified framework to analyze Gaussian process regression under important classes of computational misspecification: Karhunen-Love expansions that result in low-rank kernel approximations, multiscale wavelet expansions that induce sparsity in the covariance matrix, and finite element representations that induce sparsity in the precision matrix. Our theory also accounts for epistemic misspecification in the choice of kernel parameters.

Ruiyi Yang
Princeton University
ry8311@princeton.edu

Daniel Sanz-Alonso
University of Chicago
sanzalonso@uchicago.edu

PP4

OptiBridge: Multi-Scale Multi-Shift Bridging for Conditioning Optimization Landscapes

This paper introduces OptiBridge, a novel optimization methodology designed to address the challenges of navigating complex optimization landscapes characterized by numerous local optima, such as those found in NP-hard problems. By implementing a multi-scale, multi-shift strategy, OptiBridge significantly enhances the attraction domains of superior local optima while concurrently diminishing the prevalence of suboptimal local solutions. This will help with the performance of off-the-shelf local optimizers, such as first-order methods. We also introduce BridgeNet, a neural network architecture based on OptiBridge that learns the best shifts and scales for the OptiBridge objective to help produce an enhanced subset of local optima. Through theoretical analysis and illustrative examples, we demonstrate the efficacy of OptiBridge in improving the outcomes of local optimizers across various problem domains, particularly highlighting its potential in dominant frequency estimation and training of lightweight neural networks.

Farhang Yeganegi, Arian Eamaz, Mojtaba Soltanalian
University of Illinois - Chicago, IL
University of Illinois - Chicago, IL
fyegan2@uic.edu, aeamaz2@uic.edu, msol@uic.edu

PP4

MS76: Improved Dimension Dependence of a Proximal Algorithm for Sampling

In this proposal, I will introduce a novel sampling algorithm based on the proximal sampler framework introduced by Lee et al. (2021). This new approximate proximal sampler achieves superior complexity bounds across almost all the settings, such as strong log-concavity, log-concavity, functional inequalities, as well as those involving semi-

smooth or composite potentials. The core of our algorithm is an inexact realization of the restricted Gaussian oracle (RGO) based on approximate rejection sampling. The state-of-the-art complexity bounds are largely established by a new concentration inequality for semi-smooth functions over Gaussian distributions. For strongly log-concave distributions, our method has complexity bound $\tilde{\mathcal{O}}(\kappa d^{1/2})$ without warm start, better than the minimax bound for MALA.

Bo Yuan
Georgia Institute of Technology
byuan48@gatech.edu

PP4

Constrained Cost-Aware Multi-Fidelity Bayesian Optimization

Bayesian optimization (BO) is a widely used technique for finding materials with unprecedented properties. However, relying solely on expensive high-fidelity (HF) sources can inflate optimization expenses in complex scenarios. To address this challenge and incorporate known and unknown constraints, we introduce a novel constrained cost-aware multi-fidelity BO (C2-MFBO) framework which a few novelties. Firstly, it uses manifold-embedded Gaussian process (GP) for emulation which handles mixed input spaces and models source-dependent noise and global trends. Secondly, it leverages a composite acquisition function (AF) that quantifies the information value of high- and low-fidelity sources differently and also accommodates source-dependent constraints. Lastly, it pioneers an optimization-based stop condition instead of immature stopping after a fixed number of iterations. Through analytical and real-world examples, we will demonstrate the benefits of our approach which is publicly available via the GP+ package in Python.

Zahra Zanjani Foumani
University of California Irvine
zzanjani@uci.edu

PP4

MS77: Self Supervised Representation Learning of Text-Attributed Graphs

The emergence of pre-trained models has revolutionized various AI research domains, such as natural language processing and computer vision. They function as a foundation model to generalize across unseen domains to enable label-scarce tasks such as zero-shot and few-shot predictions. Although recently there has been growing interests in developing pre-trained graph models, their generalization capabilities, especially in unseen domains, remain limited. We aim to build a unified graph representation learning framework designed to enable effective transfer across different graph domains in a label-scarcity manner. We first propose a graph-to-text transformation technique that represent each node's neighborhood structure in a hierarchical document layout. Second, we propose a self-supervised learning framework to align the textual representations of these documents with aggregated neighborhood information derived from GNNs, integrating the representations within the textual domain. To address scalability and efficiency challenges in the pre-training stage, particularly when dealing with large neighborhood sizes which results in long context size of documents, we propose a Metropolis-Hastings random walk sampling method to approximate neighborhood structures. Extensive experiments on eight real-world datasets demonstrate that our framework possesses strong zero-shot and few-shot prediction capabilities, effectively generalizing the pre-trained model to unseen graph datasets.

Zheng Zhang
Emory University
zheng.zhang@emory.edu

PP4

MS71: Bayesian Inference of Stochastic Dynamical Systems with Inhomogeneous Poisson Noise

Discovering governing laws from noisy data is crucial for understanding the underlying dynamics. While data-driven algorithms have emerged to learn ordinary and partial differential equations, they have primarily been applied to synthetic or physical systems that are largely deterministic with limited noise. Learning stochastic differential equations (SDEs) from noisy data remains a challenge. Here, we present a framework capable of inferring SDEs with inhomogeneous Poisson noise directly from experimental time-series data. Our framework combines basis function representation and sparse Bayesian inference to discover SDEs that capture the experimentally observed ensemble statistics. By applying our framework to biological data of cell growth and division, we discover a nonlinear memory in cell division across various species. The approach presented here is also directly applicable to infer stochastic dynamical models from real-world healthcare and geoscience data.

Shijie Zhang
Massachusetts Institute of Technology
shjzhang@mit.edu

PP4

MS76: Sampling from the Mean-Field Stationary Distribution

We study the complexity of sampling from the stationary distribution of a mean-field SDE, or equivalently, the complexity of minimizing a functional over the space of probability measures which includes an interaction term. Our main insight is to decouple the two key aspects of this problem: (1) approximation of the mean-field SDE via a finite-particle system, via uniform-in-time propagation of chaos, and (2) sampling from the finite-particle stationary distribution, via standard log-concave samplers. Our approach is conceptually simpler and its flexibility allows for incorporating the state-of-the-art for both algorithms and theory. This leads to improved guarantees in numerous settings, including better guarantees for optimizing certain two-layer neural networks in the mean-field regime.

Matthew Zhang
University of Toronto
matthew.zhang@mail.utoronto.ca

Yunbum Kook
KAIST
Discrete Mathematics Group, Institute for Basic Science
yb.kook@gatech.edu

Murat Erdogdu
University of Toronto
erdogdu@cs.toronto.edu

Sinho Chewi

Institute for Advanced Study
schewi@ias.edu

Mufan Li
Princeton University
mufan.li@princeton.edu

PP4

MS61: Accelerating Approximate Thompson Sampling with Underdamped Langevin Monte Carlo

Approximate Thompson sampling with Langevin Monte Carlo broadens its reach from Gaussian posterior sampling to encompass more general smooth posteriors. However, it still encounters scalability issues in high-dimensional problems when demanding high accuracy. To address this, we propose an approximate Thompson sampling strategy, utilizing underdamped Langevin Monte Carlo, where the latter is the go-to workhorse for simulations of high-dimensional posteriors. Based on the standard smoothness and log-concavity conditions, we study the accelerated posterior concentration and sampling using a specific potential function. This design improves the sample complexity for realizing logarithmic regrets from $\tilde{\mathcal{O}}(d)$ to $\tilde{\mathcal{O}}(\sqrt{d})$. The scalability and robustness of our algorithm are also empirically validated through synthetic experiments in high-dimensional bandit problems.

Haoyang Zheng
Purdue University
zheng528@purdue.edu

Wei Deng
Morgan Stanley
weideng056@gmail.com

Christian Moya, Guang Lin
Purdue University
cmoyacal@purdue.edu, guanglin@purdue.edu

PP4

MS72: A Mean-Field Analysis of Neural Gradient Descent-Ascent: Applications to Functional Conditional Moment Equations

We study minimax optimization problems defined over infinite-dimensional function classes. We restrict the functions to the class of overparameterized two-layer neural networks and study (i) the convergence of the gradient descent-ascent algorithm and (ii) the representation learning of the neural network. We consider the minimax optimization problem stemming from estimating a functional equation defined by conditional expectations via adversarial estimation, where the objective function is quadratic in the functional space. We establish convergence under the mean-field regime by considering the continuous-time and infinite-width limit of the optimization dynamics. Under this regime, gradient descent-ascent corresponds to a Wasserstein gradient flow over the space of probability measures defined over the space of neural network parameters. We prove that the Wasserstein gradient flow converges globally to a stationary point of the minimax objective at a sublinear rate, and additionally finds the solution to the functional equation when the regularizer of the minimax objective is strongly convex. In terms of representation learning, our results show that the feature representation induced by the neural networks is allowed to deviate from the initial one by the magnitude of $O(\alpha^{-1})$, measured in terms of the Wasserstein distance. We apply our results to concrete examples including policy evaluation, nonparametric instrumental variable regression, and asset pricing.

Yuchen Zhu
Georgia Institute of Technology
yzhu738@gatech.edu ¡yzhu738@gatech.edu¿

Yufeng Zhang, Zhaoran Wang
Northwestern University
yufengzhang2023@u.northwestern.edu,
zhaoranwang@gmail.com

Zhuoran Yang, Xiaohong Chen
Yale University
zhuoran.yang@yale.edu, xiaohong.chen@yale.edu

PP5

MS80: Bias-Variance Trade-off in Multiscale Interpolation

We introduce a multiscale scheme for scattered data approximation. Our approach leverages a given interpolation scheme and constructs a hierarchical approximation over levels where a more significant portion of the data set is considered at each level. In particular, at the first level, one gets a coarse approximation from a small portion of data, while, in the following levels, we gradually correct the approximation and refine it to improve the approximation quality. This work investigates the statistical properties of error, mainly through the lens of the bias-variance trade-off. We thoroughly explore these properties numerically to approximate real and manifold-valued functions.

Asaf Abas
tel aviv university
asafabas@tau.ac.il

Nir Sharon
Tel Aviv University
nsharon@tauex.tau.ac.il

PP5

MS78: The Loss Landscape of Deep Linear Neural Networks: a Second-Order Analysis

We study the optimization landscape of deep linear neural networks with the square loss. It is known that, under weak assumptions, there are no spurious local minima and no local maxima. However, the existence and diversity of non-strict saddle points, which can play a role in first-order algorithms' dynamics, have only been lightly studied. We go a step further with a full analysis of the optimization landscape at order 2. We characterize, among all critical points, which are global minimizers, strict saddle points, and non-strict saddle points. We enumerate all the associated critical values. The characterization is simple, involves conditions on the ranks of partial matrix products, and sheds some light on global convergence or implicit regularization that have been proved or observed when optimizing linear neural networks. In passing, we provide an explicit parameterization of the set of all global minimizers and exhibit large sets of strict and non-strict saddle points.

El Mehdi Achour
RWTH Aachen University

achour@mathc.rwth-aachen.de

PP5

MS86: Entropy-Regularized Wasserstein Barycenter Functionals: Theory and Applications

We investigate functional-analytic, statistical, and algorithmic aspects related to analysis and synthesis of measures using entropy-regularized Wasserstein-2 barycenters. We highlight and contrast our results with the unregularized case, pointing specifically to dimension-independent rates of statistical estimation when using entropy-regularization under mild assumptions. Furthermore, we give theoretical justification for the use of a popular method for synthesis of free-support entropy-regularized Wasserstein-2 barycenters. Finally, we demonstrate promising results suggesting the applicability of analysis coefficients to point-cloud classification under possible corruptions and occlusions.

Brendan Mallery, James Murphy, Shuchin Aeron
Tufts University
brendan.mallery@tufts.edu, jm.murphy@tufts.edu, shuchin@ece.tufts.edu

PP5

A Deep Learning Framework for EKG Diagnosis

This project introduces a deep-learning framework for multi-label classification of EKG data. The framework converts EKG wavelet data into images to be used as inputs into a specially-tuned CNN model. To improve model performance, a data augmentation method is used to increase the data set size before training the network, which is done through the implementation of a generative adversarial network (GAN). The framework is implemented in PyTorch and tested on the MIT-BIH Arrythmia database. Classification performance using our augmentation framework shows is comparable to that of previous state-of-the-art (SOTA) image classification models. As a persistent and pronounced issue in health-related datasets arises from both the quality and quantity of existing data, we have evidence that deep-learning frameworks such as these may be effective solutions to complement the available technical tools accessible to medical professionals.

Eric Roginek
Fordham University
Fordham University
Eroginek@fordham.edu

Courtney M. King, Ali Alfatemi, Yijun Zhao
Fordham University
cking74@fordham.edu, aalfatemi@fordham.edu, yzhao11@fordham.edu

PP5

MS79: Characterizing Single-Cell Transcriptomic Signatures with Persistent Homology and Molecular Cartography

The central dogma of molecular biology follows a simple path: DNA is transcribed into RNA transcripts in the cell nucleus, and transcripts are then translated into proteins in the cell cytosol. However, protein production is not solely impacted by the level of expression of genes, but by many other regulatory processes that can be specific to the gene type and even specific to the cell spatial location. Different cells of different shapes and sizes present different RNA transcript distribution for different genes. To mathematically model and analyze this highly variable distributions, we turn to Topological Data Analysis (TDA) for a robust and comprehensive pipeline. Here, through the use of single-cell sequencing and molecular cartography technologies, we first produce detailed maps of the spatial location of individual transcripts for different genes, cell types, and organs of the soybean root and nodule. We then use persistent homology to characterize the distribution of these transcripts for each cell. Comparing these topological shape signatures reveals a new perspective on the role of the nuclear and cytoplasmic localization of transcripts as a central mechanism to control protein translation and the biology of plant cells. This work reveals the influence of the compartmentalization of transcripts as another regulatory mechanism of protein translation and a new understanding of the central dogma of molecular biology.

Erik J. Amezquita
Michigan State University
eah4d@missouri.edu

Sutton Tennant, Sandra Thibivillers
University of Missouri
setqht@missouri.edu, st5y7@missouri.edu

Sai Subhash
University of Nebraska-Lincoln
smahamkalivenkatas2@huskers.unl.edu

Samik Bhattacharya, Jäsper Klaver
Resolve BioSciences GmbH
samik.bhattacharya@resolvebiosciences.com, jasper.klaver@resolvebiosciences.com

Benjamin Smith
UC Berkeley
benjamin.smith@berkeley.edu

Marc Libault
University of Missouri
libaultm@missouri.edu

PP5

Studying Disease Reinfection Rates, Vaccine Efficacy and the Timing of Vaccine Rollout in the Context of Infectious Diseases

This research uniquely explores the varied efficacy of existing vaccines and the pivotal role of vaccination timing in the context of COVID-19. We introduce two models that account for the impact of vaccines on infections, reinfections, and deaths. We estimate model parameters under the Bayesian framework, specifically utilizing the Metropolis-Hastings Sampler. The study conducts data-driven scenario analyses for the State of Qatar, quantifying the potential duration during which the healthcare system could have been overwhelmed by an influx of new COVID-19 cases surpassing available hospital beds. Additionally, the research explores similarities in predictive probability distributions of cumulative infections, reinfections, and deaths, employing the Hellinger distance metric. Comparative analysis, utilizing the Bayes factor, underscores the plausibility of a model assuming a different susceptibility rate to reinfection, as opposed to assuming the same susceptibility rate for both infections and reinfections. Results highlight the adverse outcomes associated with delayed vaccination, emphasizing the efficacy of early vacci-

nation in reducing infections, reinfections, and deaths. Our research advocates prioritizing early vaccination as a key strategy in effectively combating future pandemics. This study contributes vital insights for evidence-based public health interventions and reinforcing preparedness for challenges posed by infectious diseases.

Elizabeth B. Amona, Indranil Sahoo, Edward Boone
Virginia Commonwealth University
amonaeb@vcu.edu, sahooi@vcu.edu, elboone@vcu.edu

Ryad Ghanam
Virginia Commonwealth University, Doha, Qatar
rghanam@vcu.edu

PP5

MS81: Joint Inference of Clonal Structure Using Single-Cell Genome and Transcriptome Sequencing Data

Latest advancements in the high-throughput single-cell genome (scDNA) and transcriptome (scRNA) sequencing technologies enabled cell-resolved investigation of tissue clones. However, it remains challenging to cluster and couple single cells for heterogeneous scRNA and scDNA data generated from the same specimen. In this study, we present a computational framework called CCNMF, which employs a novel Coupled-Clone Non-negative Matrix Factorization technique to jointly infer clonal structure for matched scDNA and scRNA data. CCNMF couples multi-omics single cells by linking copy number and gene expression profiles through their general concordance. It successfully resolved the underlying coexisting clones with high correlations between the clonal genome and transcriptome from the same specimen. We validated that CCNMF can achieve high accuracy and robustness using both simulated benchmarks and real-world applications, including an ovarian cancer cell lines mixture, a gastric cancer cell line, and a primary gastric cancer. In summary, CCNMF provides a powerful tool for integrating multi-omics single-cell data, enabling simultaneous resolution of genomic and transcriptomic clonal architecture. This computational framework facilitates the understanding of how cellular gene expression changes in conjunction with clonal genome alternations, shedding light on the cellular genomic difference of subclones that contributes to tumor evolution.

Xiangqi Bai
Stanford University
School of Medicine
xiangqi@stanford.edu

PP5

Spectral Triadic Decompositions of Real-World Networks

A fundamental problem in mathematics and network analysis is to find conditions under which a graph can be partitioned into smaller pieces. A ubiquitous tool for this partitioning is the Fiedler vector or discrete Cheeger inequality. These results relate the graph spectrum (eigenvalues of the normalized adjacency matrix) to the ability to break a graph into two pieces, with few edge deletions. An entire subfield of mathematics, called spectral graph theory, has emerged from these results. Yet these results do not say anything about the rich community structure exhibited by real-world networks, which typically have a significant fraction of edges contained in numerous densely clustered blocks. Inspired by the properties of real-world networks, we discover a new spectral condition that relates eigenvalue powers to a network decomposition into densely clustered blocks. We call this the *spectral triadic decomposition*. Our relationship exactly predicts the existence of community structure, as commonly seen in real networked data. Our proof provides an efficient algorithm to produce the spectral triadic decomposition. We observe on numerous social, coauthorship, and citation network datasets that these decompositions have significant correlation with semantically meaningful communities.

Sabyasachi Basu
University of California, Santa Cruz
sbasu3@ucsc.edu

Suman Kalyan Bera
Apple
sumankalyanbera@gmail.com

C. Seshadhri
University of California, Santa Cruz
sesh@ucsc.edu

PP5

MS87: Frequentist Confidence Intervals Via Optimization: Resolving the Burrus Conjecture

We introduce an optimization-based framework to construct confidence intervals for functionals in constrained inverse problems, ensuring valid one-at-a-time frequentist coverage guarantees. Our method builds upon the now-called strict bounds intervals which offer ways to directly incorporate any side information about parameters during inference without introducing external biases. By tying these intervals to an inversion of a constrained likelihood ratio test, we translate interval coverage guarantees into type-I error control, and characterize the resulting interval via solutions of optimization problems. Along the way, we refute the Burrus conjecture. Our framework provides a novel approach to analyze the conjecture and construct a counterexample by employing a stochastic dominance argument, which we also use to disprove a general form of the conjecture. We illustrate our framework with several numerical examples and provide directions for extensions beyond the Rust-Burrus method for non-linear, non-Gaussian settings with general constraints.

Pau Batlle
Caltech
pbatllef@caltech.edu

PP5

MS91: Learning Parameters for Mapper Graphs

The Mapper algorithm is a popular visualization technique in topological data analysis (TDA) that outputs a graph reflecting the structure of a given dataset. We illustrate Mapper's high sensitivity to its input parameters by answering the inverse problem: Given a graph G and data X, does there exist a set of Mapper parameters such that the Mapper graph of X is isomorphic to G? Additionally, we introduce G-Mapper, an iterative and statistical algorithm for optimizing parameters.

Robin Belton
Smith College, U.S.

rbelton@vassar.edu

PP5

MS92: MALADY: Multistage Active Learning with Auction Dynamics on Graphs

The choice of training points often has a significant impact on the performance of a machine learning model, particularly in semi-supervised learning (SSL) scenarios where the training set is small. Active learning is a sub-field of machine learning that helps to improve the performance of underlying machine learning methods by carefully selecting unlabeled points to be labeled, using human or oracle input in the loop. In this project, we design a multistage active learning framework that first explores and then exploits the input data space while selecting unlabeled points to be labeled. In particular, we designed a novel active learning acquisition function to measure uncertainty in the auction dynamics algorithm on the graphical framework. We validate our proposed framework with experimental results on a number of data sets.

Gokul Bhusal
Michigan State University, U.S.
bhusalgo@msu.edu

Ekaterina Rapinchuk
Michigan State University
Department of Mathematics
merkurje@msu.edu

Kevin Miller
Department of Mathematics
Brigham Young University, Provo
ksmiller@math.byu.edu

PP5

MS89: White-Box Transformers Via Sparse Rate Reduction

In this work, we contend that a natural objective of representation learning is to compress and transform the distribution of the data, say sets of tokens, towards a low-dimensional Gaussian mixture supported on incoherent subspaces. The goodness of such a representation can be evaluated by a principled measure, called sparse rate reduction, that simultaneously maximizes the intrinsic information gain and extrinsic sparsity of the learned representation. From this perspective, popular deep network architectures, including transformers, can be viewed as realizing iterative schemes to optimize this measure. Particularly, we derive a transformer block from alternating optimization on parts of this objective: the multi-head self-attention operator compresses the representation by implementing an approximate gradient descent step on the coding rate of the features, and the subsequent multi-layer perceptron sparsifies the features. This leads to a family of white-box transformer-like deep network architectures, which we call CRATE, which are mathematically fully interpretable. Experiments show that these networks, despite their simplicity, indeed learn to compress and sparsify representations of large-scale real-world image and text datasets, and achieve performance very close to highly engineered transformer-based models, including ViT and GPT2.

Sam Buchanan
Toyota Technological Institute at Chicago
sam@ttic.edu

Yaodong Yu
University of Maryland, U.S.
yyu@eecs.berkeley.edu

Druv Pai
University of California, Berkeley, U.S.
druvpai@berkeley.edu

Tianzhe Chu
ShanghaiTech University, China
tianzhechu@gmail.com

Ziyang Wu
University of California, Berkeley, U.S.
zywu@berkeley.edu

Shengbang Tong
New York University, U.S.
tsb@berkeley.edu

Hao Bai
University of Illinois Urbana-Champaign, U.S.
haob2@illinois.edu

Yuexiang Zhai
University of California, Berkeley, U.S.
simonzhai@berkeley.edu

Benjamin Haeffele
Johns Hopkins University
bhaeffele@jhu.edu

Yi Ma
University of California, Berkeley
yima@eecs.berkeley.edu

PP5

MS85: Approaches for Deep Neural Network Architecture Adaptation

This poster presents two novel strategies for progressively adapting neural network architecture along the depth. In particular, we attempt to address the following questions in a mathematically principled way: i) Where to add new capacity (layer) during the training process? ii) How to initialize the new capacity?. The first approach is based on a layerwise training strategy where a new layer is added each time and trained independently by freezing parameters in the previous layers. The algorithm uses a data-dependent regularizer that force each hidden layer to learn meaningful representation of the data. The second approach is based on defining the sensitivity for a neural network which is conceptually the derivative of a lost function with respect to infinitesimal changes in the neural network. Using an optimal control viewpoint, we show that the network derivative exists under certain conditions and a closed-form expression for the same is derived. The algorithm we derived simply determines the most sensitive location along the depth where a new layer needs to be inserted during the training process and the associated parametric initialization for the newly added layer. Numerical investigations with feed-forward networks (fully connected network and convolutional neural network) on prototype regression and classification problems demonstrate that both our proposed approaches can outperform an ad-hoc baseline network and other neural architecture adaptation strategies.

Krishnanunni C G

University of Texas at Austin
krishnanunni@utexas.edu

PP5

MS82: Autm Flow: Atomic Unrestricted Time Machine for Monotonic Normalizing Flows

Normalizing flows constitute an important class of generative models. It tries to provide a generally nonlinear bijective mapping between the base distribution and the target distribution. How to design an architecture that is computationally efficient and expressive is a central topic. We present a novel normalizing flow called AUTM that allows triangular Jacobian, tractable inverse calculation, and provable universality. We compare AUTM to popular flow architectures such as RealNVP, GLOW, NSF, FFJORD, UMNN, BNAF, etc. to demonstrate the efficiency and expressive power over machine learning and large scale image data sets.

Difeng Cai
Southern Methodist University
ddcai@smu.edu

Yuliang Ji
ByteDance
yuliang.ji@emory.edu

Huan He
University of Pennsylvania
huan.he@pennmedicine.upenn.edu

Qiang Ye
University of Kentucky
qye3@uky. edu

Yuanzhe Xi
Emory University
yxi26@emory.edu

PP5

Ordinal Pattern Preserving Surrogates for Non-Stationary Time Series

Most real-world time series exhibit non-stationary behavior, where the statistical properties of the underlying process change over time. Standard surrogate methods may not adequately capture the dynamics of the data because of the stationarity assumption. Inspired by the success of ordinal pattern based analysis of non-stationary time series, we developed a new surrogate method which preserves the ordinal patterns up to a predefined length. This new surrogate method, named "Order Preserving surrogates' is computationally efficient and works well for both stationary and non-stationary time series. We evaluate our method on both synthetic and real world time series. Compared to other existing surrogate methodologies, our Order Preserving surrogates do a better job of capturing the dynamics of various signals. This new surrogate method can be used for hypothesis testing in various applications.

Yan Cao
The University of Texas at Dallas
Department of Mathematical Sciences
yan.cao@utdallas.edu

Aprajita Singh
The University of Texas at Dallas
aprajita.singh@utdallas.edu

PP5

MS86: Largest Angle Path Distance for Multi-Manifold Clustering

We propose a novel, angle-based path metric for the multi-manifold clustering problem. This metric, which we call the largest-angle path distance (LAPD), is computed as a bottleneck path distance in a graph constructed on d-simplices of data points. Under reasonable assumptions, this distance is small between data points from the same manifold and is large for data points from different manifolds. We provide a theoretical guarantee for recovering such an LAPD gap, even when the data is polluted with noise and when the underlying manifolds are curved. When data is sampled from a collection of d-dimensional manifolds which may intersect, the method can cluster the manifolds with high accuracy and automatically detect how many manifolds are present. By leveraging fast approximation schemes for bottleneck distance, this method exhibits quasi-linear computational complexity in the number of data points. In addition to being highly scalable, the method outperforms existing algorithms in numerous numerical experiments on intersecting manifolds, and exhibits robustness with respect to noise and curvature in the data.

Haoyu Chen, Anna Little, Akil Narayan
University of Utah
Haoyu.Chen@utah.edu, little@math.utah.edu, akil.narayan@utah.edu

PP5

Techniques for the Robust Experimental Design of Nonlinear Bayesian Inverse Problems

The optimal design of experiments for Bayesian inverse problems is a field in the uncertainty quantification community that has attracted wide attention. An optimal design is one which maximizes a predefined metric of quality on the resulting inverse problem. However, such an optimal design is local to the selection of the other hyperparameters in the inverse problem, such as measurement uncertainty. Robust optimal experimental design is a recently introduced reformulation of the classical problem which considers a worst-case scenario approach regarding the uncertain parameters. However, such formulations were presented for linear(-ized) Bayesian inverse problems, in order to obtain tractable utility criteria evaluations and gradients. In this work, we present a framework tailored to infinite-dimensional Bayesian inverse problems constrained by large-scale partial differential equations for the efficient and scalable evaluation of an approximation to the expected information gain. In addition, we provide methods for its scalable differentiation leveraging eigenvalue sensitivity techniques. Together, along with an adaptation of state-of-the-art stochastic optimization techniques to the robust setting, we enable non-linear robust optimal experimental design. Extensive numerical experiments to validate the proposed approach are carried out for sensor placement in a classical nonlinear parameter identification problem.

Abhijit Chowdhary
North Carolina State University
achowdh2@ncsu.edu

Ahmed Attia

Argonne National Laboratory
aattia@anl.gov

Alen Alexanderian
North Carolina State University, U.S.
alexanderian@ncsu.edu

PP5

MS82: Topological Machine Learning for Drug Discovery

In this work, we present applications of latest topological machine learning methods in virtual screening and computer aided drug discovery.

Baris Coskunuzer
University of Texas at Dallas
coskunuz@utdallas.edu

PP5

Predictive Modeling of Complex Dynamical Systems Using Machine Learning: A Comparative Analysis

Dynamical systems exhibit intricate behaviors that challenge traditional modeling techniques. The integration of machine learning methods provides a promising avenue for understanding and predicting the evolution of such complex systems. This research endeavors to conduct a comprehensive comparative analysis of various machine learning approaches employed in predictive modeling for complex dynamical systems. The study will delve into the performance of both traditional time-series models and state-of-the-art deep learning techniques in capturing the intricate dynamics of diverse systems. We will consider applications across different domains, ranging from financial markets and climate science to biological systems. The primary objectives include evaluating the accuracy, efficiency, and robustness of these models in capturing the underlying dynamics of complex systems. The comparative analysis will encompass a meticulous examination of predictive capabilities under varying conditions and datasets. We aim to elucidate the strengths and weaknesses of each approach, shedding light on their suitability for specific types of dynamical systems. The findings of this research will contribute valuable insights into the selection of optimal modeling techniques, aiding researchers and practitioners in making informed decisions when confronted with complex dynamical systems.

Darlington S. David
University of Liberia, College of Engineering
davidds@ul.edu.lr

PP5

MS79: A Multi-Cellular Information-Theoretic Approach to Place Cell Population Learning

How the brain encodes its environment is a much-studied question with significant scientific and engineering implications. Place cells and grid cells, which fire at specific regions in an environment, enable mammals to distinguish locations and track positions. While the efficiency of grid cell coding has been extensively studied, the computational role of place cells is less well understood. This gap is partially due to the limited measure of spatial information, which, until now, has only been defined for one place cell. We derive and implement a higher-order spatial information measure, allowing for the study of the emergence of multiple place cells in a self-supervised manner. We show that emergent place cells have many desirable features, including high-accuracy spatial decoding. To our knowledge, this is the first work in which higher-order spatial information measures that depend solely on place cells' firing rates have been derived and which focuses on the emergence of multiple place cells via self-supervised learning. By enabling the quantification of spatial information for multiple place cells, we can now rigorously formulate hypotheses about place cell formation, function, and capabilities.

Jared Deighton
The University of Tennessee Knoxville
jdeighto@vols.utk.edu

Wyatt Mackey
Department of Mathematics / NIMBioS
The University of Tennesee, Knoxville
wmackey2@utk.edu

Ioannis Schizas, David Boothe
Army Research Lab
Aberdeen Proving Ground, MD, USA
ioannis.d.schizas.civ@army.mil,
david.l.boothe7.civ@army.mil

Vasileios Maroulas
University of Tennessee, Knoxville
vmaroula@utk.edu

PP5

MS91: Leveraging Derived Statistics from Differential Geometry for Understanding High-Dimensional TDA

We build a novel filtered complex, utilizing density estimation and classical linear fitting methods, from multi-dimensional Gaussian balls that serve as an approximate statistic for the local tangent space. This construction leverages density estimation, which is often more robust against outliers than traditional TDA methods. We apply this approach to large point-cloud datasets in the hopes of more faithful, humanly-intelligible visualizations.

Greg A. Depaul
UC Davis
gdepaul@ucdavis.edu

PP5

MS84: S-Step SGD and Federated SGD for Large Scale Optimization

We develop a hybrid, 2D parallel distributed-memory SGD algorithm by combining ideas from s-step methods and federated learning. The proposed algorithm attains a continuous tradeoff between accuracy and performance, while scaling beyond what is currently enabled by s-step SGD and federated SGD alone. We show empirical results that highlight better convergence and better performance when combining the s-step and federated SGD approaches into a 2D parallel algorithm. We also show theoretical parallel cost analyses which highlight the benefits of the hybrid, 2D SGD algorithm.

Aditya Devarakonda
Wake Forest University

devaraa@wfu.edu

PP5

MS78: Alternating Minimization for Regression with Tropical Rational Functions

We propose an alternating minimization heuristic for regression over the space of tropical rational functions, i.e. functions which are the difference of max-plus polynomials. The method alternates between fitting the numerator and denominator terms via tropical polynomial regression, which is known to admit a closed form solution. We demonstrate the behavior of the alternating minimization method experimentally. Our work is motivated by applications to ReLU neural networks, which are closely related to tropical rational functions. Joint work with Lars Ruthotto (Emory University).

Alex Dunbar
Emory University
alex.dunbar@emory.edu

PP5

A Graph-Based Feature Selection Algorithm

We propose a feature selection algorithm specifically designed for datasets with high dimensions and low number of samples. It makes use of autoencoder neural networks and weighted graphs for dimension reduction and data importance analysis respectively. The algorithm starts by extracting a low-dimensional representation for the input data. The resulting dimensions give us an inexact estimation of the number of features to be selected. To select the most important features, we evaluate the contribution rate of each feature based on its weight in the hidden layer of the autoencoder. By analyzing these weights, we can identify the optimal combination of features for creating this low-dimensional representation. For each node in the hidden layer (encoder), we determine the input feature with the highest weight and add it to a candidate set while removing it from the input set. Finally, we obtain a set of candidate features that are most relevant to achieve the desired low-dimensional representation. In addition, the neural network encoder layer acts as a feature extraction mechanism and enables us to obtain valuable representations from the input data. According to our tests, this algorithm provides comprehensive results for feature selection and extraction in high-dimensional datasets with limited sample size. It uses autoencoder neural networks to extract a low-dimensional representation and identify the most influential features by optimally update the weights.

Negin Bagherpour
University of Tehran
University of Tehran
negin.bagherpour@guest.ut.ac.ir

Behrang Ebrahimi
University of Tehran
behrangebrahimi13@gmail.com

PP5

MS85: Neural Differentiable Modeling with Diffusion-Based Superresolution for 2D Spatiotemporal Turbulence

In this paper, we introduce an innovative neural differentiable modeling framework designed to enhance the predictability and efficiency of spatiotemporal turbulence simulations. Our approach features hybrid differentiable modeling techniques that seamlessly integrate deep neural networks with numerical PDE solvers within a differentiable programming framework, synergizing deep learning techniques with physics-based CFD modeling. Specifically, a hybrid differentiable neural solver is constructed on a coarser grid to capture large-scale turbulent phenomena, followed by the application of a Bayesian conditional diffusion model that generates the small-scale turbulence conditioned on large-scale flow predictions. Two innovative hybrid architecture designs are studied, and their performance is evaluated through comparative analysis against conventional large eddy simulation techniques with physics-based subgrid-scale closures and purely data-driven deep neural solvers. The findings underscore the neural differentiable modeling framework's potential to significantly enhance the accuracy and computational efficiency of turbulence simulations. This study not only demonstrates the efficacy of merging deep learning with physics-based numerical solvers but also sets a new precedent for advanced CFD modeling techniques, highlighting the transformative impact of differentiable programming in scientific computing.

Jian-Xun Wang
Dept. of Aerospace and Mechanical Engineering
University of Notre Dame
jwang33@nd.edu

Xiantao Fan, Deepak Akhare
University of Notre Dame
fanxiantao1@gmail.com, dakhare@nd.edu

PP5

MS84: Anderson Acceleration in Federated Learning

Federated Learning is a decentralized machine learning approach that enables multiple local clients and a central servers to collaboratively learn a model while keeping their data locally. In this work, we introduce FedOSAA, an innovative optimization algorithm for federated learning that combines the simplicity of first-order methods with the accelerated convergence performance typically associated with second-order methods. During the local training, FedOSAA takes a few local gradient descent steps, followed by one Anderson acceleration step which accelerates the convergence while avoiding the common expense of Newton-based methods, i.e., the construction and inversion of the Hessian matrix. We establish a linear convergence rate to the global minimizer for FedOSAA on strongly convex losses. We compare FedOSAA with other state-of-the-art federated learning methods such as FedSVRG and GIANT on logistic regression problems. Numerical evidence demonstrates the superior performance of our algorithm in terms of communication and computation efficiency, which are bottlenecks in federated learning.

Xue Feng
University of California, Davis, U.S.
xffeng@ucdavis.edu

Paul Laiu
Oak Ridge National Laboratory
laiump@ornl.gov

Thomas Strohmer
University of California,Davis

Department of Mathematics
strohmer@math.ucdavis.edu

PP5

MS88: The Robust Outcome Test: A Simple, Statistically Robust Test of Discrimination

In observational studies of discrimination, the most common statistical approaches consider either the rate at which decisions are made (benchmark tests) or the success rate of those decisions (outcome tests). Both tests, however, have well-known statistical limitations, sometimes suggesting discrimination even when there is none. Despite the fallibility of the benchmark and outcome tests individually, here we prove a surprisingly strong statistical guarantee: under a common non-parametric assumption, at least one of the two tests must be correct; consequently, when both tests agree, they are guaranteed to yield correct conclusions. We present empirical evidence that the underlying assumption holds approximately in several important domains, including lending, education, and criminal justice – and that our hybrid test is robust to the moderate violations of the assumption that we observe in practice. Applying this approach to 2.8 million police stops across California, we find evidence of widespread racial discrimination.

Johann Gaebler
Harvard University
me@jgaeb.com

Sharad Goel
Kennedy School of Government
Harvard University
sgoel@hks.harvard.edu

PP5

MS91: Computing a Loss Function to Bound the Interleaving Distance for Mapper Graphs

Mapper graphs preserve the connected components of the inverse image function $f : \mathbb{X} \to \mathbb{R}$ over any given cover. Inspired by the interleaving distance for Reeb graphs, (Chambers et al.2024) extends this notion of distance to discretized mapper graphs. The distance is upper-bounded using a loss function. Unlike the NP-hard interleaving distance computation for Reeb graphs, the algorithm of the loss function has polynomial complexity. In this paper, we implement the categorical framework of mapper graphs and compute the loss function to bound the interleaving distance.

Ishika Ghosh
Michigan State University
ghoshis3@msu.edu

Erin Chambers
Saint Louis University
erin.chambers@slu.edu

Elizabeth Munch, Sarah Percival
Michigan State University
muncheli@msu.edu, perciva9@msu.edu

Bei Wang
University of Utah
beiwang@sci.utah.edu

PP5

Better Than Best Low-Rank Approximation with the Singular Value Decomposition

The Eckhart-Young theorem states that the best low-rank approximation of a matrix can be constructed from the leading singular values and vectors of the matrix. Here, we illustrate that the practical implications of this result crucially depend on the organization of the matrix data. In particular, we will show examples where a rank 2 approximation of the matrix data in a different representation more accurately represents the entire matrix than a rank 5 approximation of the original matrix data – even though both approximations have the same number of underlying parameters. Beyond images, we show examples of how flexible orientation enables better approximation of time series data, which suggests additional applicability of the findings. Finally, we conclude with a theoretical result that the effect of data organization can result in an unbounded improvement to the matrix approximation factor as the matrix dimension grows.

David Gleich
Purdue University
dgleich@purdue.edu

PP5

MS84: Federated Learning with Heterogenous Private Data Silos

Biomedical data, such as electronic medical records, is typically siloed within institutions and contains private information that limits sharing. Federated learning approaches incorporating privacy-preserving algorithms, such as differential privacy, provide an avenue to enabling the development of deep learning models from biomedical data. However, incorporating privacy mechanisms poses challenges to maintaining model accuracy, particularly in the case of underrepresented classes in heterogeneous data silos. We consider the effectiveness of centralized and distributed synthetic data generation approaches to addressing these challenges, using both naive and generative AI methods. We evaluate these approaches in the context of developing models with privacy-preserving federated learning on cancer surveillance datasets from the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) program.

John Gounley, Patrycja Krawczuk, Christopher Stanley, Heidi Hanson
Oak Ridge National Laboratory
gounleyjp@ornl.gov, krawczukpn@ornl.gov, stanleycb@ornl.gov, hansonha@ornl.gov

PP5

MS93: Learning Dynamical Systems with the Spectral Exterior Calculus

We introduce a data-driven framework for learning dynamical systems on Riemannian manifolds based on the spectral exterior calculus. Using eigenfunctions and eigenvalues of the Laplacian on smooth functions, approximated by the diffusion maps algorithm, the spectral exterior calculus represents vector fields as linear combinations of frame elements (i.e., elements of overcomplete bases), which act as generators of dynamical systems on the manifold of inter-

est. We present an implementation of this framework that consistently represents vector fields using Monte Carlo approximation from data points sampled on low-dimensional manifolds, such as the circle and 2-torus. In addition, we solve initial-value prediction problems using our vector field representations and compare the performance with solutions under the true systems.

Yanbing Gu
Dartmouth College
yanbing.gu.gr@dartmouth.edu

Suddhasattwa Das
Department of Mathematics & Statistics
Texas Tecb University
suddas@ttu.edu

Dimitrios Giannakis
Dartmouth College
Dimitrios.Giannakis@dartmouth.edu

Joanna M. Slawinska
Department of Mathematics
Dartmouth College
joanna.m.slawinska@dartmouth.edu

PP5

Nesterov Acceleration Despite Very Noisy Gradients

Momentum-based gradient descent methods use information gained along the trajectory, in addition to the local information from the gradient, in order to achieve an accelerated rate of convergence. These methods have been well-studied for convex optimization. Computing the gradient is often too expensive and it is approximated using stochastic gradient estimates in practice. However, theres a lack of theoretical analyses of accelerated methods in the setting of stochastic gradient descent, even for the simple case of convex functions. We address this gap with a novel descent algorithm (AGNES), which provably achieves acceleration for smooth convex minimization tasks with noisy gradient estimates if the noise intensity is proportional to the magnitude of the gradient. Nesterovs accelerated gradient descent does not converge under this noise model if the constant of proportionality exceeds one. AGNES fixes this deficiency and provably achieves an accelerated convergence rate no matter how small the signal to noise ratio in the gradient estimate. Empirically, we demonstrate that this is an appropriate model for mini-batch gradients in overparameterized deep learning. Finally, we show that AGNES outperforms stochastic gradient descent with momentum and Nesterovs method in the training of CNNs.

Kanan Gupta, Jonathan Siegel
Texas A&M University
kanan.g@pitt.edu, jwsiegel@tamu.edu

Stephan Wojtowytsch
University of Pittsburgh
s.woj@pitt.edu

PP5

MS93: A Reduced Operator Newton Method for Bayesian Filtering

Bayesian filtering is an iterative Bayesian inference problem, where the goal is to sample from the posterior distribution of the state of a dynamical system given past observations. A popular choice for high-dimensional Bayesian inference utilizes the concept of measure transport by finding through explicit parameterizations or nonparametrically through particle evolutions an invertible transformation of samples of a reference measure into those of the target. Despite much research effort into efficient transport-based sampling and inference in the setting where the score (gradient of log density) of the target is known, state-of-the-art methods still have an exponential scaling of the computational cost with dimension. In this work, we introduce a novel notion of low-dimensional structure that arises in the filtering problem when the underlying dynamics is chaotic, that of conditional absolute continuity. We exploit the absolute continuity of the target measures on low-dimensional manifolds and hyperbolicity of the dynamics to develop new algorithms to estimate a conditional score associated with the target. We then develop an operator-theoretic Newton-Raphson method that uses the learned scores to compute transport maps on low-dimensional manifolds. Both the intrinsic dimension reduction as well as the speed of convergence of Newton methods leads to potentially scalable solutions to filtering in high-dimensional chaotic systems, such as in geophysical applications.

Anant Gupta, Nisha Chandramoorthy
Georgia Institute of Technology
agupta886@gatech.edu, nishac@uchicago.edu

PP5

MS95: SE(3) Synchronization by Dual Quaternions

In synchronization problems, the goal is to estimate elements of a group from noisy measurements of their ratios. A popular estimation method for synchronization is the spectral method. It extracts the group elements from eigenvectors of a block matrix formed from the measurements. The eigenvectors must be projected, or "rounded", onto the group. The rounding procedures are constructed ad hoc and increasingly so when applied to synchronization problems over non-compact groups. We developed a spectral approach to synchronization over the non-compact group SE(3), the group of rigid motions of $\mathbb{R}^3$. We based our method on embedding SE(3) into the algebra of dual quaternions, which has deep algebraic connections with the group SE(3). These connections suggest a natural rounding procedure considerably more straightforward than the current state-of-the-art for spectral SE(3) synchronization, which uses a matrix embedding of SE(3). We show by numerical experiments that our approach yields comparable results to the current state-of-the-art in SE(3) synchronization via the spectral method. Thus, our approach reaps the benefits of the dual quaternion embedding of SE(3), while yielding estimators of similar quality.

Ido Hadi
Tel Aviv University
idohadi@mail.tau.ac.il

PP5

MS86: Medical Image Classification by Graph Clustering Methods Applied to a Neural Network Approach

We are given a set of medical image data and would like to classify these images. There are many neural network algorithms and topological machine learning algorithms to do the image classification. However, this problem can be formulated as a graph clustering problem. The vertices of

a graph are images and the vertices within the same cluster can be assumed to share similar features and properties, thus making the applications of graph clustering techniques very useful for image classification. We shall first use a box spline based wavelet-framelet method to clean the images and help building up the adjacency matrix for the given image data. Recently, the approach based on sparse solutions of linear systems for graph clustering finds clusters more efficient than the traditional spectral clustering method. We propose to use the two newly developed graph clustering methods based on sparse solution methods for linear system for image classification. The performance of our graph clustering methods will be shown to be very effective to classify an image. For the focus of this presentation, we also found ways to incorporate these clustering methods into a convolutional neural network (CNN) to perform more detailed image data classification. Additionally, the images have been pre-processed for these techniques in ways that have not been commonly applied in CNNs of this type. Numerical results will be demonstrated.

Jackson Hamel
University of Georgia department of mathematics
Research under Dr. Ming-Jun Lai
jrhamel314@uga.edu

PP5

MS90: Geometric Deep Learning for Unregistered Shape Data

We will present various geometric deep frameworks designed to generate feature vector representations of 3D geometric data. Specifically, we will delve into techniques that seek to produce feature vectors invariant to parameterization, ensuring that different observations of the same shape (viewed as different parameterizations of the geometric data) are mapped to identical feature vectors. Further we will demonstrate the utility of these feature vectors in a variety of downstream learning tasks such as classification, registration, and shape reconstruction.

Emmanuel Hartman
University of Houston
ehartman@alumni.unca.edu

Martin Bauer
Florida State University
bauer@math.fsu.edu

Nicolas Charon
University of Houston
ncharon@central.uh.edu

PP5

Forecasting Solar-Geomagnetic Activity Using Ai: A Markov Chain Monte Carlo Approach

The compatibility of solar and geomagnetic activity is calculated using the Markov chain Monte Carlo (MCMC) technique, an application of artificial intelligence (AI). The study covers various sunspots durations (1749-2017, 1932-2017, & 2009-2017) and related geomagnetic (Ap) durations (1932-2017) by employing a Markov 4-states transition matrix. The classification of the matrix is determined by the maximum to minimum range of each cycle in the selected data sets. This AI-driven analysis establishes a behavioral relationship and draws samples from complex distributions of sunspots and Ap index using the MCMC technique. The expected number of sunspots and geomagnetic activity for each state remains almost constant, demonstrating the AI's ability to analyze and predict patterns in complex datasets. Significant results of predicted frequencies during 20, 40, and 80 years are aligned with observed frequencies of sunspot and Ap index, indicating the AI's effectiveness in predictive modeling. Based on model-generated results, the study suggests that the future behavior of the studied parameters will follow a similar progression, showcasing the application of AI in forecasting natural phenomena. Additionally, the effects of the sun on geomagnetic index are confirmed, and the frequency for upcoming cycles can be predicted by this AI-driven study.

Danish Hassan
National Textile University Karachi Campus Pakistan
danish10ansari@gmail.com

PP5

Predicting Scaling Laws with Approximation and Statistical Theory for Transformer Neural Networks on Intrinsically Low-Dimensional Data

During training, transformer generalization error can be empirically predicted as a power-law in terms of the number of training samples and the network size. We establish approximation and statistical theory for transformer neural networks that relates the exponent of the power-law to the intrinsic dimension of the data manifold.

Alexander Havrilla
Georgia Tch
ahavrilla3@gatech.edu

Wenjing Liao
Georgia Institute of Technology
wliao60@gatech.edu

PP5

MS94: Tensor Kernel in Causal Inference

This study introduces a novel approach to causal inference by applying tensor decomposition kernels within Support Vector Machines (SVMs) for synthetic control. Traditional synthetic control methods often struggle with high-dimensional and complex datasets. Our approach uses tensor decomposition to enhance SVM's ability to capture multi-dimensional structural relationships, improving the accuracy of treatment effect estimates and enabling the detection of interaction effects across multiple variables. We validated our method through experiments with both simulated and real-world datasets, demonstrating superior precision in causal estimates compared to traditional methods. The results confirm the effectiveness of tensor decomposition kernels in handling complex causal scenarios, making this approach a valuable tool for researchers in various fields such as economics and epidemiology. This study advances the application of machine learning techniques in causal inference, particularly in settings involving intricate variable interactions.

Huan He
University of Pennsylvania
huan.he@pennmedicine.upenn.edu

Tianshi Xu
Emory University
tianshi.xu@emory.edu

Yong Chen

University of Pennsylvania
ychen123@upenn.edu

PP5

Continuous Partitioning for Graph-Based Semi-Supervised Learning

Laplace learning algorithms for graph-based semi-supervised learning have been shown to be degenerate in the low label-regime and thus yield poor performance at low label rates and in imbalanced class settings. Furthermore, the performance of these algorithms is sensitive to post-hoc decision rules that map continuous predictions to labels. We propose a framework for graph-based semi-supervised learning based on continuous nonconvex quadratic programming which provably obtains integer solutions. Our approach, CutSSL, is motivated by an exact quadratic relaxation of a cardinality-constrained minimum-cut graph partitioning problem, a nondegenerate problem at low label rates. Furthermore, we show our formulation is related to an optimization problem whose approximate solution is the mean-shifted Laplace learning heuristic, thus providing new insight into the improved performance of this heuristic over Laplace learning at low label rates. We introduce a scalable algorithm based on ADMM and demonstrate that CutSSL surpasses the current state-of-the-art in node classification tasks on k-nearest neighbor graphs derived from classic image classification benchmarks and large real-world citation and product networks across a variety of label rates, class imbalance, and label imbalance regimes.

Chester Holtz
University of California San Diego
chholtz@ucsd.edu

Pengwen Chen
National Chung Hsing University
pengwen@email.nchu.edu.tw

Zhengchao Wan, Chung-Kuan Cheng
University of California, San Diego
z5wan@ucsd.edu, ckcheng@ucsd.edu

Gal Mishne
UC San Diego
gmishne@ucsd.edu

PP5

MS96: Simple Arithmetic Operation in Latent Space Can Generate a Novel Three Dimensional Graph Metamaterials

Recent advancements in AI-based design strategies for metamaterials have revolutionized the creation of customizable architectures across nano- to macro-scale dimensions, achieving mechanical behaviors surpassing the inherent properties of constituent materials. However, these methods' growing complexity challenges the generation of diverse metamaterials without significant human and computational resources, hindering adoption. Addressing this, our study introduces a design strategy generating three-dimensional graph metamaterials using simple arithmetic operations in latent space. By integrating hidden representations of disentangled latent space and latent diffusion processes, our approach comprehensively understands complex design spaces, generating diverse graph metamaterials. This versatile methodology creates structures from repetitive lattice structures to functionally graded mechanical metamaterials and serves as an inverse design strategy for diverse lattice structures, including crystalline and trabecular bone structures. This foundational step advances comprehension of the intricate latent design space, potentially establishing a unified model for traditional generative models in mechanical metamaterials.

YoungJoon Hong
KAIST
hongyj@kaist.ac.kr

PP5

Suboptimality Bounds for Trace-Bounded Sdps Enable a Faster and Scalable Low-Rank Sdp Solver Sdplr+

Semidefinite programs (SDPs) and their solvers are powerful tools with many applications in machine learning and data science. Designing scalable SDP solvers is challenging because by standard the positive semidefinite decision variable is an $n \times n$ dense matrix, even though the input is often $n \times n$ sparse matrices. However, the information in the solution may not correspond to a full-rank dense matrix as shown by Barvinok and Pataki. Two decades ago, Burer and Monteiro developed an SDP solver `SDPLR` that optimizes over a low-rank factorization instead of the full matrix. This greatly decreases the storage cost and works well for many problems. The original solver `SDPLR` tracks only the primal infeasibility of the solution, limiting the technique's flexibility to produce moderate accuracy solutions. We use a suboptimality bound for trace-bounded SDP problems that enables us to track the progress better and perform early termination. We then develop `SDPLR+`, which starts the optimization with an extremely low-rank factorization and dynamically updates the rank based on the primal infeasibility and suboptimality. This further speeds up the computation and saves the storage cost. Numerical experiments on Max Cut, Minimum Bisection, Cut Norm, and Lovsz Theta problems with many recent memory-efficient scalable SDP solvers demonstrate its scalability up to problems with million-by-million decision variables and it is often the fastest solver to a moderate accuracy of 10^{-2}.

Yufan Huang, David Gleich
Purdue University
huan1754@purdue.edu, dgleich@purdue.edu

PP5

Looking Beyond Geometric and Probability-Based Distances for Multiclass Classification in Network Communication Systems

Uncertainty quantification and outlier detection in input datasets is fundamental to improve the efficiency of the multiclass classification problem. Anomaly detection in network communication systems uses multiclass classification algorithms extensively. Outliers and anomalies present themselves as noise in training data and hence can have a disproportionate impact on accuracy. At the same time, if the outliers (or noise) follow the same probability distribution in both train and test datasets, they can go undetected while still impacting the accuracy. Both Geometric distances and Probability distribution-based distances have been used to detect outliers in the network communication datasets-each with its own merits and drawbacks. In this work, we study three distinct classes of distance measures for outlier/uncertainty detection, namely, (1) Geometric,

(2) Probability distribution-based, and (3) Optimal Transport (OT)-based distance called Wasserstein distance. Our problem is set in the anomaly and intrusion detection area of network communication systems. We employ each of these distance measures as cost functions in multiclass classification and clustering problem, aimed at detecting anomalies. We also investigate a combined formulation of the distance measures and present use cases where they may outperform individual distance measures.

Venkat Sai Suman Lamba Karanam
University of Nebraska-Lincoln
saisuman5312@gmail.com

Zahmeeth S. Sakkaff
University of Nebraska, Lincoln
sszahmeeth@gmail.com

PP5

MS95: Generic Orbit Recovery from Invariants of Very Low Degree

Motivated by the multi-reference alignment problem and questions in equivariant neural networks we study the problem of recovering the generic orbit in a unitary representation of a compact group from invariant tensors of degree three or less. We explore the similarities and differences between the descriptive power of low degree polynomial and unitary invariant tensors and provide some evidence that in many cases of interest they have similar descriptive power. In particular we prove that for the regular representation of a finite group polynomial invariants of degree at most three separate generic orbits. We also investigate these questions for subregular representations of finite groups and prove that for the defining representation of the dihedral group, polynomial invariants of degree three and less separate generic orbits.

Joshua Katz
University of Missouri-Columbia
jkgbh@missouri.edu

Dan Edidin
Department of Mathematics
University of Missouri
edidind@missouri.edu

PP5

Parametric Kernel Low-Rank Approximations Using Tensor Train

Kernel matrices are prevalent in applied mathematics with applications in scientific computing (integral equations) and data science (Gaussian processes). However, kernel matrices arising from practical applications are often dense, large, and depend on specific hyperparameters; thus, there is a need for methods to compute and store low-rank approximations of parametric kernel matrices efficiently. We propose a method divided into two phases, an offline phase and an online phase, where the offline phase dominates the computational cost, and the online phase is relatively inexpensive. During the offline phase, we leverage tensorized multi-variate Chebyshev function approximation and the Tensor Train (TT) decomposition of the coefficient tensor, efficiently computed via TT cross approximation. During the online phase, we instantiate a kernel matrix for a particular parameter and compute and store its low-rank approximation. Our method has linear complexity in terms of the size of the kernel matrix, and its utility and efficiency are demonstrated by applying it to various kernels arising from different application areas, varying spatial configurations of the source and target points, and varying properties of the kernel matrix (symmetry and positive semidefinite).

Abraham W. Khan
North Carolina State University, Mathematics Department
awkhan3@ncsu.edu

Arvind Saibaba
North Carolina State University
asaibab@ncsu.edu

PP5

MS94: Efficient Training of Gaussian Processes with Tensor Product Structure

Gaussian processes with the covariance matrix given as the sum of possibly multiple Kronecker products appear in spatio-temporal magnetoencephalography or climate data sets. This structure allows the covariance matrix to be identified as a tensor, which is used to represent this operator and the training data in the tensor train format. Determining the optimal set of hyperparameters of a Gaussian process based on a large amount of training data requires both linear system solves and trace estimation. In particular, solving a linear system with the covariance tensor is a major bottleneck and requires appropriate preconditioning.

Josie König
University of Potsdam
josie.koenig@uni-potsdam.de

Martin Stoll
Chemnitz University of Technology
martin.stoll@mathematik.tu-chemnitz.de

Max Pfeffer
University of Göttingen
m.pfeffer@math.uni-goettingen.de

PP5

Data Assimilation Beyond the Linear and Local Gaussian Setting

In the filtering setting, we seek to infer the state distribution conditioned on all the observations available up to that time. Ensemble filters sequentially assimilate observations by updating a set of samples over time. Despite their empirical success, Gaussian filters such as the ensemble Kalman filter (Evensen, J. Geophys. Res., 1994) operate under simplifying assumptions: linearity of the observation model and Gaussianity of the forward process and observation error. Also it is often assumed that the observations to assimilate are local functions of the state, i.e, that the observations only depend on the state variables that are close by in physical distance. These assumptions are limiting and not representative of many practical settings, thus leading to fundamentally biased inference. Indeed, we often deal with non-Gaussian distributions as well as nonlocal observations given by integrals of linear and nonlinear functions of the state, such as radiance measured by satellites, fluxes through surfaces, or solutions of elliptic partial differential equations. Leveraging tools from measure transport (Spantini et al., SIAM Review, 2022), we introduce an unifying framework to analyze ensemble filters. Building on the flexibility of this formulation, we present

our recent advances to build ensemble filters able to leverage nonlinear, non-local, and non-Gaussian features of the filtering problems of interest.

Mathieu Le Provost, Jan Glaubitz, Youssef M. Marzouk
Massachusetts Institute of Technology
n/a, Glaubitz@mit.edu, ymarz@mit.edu

PP5

MS89: The Expressive Power of Low-Rank Adaptation

Low-Rank Adaptation (LoRA), a parameter-efficient fine-tuning method that leverages low-rank adaptation of weight matrices, has emerged as a prevalent technique for fine-tuning pre-trained models such as large language models and diffusion models. Despite its huge success in practice, the theoretical underpinnings of LoRA have largely remained unexplored. This paper takes the first step to bridge this gap by theoretically analyzing the expressive power of LoRA. We prove that, for fully connected neural networks, LoRA can adapt any model f to accurately represent any smaller target model $\overline{f}$ if LoRA-rank $\geq$ (width of f) $\times \frac{\text{depth of } \overline{f}}{\text{depth of } f}$, under a mild assumption. We quantify the approximation error when the LoRA-rank is lower than the threshold. For Transformer networks, we show any model can be adapted to a target model of the same size with rank-($\frac{\text{embedding size}}{2}$) LoRA adapters. Our study reveals numerous theoretical insights on hyperparameter tuning and algorithm development for LoRA, all of which are empirically validated.

Kangwook Lee, Yuchen Zeng
University of WisconsinMadison
kangwook.lee@wisc.edu, yzeng58@wisc.edu

PP5

MS89: Transformers Are Uninterpretable with Myopic Methods: a Case Study with Bounded Dyck Grammars

Interpretability methods aim to understand the algorithm implemented by a trained model (e.g., a Transofmer) by examining various aspects of the model, such as the weight matrices or the attention patterns. In this work, through a combination of theoretical results and carefully controlled experiments on synthetic data, we take a critical view of methods that exclusively focus on individual parts of the model, rather than consider the network as a whole. We consider a simple synthetic setup of learning a (bounded) Dyck language. Theoretically, we show that the set of models that (exactly or approximately) solve this task satisfy a structural characterization derived from ideas in formal languages (the pumping lemma). We use this characterization to show that the set of optima is qualitatively rich; in particular, the attention pattern of a single layer can be "nearly randomized', while preserving the functionality of the network. We also show via extensive experiments that these constructions are not merely a theoretical artifact: even after severely constraining the architecture of the model, vastly different solutions can be reached via standard training. Thus, interpretability claims based on inspecting individual heads or weight matrices in the Transformer can be misleading.

Yuchen Li
Carnegie Mellon University
yuchenl4@cs.cmu.edu

Kaiyue Wen
Tsinghua University
wenky20@mails.tsinghua.edu.cn

Bingbin Liu, Andrej Risteski
Carnegie Mellon University
bingbinl@cs.cmu.edu, aristesk@andrew.cmu.edu

PP5

Energy Dissipation Rate-Based Sampling: An Adaptive Refinement Strategy for Solving Thermodynamically Consistent Models Using Pinns

We introduce Energy Dissipation Rate-based Sampling (EDRS), an innovative adaptive refinement technique for Physics-Informed Neural Networks (PINNs) aimed at significantly enhancing their efficacy in tackling thermodynamically consistent models. The core is the strategic utilization of energy dissipation rate density as a key metric for selectively resampling critical collocation points, thereby refining the model's accuracy. EDRS notably outperforms the traditional residual-based adaptive refinement approach, demonstrating a sixfold improvement in relative mean square error for the Allen-Cahn equation. Leveraging the inherent mesh-free nature of PINNs, we deploy neural networks to adeptly approximate the solutions of thermodynamically consistent phase field models with dynamic boundary conditions in arbitrary domains. Our primary aim is to meticulously investigate how dynamic boundary conditions influence the overall dynamics of the bulk material. We conduct thorough simulations across both disk and ellipse-shaped domains, analyzing how static versus dynamic boundary conditions distinctly affect the system's behavior. This study not only underscores the potential of EDRS in enhancing the computational performance of PINNs but also enriches our comprehension of the pivotal role dynamic boundary conditions play in influencing the dynamic behavior of thermodynamic systems, offering valuable insights for future studies in computational physics and engineering.

Chunyan Li
The Pennsylvania State University
cvl6109@psu.edu

Wenkai Yu
Beijing Computational Science Research Center
yuwenkai@csrc.ac.cn

Qi Wang
University of South Carolina
qwang@math.sc.edu

PP5

Filtrated Reinforcement Learning for Infinite-Dimensional Systems Using Moment Kernelization

Despite the prosperity of research into reinforcement learning (RL), methods for designing optimal control policies for high-dimensional deterministic systems remain underdeveloped. The primary challenge lies in the curse of dimensionality, which results in high computational costs and low learning accuracy. To fill this literature gap, in this work, we study reinforcement learning problems for control systems defined on an infinite-dimensional function space. This class of systems, referred to as the ensemble system, arises in diverse scientific and application domains, ranging from quantum science and robotics to neuroscience. We

develop a moment kernel transform, which maps an ensemble system to an associated moment system defined on a reproducing kernel Hilbert space (RKHS) consisting of infinite sequences. The RKHS structure enables the approximation of an ensemble system using a finite-dimensional truncated moment system. We further reveal a filtration structure formed by a sequence of truncated moment systems with increasing truncation orders. Using this structure, we show that the sequences of control policies and value functions learned from the sequence of systems converge to those of the ensemble system. This gives rise to a filtrated RL algorithm with a convergence guarantee. The computational efficiency and performance of the proposed algorithm are demonstrated using emerging applications in quantum science and robotics.

Jr-Shin Li, Wei Zhang
Washington University in St. Louis
jsli@wustl.edu, wei.zhang@wustl.edu

PP5

MS78: Real Circles Tangent to Three Conics

In this paper we study circles tangent to conics. We show there are generically 184 complex circles tangent to three conics in the plane and we characterize the real discriminant of the corresponding polynomial system. We give an explicit example of 3 conics with 136 real circles tangent to them. We conjecture that 136 is the maximal number of real circles. Furthermore, we implement a hill-climbing algorithm to find instances of conics with many real circles, and we introduce a machine learning model that, given three real conics, predicts the number of circles tangent to these three conics.

Julia Lindberg
University of Texas-Austin
julia.lindberg@math.utexas.edu

PP5

MS83: Hyperparameter Tuning via Trajectory Predictions: Stochastic Prox-linear Methods in Matrix Sensing

Motivated by the desire to understand stochastic algorithms for nonconvex optimization that are robust to their hyperparameter choices, we analyze a mini-batched prox-linear iterative algorithm for the problem of recovering an unknown rank-1 matrix from rank-1 Gaussian measurements corrupted by noise. We derive a deterministic recursion that predicts the error of this method and show, using a non-asymptotic framework, that this prediction is accurate for any batch-size and a large range of step-sizes. In particular, our analysis reveals that this method, though stochastic, converges linearly from a local initialization with a fixed step-size to a statistical error floor. Our analysis also exposes how the batch-size, step-size, and noise level affect the (linear) convergence rate and the eventual statistical estimation error, and we demonstrate how to use our deterministic predictions to perform hyperparameter tuning (e.g. step-size and batch-size selection) without ever running the method. On a technical level, our analysis is enabled in part by showing that the fluctuations of the empirical iterates around our deterministic predictions scale with the error of the previous iterate.

Mengqi Lou
Georgia Institute of Technology
mlou30@gatech.edu

PP5

MS83: On Averaging and Extrapolation for Gradient Descent

This work considers the effect of averaging, and more generally extrapolation, of the iterates of gradient descent in smooth convex optimization. After running the method, rather than reporting the final iterate, one can report either a convex combination of the iterates (averaging) or a generic combination of the iterates (extrapolation). For several common stepsize sequences, including recently developed accelerated periodically long stepsize schemes, we show averaging cannot improve gradient descent's worst-case performance and is, in fact, strictly worse than simply returning the last iterate. In contrast, we prove a conceptually simple and computationally cheap extrapolation scheme strictly improves the worst-case convergence rate: when initialized at the origin, reporting $(1+1/\sqrt{16Nlog(N)})x_N$ rather than xN improves the best possible worst-case performance by the same amount as conducting $O(\sqrt{N/log(N)})$ more gradient steps. Our analysis and characterizations of the best-possible convergence guarantees are computer-aided, using performance estimation problems. Numerically, we find similar (small) benefits from such simple extrapolation for a range of gradient methods.

Alan Luner
Johns Hopkins
aluner1@jhu.edu

PP5

MS82: Deep Optimal Control Approach to Path Estimation of Nonlinear Filtering

We developed a novel deep learning based numerical solution of the nonlinear filtering problem. The approach begins by establishing an estimation measure via a controlled Markov process and then measuring the discrepancy between the posterior distribution of the nonlinear filter and the controlled measure using Kullback-Leibler (KL) divergence. This framework leads to a variational problem for addressing the nonlinear filtering. Practically, the variational objective can often be articulated as the sum of a running cost and a terminal cost with respect to the controlled Markov process. To solve this reformulated stochastic control problem, we introduce a training scheme for Deep FBSDE algorithm, which utilizes artificial neural networks to derive the optimal controller. We numerically demonstrate the scalability and efficiency of our method in addressing the nonlinear filtering problem across both finite and infinite state space signal processes.

George Yin, Xiaohang Ma
University of Connecticut, Storrs
gyin@uconn.edu, xiaohang.ma@uconn.edu

Qing Zhang
University of Georgia
Department of Mathematics
qz@uga.edu

PP5

Neural Networks for Nonlinear Dynamics: The Im-

portance of Initial Transients

The numerical simulation of high-dimensional nonlinear systems often demands extensive computational resources, spanning hours, days, or even weeks. To alleviate this computational burden, we may construct a reduced-order model capable of describing the essential input-output behavior of the original system. One technique for reduced-order modeling involves approximating each chart of the system's slow manifold with an autoencoder neural network. Subsequently, the dynamics on each chart's image are approximated by a neural ordinary differential equation. Once trained, the reduced-order model becomes operational. However, this method relies solely on post-transient data, limiting its effectiveness in modeling trajectories that originate off the slow manifold. To overcome this limitation, we propose an extended learning curriculum that incorporates a new training stage capable of modeling fast transients. Initially, we select a sampled trajectory point near the slow manifold. Then, utilizing that point as an initial condition, we simulate the reduced-order model backward in time. Finally, we utilize the backward time predictions as targets for a neural network responsible for encoding fast transients. To illustrate our approach, we construct two-dimensional reduced-order models for the complex GinzburgLandau equation in its supercritical regime and a Kolmogorov flow with a traveling wave solution.

Gregory Macchio
Princeton University
gm0796@princeton.edu

Clarence Rowley
Princeton University
Department of Mechanical and Aerospace Engineering
cwrowley@princeton.edu

PP5

Spatiotemporal Causal Inference with Mechanistic Ecological Models: Evaluating Targeted Culling on Chronic Wasting Disease Dynamics in Cervids

Spatiotemporal (ST) causal inference is needed to detect the effect of interventions on indirectly measured epidemiological outcomes going beyond studying ST correlations. Chronic wasting disease (CWD) causes neurological degeneration and death to white-tailed deer in Wisconsin. Targeted culling (TC) involves removing deer after traditional hunting seasons in high CWD prevalence areas. Evaluating the causal effect of TC in the spread and growth of CWD is an unresolved research and CWD management question that can guide surveillance efforts. Reaction-diffusion partial differential equations (PDEs) can mechanistically model underlying ST dynamics of wildlife diseases allowing us to make inference in unobserved epidemiological quantities. These models indirectly regress ST covariates on diffusion and growth rates parameterizing such PDEs obtaining associational conclusions. In this work, we develop an innovative method to obtain causal estimators for TC interventions on CWD epidemiological processes using inverse-probability-of-treatment-weighting by means of marginal structural models embedded in the PDE fitting process. Moreover we establish a novel scheme for sensitivity analysis under unmeasured confounder for testing the significance of the effect on indirectly measured epidemiological outcomes. Our methods can be used to study ST interventions in epidemiological evolution of infectious diseases helping to mitigate public health implications and disease burden

Juan Francisco Mandujano Reyes
University of Wisconsin-Madison
mandujanorey@wisc.edu

Ting Fung Ma
Department of Statistics
University of South Carolina
tingfung@mailbox.sc.edu

Ian P. McGahan
Department of Statistics
University of Wisconsin - Madison
imcgahan@wisc.edu

DANIEL J. Storm
Wisconsin Department of Natural Resources
danielj.storm@wisconsin.gov

DANIEL P. Walsh
U.S. Geological Survey
Montana Cooperative Wildlife Research Unit
dwalsh@usgs.gov

JUN Zhu
Department of Statistics
University of Wisconsin - Madison
junzhu@wisc.edu

PP5

MS87: Trustworthy Scientific Inference from Generative Models

Scientific inference often involves using data to infer internal parameters that determine a complex physical phenomenon. This data usually comes in the form of a labeled set collected either i) from a mechanistic model that implicitly encodes the likelihood function, or more generally from a statistical model that cannot be evaluated; or ii) from observational studies where labels can be measured with high precision. Both of these are Likelihood-Free Inference (LFI) settings, where scientists are increasingly leveraging machine learning methods, such as neural density estimators and generative models, to infer parameters of interest given a new observation from the same intractable likelihood model. However, high-posterior density (HPD) regions derived from neural density estimators do not necessarily have a high probability of including the true parameter of interest, even if the posterior is well-estimated and the labeled dataset has the same distribution as the target. In this talk, I will present new LFI methodology and algorithms to leverage neural density estimators and produce confidence regions of parameters of interest that have (i) nominal frequentist coverage for any value of the true unknown parameters, and (ii) small average area (yielding high constraining power) if the prior is well-specified. I will illustrate our methods on examples from astronomy and high-energy physics, and discuss where we stand and what challenges still remain.

Luca Masserano
Carnegie Mellon University

lmassera@andrew.cmu.edu

PP5

Machine Learning Competency Measures

Current approaches to machine model application involve use of model confidence in decision making. The goal of model competence estimation is to provide the end user with more than model confidence, and to calibrate trust in a specific decision with a metric for how likely it is that the input is truly outside the prediction space of the algorithm. Confidence estimation along with out of distribution indicators are both factored in to provide a measure of model prediction competence as an indicator of the predictive capability of a model for a current input.

Francesca Mcfadden
Johns Hopkins University
francesca.mcfadden@jhuapl.edu

PP5

A Hybrid of Half-Quadratic Minimization Algorithm and Method of K-means for Edge Detection of Texts

We propose an improved edge-detection method for text images. First, we formulate an optimization problem for image denoising. The minimization problem, which utilizes a nonlinear regularization term, is solved using a half-quadratic minimization technique. In this approach, we transform the nonlinear objective function into two simpler subproblems: one being a quadratic problem, while the other has an explicit solution. The optimality condition of the quadratic problem results in an elliptic partial differential equation, which is solved using a Finite Element Method. After denoising the image, we employ the method of K-means to separate the text from the image. The final step involves utilizing any edge detection algorithm to obtain the edges of the text images. We implement our proposed approach on images containing text with varying levels and types of noise and compare our results with other edge detection algorithms.

Renier Mendoza
Institute of Mathematics
University of the Philippines- Diliman
rmendoza@math.upd.edu.ph

Kristine Rey Recio, Rolly Czar Joseph Castillo
Institute of Mathematics
University of the Philippines - Diliman
kristinereyr@gmail.com, rtcastillo1@up.edu.ph

PP5

MS92: Dirichlet Active Learning: Exploration and Exploitation Guarantees

This work introduces Dirichlet Active Learning (DiAL), a Bayesian-inspired approach to the design of active learning algorithms. Our framework models feature-conditional class probabilities as a Dirichlet random field and lends observational strength between similar features in order to calibrate the random field. This random field can then be utilized in learning tasks: in particular, we can use current estimates of mean and variance to conduct classification and active learning in the context where labeled data is scarce. We demonstrate the applicability of this model to low-label rate graph learning by constructing "propagation operators' based upon the graph Laplacian, and offer computational studies demonstrating the method's competitiveness with the state of the art. Finally, we provide rigorous guarantees regarding the ability of this approach to ensure both exploration and exploitation, expressed respectively in terms of cluster exploration and increased attention to decision boundaries.

Kevin Miller
Department of Mathematics
Brigham Young University, Provo
ksmiller@math.byu.edu

Ryan Murray
North Carolina State University
rwmurray@ncsu.edu

PP5

MS90: Model-Constrained AutoEncoder Approach for Solving Forward and Inverse Problems

Obtaining fast and accurate solvers for partial differential equations (PDEs) and inverse problems is crucial across a broad spectrum of engineering disciplines. We introduce a novel methodology that employs a model-constrained autoencoder to learn both the forward surrogate and inverse mappings for PDE-constrained inverse problems simultaneously. This approach is particularly effective in low-data regimes, where purely data-driven machine-learning methods tend to overfit due to the absence of underlying physical laws. We validate the generalization, accuracy, and effectiveness of our model through its application to the 2D Poisson equation, 2D Burgers' equation, and 2D Navier-Stokes equation. Our results demonstrate substantial enhancements in computational efficiency and solution fidelity, marking a significant advancement in the field.

Hai Nguyen
The University of Texas at Austin
hainguyen@utexas.edu

PP5

MS93: Regularity and Gradients of Optimal Transportation Based Loss Functions

Optimal transportation distances become ever more popular in learning problems involving probability measures. In a typical setting, one has a loss function that measures an optimal transportation distance between a parametric model and data. Thus, gradient-based optimization methods require differentiability analysis and computations of gradients of such loss functions. Here, we provide such analysis and discuss gradient formulas. In particular, we find a new sufficient condition for the differentiability via the geometry of optimal transport plans and prove qualitative approximation results for the gradient formula.

Levon Nurbekyan
Department of Mathematics
Emory University
levon.nurbekyan@emory.edu

Yunan Yang
Cornell University
yunan.yang@cornell.edu

Elisa Negrini
University of California, Los Angeles
enegrini@math.ucla.edu

Robert Martin
U.S. Army Research Office
robert.s.martin163.civ@army.mil

Mirjeta Pasha
Massachusetts Institute of Technology
In Fall 2024 Virginia Tech
mpasha@mit.edu

PP5

MS93: Learning Statistically Accurate Chaotic Models with Neural ODEs

To address the inherent limitations of Neural Ordinary Differential Equations (Neural ODEs) such as instability or limited expressiveness, recent works have proposed various methods, including innovative adjoint techniques and approaches focusing on latent space representations. Nonetheless, can neural dynamical models learn the true dynamics and, in particular, reproduce associated statistical measures? Our initial findings indicate that the usual generalization error of Neural ODEs is not reflective of how well the governing dynamics is learned. In this work, we define notions of generalization that are more appropriate for ergodic systems, such as ergodic chaotic systems and stochastic systems, utilizing metrics derived from dynamical systems theory. Additionally, we introduce and analyze a regularization technique aimed at enabling Neural ODEs to learn invariant measures. Our analysis provides sufficient conditions for when neural network parameterizations can accurately learn statistical properties of dynamical systems. Our empirical results and analysis encompass a wide variety of ergodic systems and neural network parameterizations, including via Fourier neural layers, ResNets and RNNs. By combining shadowing theory from dynamical systems with statistical learning, our work advances the foundations of effective neural network modeling of complex dynamics.

Jeongjin Park
Georgia Institute of Technology
Computational Science and Engineering Department
jpark3141@gatech.edu

Nicole Yang
Emory University
tianjiao.yang@emory.edu

Nisha Chandramoorthy
Georgia Institute of Technology
nishac@uchicago.edu

PP5

MS87: Optimal Ridge Regularization for Out-of-Distribution Prediction

We study the behavior of optimal ridge regularization and optimal ridge risk for out-of-distribution prediction, where the test distribution deviates arbitrarily from the train distribution. We establish general conditions that determine the sign of the optimal regularization level under covariate and regression shifts. These conditions capture the alignment between the covariance and signal structures in the train and test data and reveal stark differences compared to the in-distribution setting. For example, a negative regularization level can be optimal under covariate shift or regression shift, even when the training features are isotropic or the design is underparameterized. Furthermore, we prove that the optimally tuned risk is monotonic in the data aspect ratio, even in the out-of-distribution setting and when optimizing over negative regularization levels. In general, our results do not make any modeling assumptions for the train or the test distributions, except for moment bounds, and allow for arbitrary shifts and the widest possible range of (negative) regularization levels.

Pratik Patil
Berkeley
pratikpatil@berkeley.edu

PP5

MS80: Learning Cryo-Em Structures Via the Method of Moments (Poster)

We extend our recent work by introducing deep neural network priors to learn the moment inversion map for cryo-EM. Our neural networks output the volumes along the distribution of rotations and shift variance, with moments being the input. We also generalize the neural network-based reconstruction algorithm to perform the reconstruction even with shifts in the data, which we demonstrate on simulated and biological volumes.

Sounak Paul
University of Chicago
paulsounak96@uchicago.edu

PP5

Normalizing Flows for Simulation-Based Inference of Non-Markovian and Multiscale Stochastic Processes

Inverse modeling of arbitrary stochastic processes remains an open challenge. When it is possible, inverse modeling often requires problem-specific tools due to incalculable or intractably costly likelihoods. Meanwhile, given a model and parameters, forward simulation is usually straightforward, computationally cheap, and often parallelizable. Recent developments in likelihood-free or simulation-based inference leverage neural network models, particularly normalizing flows, provide a method to efficiently approximate likelihoods based only on realizations of the underlying stochastic process (samplable from forward simulations). Here, we apply normalizing flows to address easy to simulate dynamics for which likelihood calculation has been previously challenging, including: non-Markovian protein production kinetics across cell division, and a large state space model of multi-stage gene expression. In both cases, we see that normalizing flows can accurately approximate likelihoods across different measurement ranges, making it ideal for multi-scale processes.

Pedro Pessoa, Max Schweiger, Steve Presse
Arizona State University
ppessoa@asu.edu, mcschwei@asu.edu, spresse@asu.edu

PP5

MS79: Hebbian Learning of Cyclic Features of Neural Code

Cyclic structures are a class of mesoscale features ubiquitous in both experimental stimuli and the activity of neural populations encoding them. Important examples include encoding of head direction, grid cells in spatial navigation, and orientation tuning in visual cortex. The central question of our present work is: how does the brain faithfully transmit cyclic structures between regions? Is this a

generic feature of neural circuits, or must this be learned? If so, how? While cyclic structures are difficult to detect and analyze with classical methods, tools from algebraic topology have proven to be particularly effective in understanding cyclic structures. Recently, work of Yoon et al. develops a topological framework to match cyclic coding patterns in distinct populations that encode the same information. We leverage this framework to show that, beginning with a random initialization, Hebbian learning robustly supports the propagation of cyclic structures through feedforward networks. Moreover, the efficacy of this learning mechanism can be modulated by the relative sizes of the input and output neural populations, suggesting that high-dimensional "unpackings" are necessary to propagate complex geometry in the context of naive neural connectivity. This is joint work with Chad Giusti and is supported by the Air Force Office of Scientific Research under award FA9550-21-1-0266.

Nikolas Schonsheck
University of Delaware
nischon@udel.edu

Chad Giusti
Oregon State University
chad.giusti@oregonstate.edu

PP5

MS84: Distributed Machine Learning Strategies Using Dynamical Low-Rank Approximation

Computational cost on client machines and communication latency are central challenges for federated learning, where the cost of communicating the whole weight matrices and training the entire model on resource-constraint edge devices is prohibitively expensive. This work introduces Federated dynamical low-rank training (FeDLRT), a technique for improving the computing and communication efficiency of federated learning in training neural networks across decentralized devices. Drawing from dynamical low-rank approximation, FeDLRT enables provably robust optimization using only a low-rank factorization on the clients and server of a federated learning setup. FeDLRT offers several key benefits. It focuses on automatic rank adaptation of the low-rank factors of the weight matrices, a feature that significantly reduces communication and computational costs while maintaining model performance. Additionally, it incorporates variance reduction techniques and provides theoretical proof of global loss descent in the federated learning setting. Numerical experiments further demonstrate the effectiveness of FeDLRT in enhancing the efficiency and scalability of federated learning for neural networks.

Steffen Schotthoefer, Paul Laiu
Oak Ridge National Laboratory
schotthofers@ornl.gov, laiump@ornl.gov

PP5

MS94: Acceleration Methods for Scientific and Data Science Applications

Recent years have seen increasing interest in general-purpose 'acceleration' methods that improve the convergence rate of fixed point iterations. Anderson Acceleration (AA) is a well-known example of such a method. In this poster, we introduce a new class of nonlinear acceleration algorithms based on extending conjugate residual-type procedures for linear equations. Our main algorithm exhibits similarities with both Anderson Acceleration and inexact Newton methods. Experimental results demonstrate the algorithm's effectiveness across diverse applications, including simulation and deep learning.

Mitchell Scott
Emory University
mitchell.scott@emory.edu

ziyuan tang
University of Minnesota
tang0389@umn.edu

Tianshi Xu
Emory University
tianshi.xu@emory.edu

Yuanzhe Xi
Department of Mathematics
Emory University
yuanzhe.xi@emory.edu

Yousef Saad
University of Minnesota
saad@umn.edu

PP5

Navigating the Exploitation-Exploration Challenge in Genetic Algorithms: A Novel Human-Centered Approach

Genetic Algorithms are potent tools for solving complex computational challenges.However, one persistent challenge for algorithm designers is striking the delicate balance between exploitation and exploration, essential for effective search navigation. In this presentation, a new interactive approach for handling the exploitation-exploration trade-off dynamics within the search process of Genetic algorithms is proposed.Unlike the conventional approach, the search process will not be compromised of a single-phase nor the decision-maker tuning efforts will be distributed among the algorithms traditional parameters such as defining new crossover and mutation operators internal to the algorithm to influence its search navigation. Instead, a human-centered two-phase search process, compromised of a global search phase followed by a local search phase will be utilized. In this framework, the designer plays the central role in directing the algorithm's search navigation through the focused tuning efforts of a new search space size control parameter external to the algorithm which proves itself to be the most dominant parameter in-effect to the algorithms navigation process. The broad applicability of this approach holds the key to unlocking unexplored optimization capabilities across the class of evolutionary algorithms, suggesting a potentially transformative step in the field. We demonstrate the power of this new approach on well-known benchmark problems in optimization.

Ehsan Shams
Alexandria University
ehsan.ali@alexu.edu.eg

PP5

MS95: Refining Particle Picking in Cryo-Em

Single-particle cryo-electron microscopy (cryo-EM) is a powerful technique for determining high-resolution protein structures in their near-native states. In cryo-EM,

the experimental data consists of projection images, often called micrographs, each containing numerous individual protein projections known as particles. Due to the low signal-to-noise ratio (SNR) typically exhibited by micrographs, many particles are required to achieve a three-dimensional (3D) protein reconstruction at near-atomic resolution. Identifying the tomographic projections contained in each micrograph is known as particle picking. Many semi-automatic and automatic methods have been developed to address this problem. In this poster, we propose adding a new step to the computational pipeline, which refines the output of a particle picker and facilitates the conventional following steps of 2D classification and 3D recovery by centering the picked images and eliminating outliers. The poster is based on joint work with Ayelet Heimowitz (Ariel University) and Lev Kapnulin (Tel Aviv University).

Nir Sharon
Tel Aviv University
nsharon@tauex.tau.ac.il

PP5

MS90: Efficient Clustering on Riemannian Manifolds Using Fréchet Embeddings

Symmetric Positive Definite (SPD) matrices, in particular correlation matrices, appear in various applications from applied mathematics and engineering, most notably in neuro-imaging applications such as Diffusion Tensor Imaging and Functional Magnetic Resonance Imaging, where they are employed to model the strength of neural connections between different brain sites. Mathematically, the set of SPD matrices is not a vector subspace of the Euclidean space under standard matrix addition and scalar multiplication, however it possesses a smooth manifold structure that can be endowed with a Riemannian metric. Hence the proper measure of similarity between SPD matrices is not an Euclidean distance but a Riemannian distance that can capture the intrinsic geometrical structure of the underlying space. Unfortunately, computation of distances in the Riemannian manifold of SPD matrices is rather expensive with the result that it is exceedingly expensive to compute k-means of SPD matrices directly in the Riemannian setting. Here we present and demonstrate a novel approach to efficiently cluster data on the space of SPD matrices taking advantage of a specially designed Frchet embedding.

Robert Azencott, Nicolas Charon, Demetrio Labate
University of Houston
robertazencott@gmail.com, ncharon@central.uh.edu,
dlabate@math.uh.edu

Andreas Mang
University of Houston
Department of Mathematics
amang@Central.UH.EDU

Ji Shi
University of Houston
shijibest@gmail.com

PP5

MS89: Learning Associative Memories with Gradient Descent

This work focuses on the training dynamics of one associative memory module storing outer products of token embeddings. We reduce this problem to the study of a system of particles, which interact according to properties of the data distribution and correlations between embeddings. Through theory and experiments, we provide several insights. In overparameterized regimes, we obtain logarithmic growth of the "classification margins.' Yet, we show that imbalance in token frequencies and memory interferences due to correlated embeddings lead to oscillatory transitory regimes. The oscillations are more pronounced with large step sizes, which can create benign loss spikes, although these learning rates speed up the dynamics and accelerate the asymptotic convergence. In underparameterized regimes, we illustrate how the cross-entropy loss can lead to suboptimal memorization schemes. Finally, we assess the validity of our findings on small Transformer models. Associated minisymposium: MS0 Mathematical Principles in Foundation Models

Berfin Simsek
NYU
Flatiron CCM
simsekberfin95@gmail.com

Vivien Cabannes
Meta
vivien.cabannes@gmail.com

Alberto Bietti
New York University
alberto.bietti@gmail.com

PP5

MS90: Sparsity Promoting Eeg Source Localization with Time and Spatial Regularization

Source localization is crucial for interpreting electroencephalograms and requires solving large and extremely ill-posed inverse problems. As standard approaches are intractable for such large dynamic datasets at scale, we propose a source localization procedure utilizing sparsity promoting spatial and temporal regularization. We propose using computationally efficient variable projected augmented Lagrangian methods for its solution. In a simulation study, we confirm that our approach can successfully locate regions of brain activity for inverse problems with 11,500 data points but 3,719,100 unknowns.

Jack Michael Solomon
Emory University
jack.michael.solomon@emory.edu

Matthias Chung
Emory University, U.S.
matthias.chung@emory.edu

Rosemary A. Renaut
Arizona State University
School of Mathematical and Statistical Sciences
renaut@asu.edu

PP5

Learning Selection Strategies in Buchberger's Algorithm

Studying the set of exact solutions of a system of polynomial equations largely depends on a single iterative algorithm, known as Buchberger's algorithm. Optimized versions of this algorithm are crucial for many computer algebra systems (e.g., Mathematica, Maple, Sage). We in-

troduce a new approach to Buchberger's algorithm that uses reinforcement learning agents to perform S-pair selection, a key step in the algorithm. We then study how the diffculty of the problem depends on the choices of domain and distribution of polynomials, about which little is known. Finally, we train a policy model using proximal policy optimization (PPO) to learn S-pair selection strategies for random systems of binomial equations. In certain domains, the trained model outperforms state-of-the-art selection heuristics in total number of polynomial additions performed, which provides a proof-of-concept that recent developments in machine learning have the potential to improve performance of algorithms in symbolic computation.

Michael Stillman
Cornell University
mike@math.cornell.edu

PP5

Binary Malware Attribution Using Llm Embeddings and Topological Descriptors

In malware authorship attribution, there's a scarcity of easily-obtained, expressive, stylistically salient features. One has nothing to go off of except the binary malware executable; this has traditionally constrained static analysis to producing simple statistics on the decompiled binary. Dynamic analysis can obtain more descriptive features by running the malware executable in a sandbox, but this is both resource-intensive and has obvious risks. In this talk, we introduce two new sources of expressive static features which, when combined, achieve classification accuracies competitive with leading dynamic analysis models. Our features stem from a hypothesis that coding style is found at or around the level of the function. To capture features, we leverage Large Language Model (LLM) embeddings, persistent homology, and graph neural networks (GNNs) to identify stylistically-salient geometric features of the program architecture. We demonstrate that this new paradigm of graph learning combined with LLM code-embeddings can easily be extended to encompass and enhance existing featurizations.

Branden Stone, Kincaid MacDonald, Ajai Ruparelia
Georgia Tech Research Institute
branden.stone@gtri.gatech.edu,
kincaid.macdonald@gtri.gatech.edu,
ajai.ruparelia@gtri.gatech.edu

Boomer Rogers
Georgia Tech
herman.rogers@gatech.edu

Adonis Bovell
Georgia Tech Research Institute
adonis.bovell@gtri.gatech.edu

PP5

MS80: Phase Recovery with Moment Constraints for Multireference Alignment

Multireference alignment (MRA) is a computational model used to recover a signal that has undergone random circular translations and subjected to noise. This problem is commonly addressed through two types of methods. Techniques such as template alignment focus on estimating the unknown shifts of observed signals, allowing for the reconstruction of the original signal while accounting for these translations. Alternatively, methods such as expectationmaximization (EM) and the method of moments estimate the signal without explicitly determining the unknown shifts. In our approach, we define a loss function by averaging the distance of a template to the aligned signals. Crucially, the loss function is constrained to templates whose power spectrum matches an empirical estimate. We then use a gradient-based approach to estimate the true signal. Our method offers increased efficiency compared to EM and demonstrates improved accuracy over method of moments applied to the bispectrum.

Emanuel Ström
KTH (Royal Institute of Technology, Stockholm)
emastr@kth.se

Vahid Shahverdi
KTH Royal Institute of Technology
vahidsha@kth.se

PP5

MS95: Moment Invariant Distribution Learning

We consider the scenario where data is collected from M sub-populations to produce M data batches, but the sample sizes $\{n_i\}_{i=1}^M$ of the batches are small, i.e. $n_i \ll M$. We assume that local environmental factors affect only the first and second moments of the sub-populations, so that the i^{th} batch consists of n_i independent, identically distributed observations from $f_{\sigma_i,\mu_i}(x) = \frac{1}{\sigma_i} f\left(\frac{x-\mu_i}{\sigma_i}\right)$ for some universal distribution f. Our goal is to reliably recover the underlying distribution f by aggregating the sparse data. This is a highly relevant problem, as it allows for reliable nonparametric estimates of the density f in settings where traditional approaches fail, and more broadly leads to more precise comparison across sub-populations in settings where little data is available. We approach this problem with tools from signal processing for solving multireference alignment models with dilation, including Fouier moments and unbiasing procedures.

Anna Little, Mikhail Sweeney
University of Utah
little@math.utah.edu, sweeney@math.utah.edu

PP5

MS86: Modified Nystrom Approximation Via Low-Rank Matrix Sensing

We study a modified version of the Nystrom problem in which the $(1, 2)$ block of the matrix is incomplete. We demonstrate how this modified version is equivalent to a low-rank matrix sensing problem.

Abiy F. Tasissa
Tufts University
Abiy.Tasissa@tufts.edu

PP5

MS86: Graph Clustering Techniques with Sparse Solution Methods and Application to Medical Image Classification

Given medical image datasets, we are interested in classifying the images for potential application in disease diagnosis. There are many neural network models and topological machine learning algorithms to do the image classification. However, this problem can also be formulated as a graph clustering problem. The vertices of a graph

are images and the vertices within the same cluster can be assumed to share similar features and properties, thus making the applications of graph clustering techniques very useful for image classification. We shall first use a box spline based wavelet-framelet method to clean the images and help building up the adjacency matrix for the given image data. Finding clusters is one of major research interests in graph analysis. Recent approach based on sparse solutions of linear systems for graph clustering finds clusters more efficiently than the traditional spectral clustering method. We thus propose to use the two newly developed graph clustering methods based on sparse solution methods for linear system for image classification. The performance of our graph clustering methods will be shown to be very effective to classify an image. Numerical results will be demonstrated in the end of the paper.

Ye Tian, Ming-Jun Lai
University of Georgia
yt95681@uga.edu, mjlai@uga.edu

PP5

An Optimal Weighted Least Squares Framework for Operator Learning

We consider the problem of learning an operator $K : L^2_\nu \to L^2_\eta$ from a finite set of M observations $g^i = K(f^i)$, where the f^i are independent samples drawn from a tailored probability measure μ on L^2_ν. For a fixed probability measure ρ on L^2_ν, the admissible class of operators, $\mathcal{B}_\rho$, is taken to be the Hilbert space of Borel-measurable maps $A : (L^2_\nu, \mathcal{F}, \rho) \to (L^2_\eta, \mathcal{B}(L^2_\eta))$ satisfying $\|A\|^2_{\mathcal{B}_\rho} = \int_{L^2_\nu} \|A(f)\|^2_{L^2_\eta} \rho(df) < \infty$, and the reconstruction error is measured in this norm. Given an N dimensional subspace $V_N \subset \mathcal{B}_\rho$, we establish probabilistic accuracy and stability results for general weighted least squares approximations in V_N, and we show that there exists a sampling measure μ and weight w for which optimal stability and accuracy can be achieved for a given sample size M. We provide an explicit optimal sampling measure when ρ is Gaussian and V_N comprises both linear and non-linear operators. We highlight the effectiveness of this method in several numerical experiments: these include learning the PDE solution operator to Darcy flow, Naiver-Stokes, and Burgers equations. Our framework offers similar accuracy to Fourier Neural Operators as well as several advantages such as interpretability, *a priori* error estimation with optimal sample efficiency, and simplicity of implementation.

John M. Turnage
University of Utah
turnage@math.utah.edu

PP5

MS96: Autoequivariant Network Search Via Group Decomposition (Poster)

Recent works show that group equivariance as an inductive bias improves neural network performance for both classification and generation. However, designing group-equivariant neural networks is challenging when the group of interest is large and is unknown. Moreover, inducing equivariance can significantly reduce the number of independent parameters in a network with fixed feature size, affecting its overall performance. We address these problems by proving a new group-theoretic result in the context of equivariant neural networks that shows that a network is equivariant to a large group if and only if it is equivariant to smaller groups from which it is constructed. Using this result, we design a novel fast group equivariant construction algorithm, and a deep Q-learning-based search algorithm in a reduced search space, yielding what we call autoequivariant networks (AENs). AENs find the right balance between equivariance and network size when tested on new benchmark datasets, G-MNIST and G-Fashion-MNIST, obtained via group transformations on MNIST and Fashion-MNIST respectively that we release. Extending these results to group convolutional neural networks, where we optimize between equivariances, augmentations, and network sizes, we find group equivariance to be the most dominating factor in all high-performing GCNNs on several datasets like CIFAR10, SVHN, RotMNIST, ASL, EMNIST, and KMNIST.

Lav Varshney
University of Illinois at Urbana-Champaign
varshney@illinois.edu

PP5

MS85: Neural Networks for Inference in Optimal Control Governed by the FitzHugh-Nagumo Model

We investigate the use of neural networks (NN) for the estimation of hidden model parameters and uncertainty quantification from noisy observational data for inverse parameter estimation problems governed by a system of nonlinear ordinary differential equations (ODEs). The underlying ODE is the FitzHugh-Nagumo model. The considered problem exhibits significant mathematical and computational challenges for classical parameter estimation methods, including strong non-linearities, non-convexity, and sharp gradients of the optimization landscape. We explore the use of NNs to approximate the reconstruction maps for parameter estimation from observational data. The considered data are time series of the spiking membrane potential of a biological neuron. We infer parameters controlling the dynamics of the model, noise parameters of autocorrelated additive noise and noise modelled via stochastic differential equations, as well as the covariance matrix of the posterior distribution to expose uncertainties. Our approach is motivated by formulating parameter estimation as a Bayesian inverse problem. We report results for different NN architectures and study the influence of noise on predication accuracy. We also report results for training NNs on dedicated hardware. Our results demonstrate the NNs are a promising tool to estimate parameters of the dynamical system, stochastic processes, as well as uncertainties as they propagate though our system.

German Villalobos
Department of Mathematics
University of Houston
gvillalobos@uh.edu

Johann Rudi
Virginia Tech
jrudi@vt.edu

Andreas Mang
University of Houston
Department of Mathematics
amang@Central.UH.EDU

PP5

MS81: Optimal Transport for Single-Cell Hetero-

geneous Data Analysis

Advances in single-cell technologies enable comprehensive studies of heterogeneous cell populations that make up tissues, the dynamics of developmental processes, and the underlying regulatory mechanisms that control cellular functions. The computational integration of single-cell datasets is drawing heavy attention toward making advancements in machine learning and data science. Optimal transport (OT) is a powerful tool in the analysis of complex data, as it learns an optimal cost-effective mapping between data distributions. In this poster, I will report our recent work on developing OT-based data analysis methods for single-cell multi-omics integration and dynamic inference of time series single-cell data.

Lin Wan
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
lwan@amss.ac.cn

PP5

MS96: Signal Processing with Implicit Neural Fields: Editing, Manipulation, and Understanding (Poster)

Implicit Neural Representations (INRs) use multi-layer perceptrons to encode continuous multi-media data, showing promise in various computer vision tasks. However, editing and processing INRs remain challenging since signals are represented by neural network parameters. Existing methods manipulate these continuous representations by processing their discretized instances, losing INR's compactness and continuity. This work explores directly modifying INRs without explicit decoding by proposing an implicit neural signal processing network, INSP-Net, using differential operators on INRs. The key insight is that spatial gradients of neural networks can be computed analytically and are translation-invariant. Additionally, any continuous convolution filter can be approximated by a linear combination of high-order differential operators. INSP-Net applies signal processing operators as weighted compositions of computational graphs corresponding to INRs' high-order derivatives, with data-driven learned weighting parameters. Building on INSP-Net, we introduce INSP-ConvNet, a Convolutional Neural Network that implicitly operates on INRs. Experiments demonstrate the effectiveness of INSP-Net and INSP-ConvNet in low-level image and geometry processing tasks (e.g., blurring, deblurring, denoising, inpainting, smoothening) and high-level tasks on implicit fields like image classification

Peihao Wang
University of Texas at Austin
peihaowang@utexas.edu

PP5

MS79: Characterizing Lymphedema Risk Using MRI Topological Data Analysis

Breast cancer-related lymphedema (BCRL) is the most common post-treatment complication affecting breast cancer survivors, occurring in 20% of survivors. BCRL is a result of lymph fluid backup due to the damage or removal of lymph nodes during breast cancer treatment. This chronic condition can negatively impact patients quality of life due to its intense pain and reduction in mobility. While risk factors for developing this condition have been previously identified, these methods remain insufficient for predicting individualized patient risk. To improve these methods, we leverage topological data analysis to identify anatomical variations in the breast tissue leading to increased lymphedema risk using preoperative MRIs in a primary cohort of 364 breast cancer patients. The resulting model seeks to improve lymphedema risk scoring and identify regions of interest to guide clinical decision-making. Our results are benchmarked against both traditional machine learning (incorporating demographic and treatment information) and deep learning models (incorporating MRI data). With improved BCRL prediction, patients at the greatest risk of developing this condition can be identified as candidates for preventative surgeries, such as lymphovenous bypass surgery, which have historically not been cost-effective to implement widely.

Jeremy Watts
University of Tennessee
gkm819@vols.utk.edu

Vasileios Maroulas
University of Tennessee, Knoxville
vmaroula@utk.edu

Anahita Khojandi
University of Tennessee
khojandi@utk.edu

Stefanos Boukovalas, Devin Clegg
University of Tennessee Medical Center
sboukovalas@utmck.edu, dclegg1@utmck.edu

PP5

MS92: Fast, Effective Data Reduction Through Novelty Sampling

We present a new algorithm for reducing a large data set to a small number of landmark data points. The landmarks are randomly selected, yet they account for nearly all the "novelty" in the data. To generate landmarks, we randomly propose data points and accept/reject with probabilities depending on the previous selections. After the generation step, the landmarks can be used to quickly make predictions and find clusters in the data. Landmark-based learning has a memory footprint which is independent of the data size, so the approach is suitable for distilling large data sets with $N >= 10^9$ data points.

Robert Webber
California Institute of Technology
rwebber@ucsd.edu

PP5

Interpretable Function Approximation Using Random Fourier Features and ANOVA-Boosting

We study the problem of scattered-data approximation, where we have given sample points and the corresponding function evaluations. In the random Fourier feature approach, we draw frequencies at random and learn coefficients from the given data to construct the approximant. We use the classical analysis of variance (ANOVA) decomposition for approximating high-dimensional functions of low effective dimension. Thereby we give a relation between the Fourier transform of the function and the ANOVA terms. In the case for dependent input variables, the ANOVA decomposition is generalized with the aim to detect the structure of the function, i.e. to find which input variables and variable interactions are important. This in-

formation is then used to boost random Fourier feature algorithms. Furthermore, we generalize already existing random Fourier feature models to an ANOVA setting, where terms of different order can be used. Our algorithms have the advantage of interpretability, meaning that the influence of every input variable is known, even for dependent input variables. We give theoretical as well as numerical results that our algorithms perform well for sensitivity analysis. The ANOVA-boosting step reduces the approximation error of existing methods significantly.

Laura Weidensager, Daniel Potts
Chemnitz University of Technology
laura.weidensager@math.tu-chemnitz.de,
daniel.potts@math.tu-chemnitz.de

PP5

MS90: Biophysics-Based Data Assimilation of Longitudinal Tau and Amyloid-β Pet Scans

Misfolded tau and amyloid-β (Aβ) are hallmark proteins of Alzheimer's Disease (AD). Interpreting and combining this data beyond statistical correlations remains a challenge. Biophysical models offer a complementary avenue to assimilating such complex data. To this end, we introduce a mathematical model that tracks the dynamics of four species (normal and abnormal tau and Aβ) and uses a graph to approximate their spatial coupling. The graph nodes represent gray matter regions of interest (ROI), and the edges represent tractography-based connectivity between ROIs. We model interspecies interactions, migration, proliferation, and clearance. Our biophysical model has seven unknown scalar parameters plus unknown initial conditions for tau and Aβ. Using imaging MRI, tau and Aβ scans, we can calibrate these parameters by solving an inverse problem. We propose an inversion algorithm that stably reconstructs the unknown parameters. We verify and test its numerical stability in the presence of noise using synthetic data. We discovered that the inversion is more stable when using multiple scans. Finally, we apply the overall methodology on 334 subjects from the ADNI dataset and compare it to a commonly used tau-only model calibrated by a single PET scan. We report the R^2 and relative fitting error metrics. The proposed method achieves $R^2 = 0.82$ compared to $R^2 = 0.64$ of the tau-only single-scan reconstruction.

Zheyu Wen, Ali Ghafouri
University of Texas at Austin
zheyw@utexas.edu, ghafouri@utexas.edu

George Biros
University of Texas at Austin
Oden Institute
biros@oden.utexas.edu

PP5

MS78: Geometry of Polynomial Neural Networks

We study the expressivity and learning process for polynomial neural networks (PNNs) with monomial activation functions. The weights of the network parametrize the neuromanifold. In this paper, we study certain neuromanifolds using tools from algebraic geometry: we give explicit descriptions as semialgebraic sets and characterize their Zariski closures, called neurovarieties. We study their dimension and associate an algebraic degree, the learning degree, to the neurovariety. The dimension serves as a geometric measure for the expressivity of the network, the learning degree is a measure for the complexity of training the network and provides upper bounds on the number of learnable functions. These theoretical results are accompanied with experiments.

Jiayi Li
UCLA
jiayi.li@g.ucla.edu

Maximilian Wiesman
MPI MiS Leipzig
Germany
wiesmann@mis.mpg.de

Kaie Kubjas
Aalto University
kaie.kubjas@aalto.fi

PP5

MS85: Reservoir Computing for System Identification and Predictive Control with Limited Data

Model predictive control (MPC) is an industry standard control technique. MPC functions by iteratively solving an open-loop optimization problem to guide a system towards a desired state or trajectory. Consequently, an accurate forward model of system dynamics is critical for the efficacy of MPC and much recent work has been aimed at the use of neural networks to act as data-driven surrogate models to enable MPC. Perhaps the most common network architecture applied to this task is the recurrent neural network (RNN) due to its natural interpretation as a dynamical system. In this work, we assess the ability of RNN variants to both learn the dynamics of exemplar control systems and serve as surrogate models for MPC. We find that echo state networks (ESNs) have a variety of benefits over competing architectures, namely reductions in computational complexity, longer valid prediction times, and reductions in cost of the MPC objective function.

Jan P. Williams
Department of Mechanical Engineering
University of Washington
jmpw1@uw.edu

J. Nathan Kutz
University of Washington, Seattle
Dept of Applied Mathematics
kutz@uw.edu

Krithika Manohar
University of Washington
kmanohar@uw.edu

PP5

Restnet: A Residual Multi-Transformation Network Architecture

Residual networks (ResNets) allowed the use of deeper networks ameliorating the performance degradation problem of deep neural networks. ResNets create a residual connection consisting of a copy of the input data which is added posteriorly in the network after skipping several layers. In addition to branching out a copy and skipping layers, the ResNeXt technique introduces a multi-branch architecture adding the dimension of "cardinality" to artificial network designs. However, cardinality only refer to branching out several network blocks of similar structure. Initially, block

similarity allows for block concatenation increasing the efficiency of the ResNeXt. In this work we discuss breaking block similarity by introducing different types of non-linear activation functions for each branch. We call this method ResTNet. Adding an extra branch to ResNet-18 creates a ResTNet18 structure with small complexity increment in between ResNet18 and ResNet34. Classification loss improves drastically just by introducing a single extra branch compared to corresponding ResNet structure. We compare ResNet and ResTNet by training on satellite imagery from the publicly available Rare Planes Dataset. This dataset consists of ground images taken from a Maxar-5 satellite, 18k annotations and 20 different classes. Distribution Statement A: Approved for Public Release. Distribution is Unlimited. Public Affairs Release Approval # AFRL-2024-2314.

Victor M. Vergara
Blue Halo LLC
victor.vergara@bluehalo.com

Evan Kain
Air Force Research Laboratory
evan.kain.1@spaceforce.mil

Jeremy Wojcik
Blue Halo LLC
jeremy.wojcik@bluehalo.com

PP5

MS81: CTRhythm: Accurate Atrial Fibrillation Detection From Single-lead Rhythm by Convolutional Neural Network and Transformer Integration

Atrial Fibrillation (AF), a common supraventricular arrhythmia that affects about 30 million people globally. Electrocardiogram (ECG) analysis serves as a common diagnostic approach. Widespread adoption of wearable devices monitoring heart rhythm prompted the development of AF detection models for single-lead ECGs. Current state-of-the-art methods for AF detection, such as convolutional neural network (CNN) and convolutional recurrent neuralnetwork (CRNN) based models, focus on capturing local patterns only, despite the fact heart rhythms exhibit rich long-range dependencies. To address this limitation, we propose a novel method for single-lead ECG rhythm classification, termed CNN-Transformer Rhythm Classifier (CTRhythm), which integratesCNN with a Transformer encoder to effectively captureboth local and global patterns. CTRhythm achieved an overall F1 score of 0.8303 and an area under the Receiv-er Operating Characteristic Curve (AUROC) at 0.952, outperforming the baseline deep learning models on the golden standard CINC2017 dataset. In two external validation datasets, CTRhythm achieved overall F1scores of 0.929 and 0.932, respectively, showing its strong generalization capabilities. CTRhythm is freely available at https://github.com/labxscut/CTRhythm.

Li C. Xia
South China University of Technology
lcx.scut@outlook.com

PP5

Leveraging Neural Fields for Geophysical Inverse Problems

The recent surge in test time learning (TTL) has garnered substantial attention from researchers, particularly in the context of incorporating machine learning algorithms into the inversion process. The deep image prior (DIP) method and coordinate-based representations (e.g., neural fields) have shown that neural networks (NN), without any prior learning, can produce good inversion results. In this work, we will discuss the progress in utilizing neural fields in the geophysical inverse problems. Neural fields use neural networks to map a coordinate to the corresponding physical property value at that coordinate. We formulate the inverse problem in terms of the NN weights, which allows us to take advantage of searching over the high-dimensional space. Furthermore, parameterizing the inverse problems in a continuous setting naturally introduces smoothing effects. We demonstrate the use of neural fields in seismic tomography inversions and direct current resistivity inversions. The results show that this TTL approach can eliminate unwanted artifacts in the recovered subsurface physical property model caused by the sensitivity of the survey and physics. We also find that our results are better than the conventional inversion results in some cases in terms of the recovery of the boundaries and physical property values of the main targets. Our work illustrates that the inductive bias brought by neural fields can be beneficial in geophysical inversion.

Anran Xu, Lindsey Heagy
University of British Columbia
anranxu@student.ubc.ca, lheagy@eoas.ubc.ca

PP5

MS94: Hpytorch: Data-Driven Hierarchical Matrix-Based Gaussian Process

Gaussian processes (GPs) are a powerful tool for many applications, however, their performance heavily relies on selecting suitable hyperparameters. In this poster, we introduce HPyTorch, a scalable preconditioned GP package that addresses the hyperparameter selection problem while harnessing hardware acceleration capabilities through OpenMP and CUDA. HPyTorch is designed to exhibit robust performance across a wide range of hyperparameters, owing to its utilization of data-driven matrix routines that dynamically adjust components in response to the spectrum decay of the kernel matrices. Moreover, when employing gradient-based optimization methods, our package eliminates the need for auto-differentiation, further enhancing its computational efficiency. We evaluate the performance of HPyTorch on several real-world datasets, demonstrating that it outperforms state-of-the-art packages in terms of both accuracy and efficiency.

Tianshi Xu
Emory University
xuxx1180@gmail.com

Hua Huang
Georgia Institute of Technology
huangh223@gatech.edu

Edmond Chow
School of Computational Science and Engineering
Georgia Institute of Technology
echow@cc.gatech.edu

Yuanzhe Xi
Department of Mathematics
Emory University

yuanzhe.xi@emory.edu

PP5

MS87: Bayesian Constraint Relaxation: Distance and Divergence Penalization

We consider regularizing the squared distance to set-based constraints for several statistical tasks that can be cast as constrained optimization. These distance-to-set penalties are more flexible than many existing algebraic and regularization penalties, and often avoid drawbacks that arise from alternatives such as shrinkage. Moreover, these translate naturally to a flexible class of priors for incorporating relaxed constraints within a Bayesian framework. We show how they are amenable to gradient-based samplers, and derive a natural extension to (Bregman) divergence-to-set priors. We discuss several examples showcasing how this generalization improves performance by making use of information geometry of the data generating mechanism.

Jason Xu
Duke University
jason.q.xu@duke.edu

PP5

Distributed Deep Learning Optimization Using Second-Order Methods with Dynamic Weight Sharing and Hessian Variance Reduction

In recent years, deep learning has proven to be the most successful tool across various domains, such as computer vision and natural language processing, and the scalability of these models under growing data demands has become crucial. The increasing complexity of deep learning models and the exponential growth of data necessitates the distributed systems for training while managing increased computational loads. Nevertheless, this distributed approach faces many challenges. Among these are hardware limitations, computational overheads, and communication management. Our research introduces a method employing second-order optimization methods for distributed learning environments with asynchronous updates across worker and master nodes. We integrate dynamic weighting schemes that adjust updates when a worker node fails, alongside leveraging data overlap techniques to reduce variance for Hessian diagonal approximation.

Yuesheng Xu
Lehigh University
yux220@lehigh.edu

Arielle K. Carr
Lehigh University, U.S.
arg318@lehigh.edu

PP5

MS96: Generative Downscaling of PDE Solvers with Physics-Guided Diffusion Models (Poster)

Solving partial differential equations (PDEs) on fine spatio-temporal scales for high-fidelity solutions is critical for numerous scientific breakthroughs. Yet, this process can be prohibitively expensive, owing to the inherent complexities of the problems, including nonlinearity and multiscale phenomena. To speed up large-scale computations, a process known as downscaling is employed, which generates high-fidelity approximate solutions from their low-fidelity counterparts. In this paper, we propose a novel Physics-Guided Diffusion Model (PGDM) for downscaling. Our model, initially trained on a dataset comprising low-and-high-fidelity paired solutions across coarse and fine scales, generates new high-fidelity approximations from any new low-fidelity inputs. These outputs are subsequently refined through fine-tuning, aimed at minimizing the physical discrepancies as defined by the discretized PDEs at the finer scale. We evaluate and benchmark our model's performance against other downscaling baselines in three categories of nonlinear PDEs. Our numerical experiments demonstrate that our model not only outperforms the baselines but also achieves a computational acceleration exceeding tenfold, while maintaining the same level of accuracy as the conventional fine-scale solvers.

Wuzhe Xu
University of Massachusetts Amherst
wuzhexu@umass.edu

PP5

All Models Are Wrong But a Set of Them Is Useful

In real-world scenarios, prediction performance often relies on a single, complex model that sacrifices interpretability. However, a groundbreaking concept called the Rashomon set promotes the idea of using multiple equally effective models instead. The Sparse Wrapper Algorithm (SWAG) is a cutting-edge approach that blends screening and wrapper techniques to create a set of low-dimensional models with generalizable predictive capabilities. SWAG operates through a forward step process, where users select a modeling mechanism and the algorithm evaluates low-dimensional models. It then systematically builds larger models based on the best-performing ones from previous steps, resulting in a collection of models termed 'SWAG models.' This method's flexibility empowers practitioners to choose models that align with their specific needs or domain expertise without compromising accuracy. Furthermore, SWAG effectively addresses common data challenges such as missing values, outliers, and collinearity. The benefits of SWAG extend to decision-makers in various fields like genomics, engineering, and neurology. By constructing a network that illuminates attribute interactions, SWAG provides a deeper and more insightful perspective for informed decision-making.

Yagmur Yavuz Ozdemir
Auburn University
yzy0096@auburn.edu

Roberto Molinari
Auburn University
Department of Mathematics and Statistics
robmolinari@auburn.edu

Gaetan Bakalli
EM Lyon Business School
bakalli@em-lyon.com

Stéphane Guerrier, Samuel Orso, Cesare Miglioli
University of Geneva
stephane.guerrier@unige.ch, samuel.orso@unige.ch,
cesare.miglioli@unige.ch

Nabil Mili
University of Lausanne

nabil.mili@unil.ch

PP5

MS81: False Discovery Rate Control in Multiple Regression Models with Unobserved Confounders

False discovery rate (FDR) control in large-dimensional cross-sectional datasets remains challenging due to confounders, especially batch effects or unobserved variables in genome and brain imaging datasets. We utilized the factor model to characterize and eliminate these unobserved confounders, thus contributing to FDR control. This novel statistical framework involves two steps to eliminate the common dependence among variables caused by confounders. Firstly, we estimated the Negative Control (NC) set to address false negatives. Secondly, we extracted information on unobserved effects from the estimated NC set and constructed pairwise asymptotically independent test statistics. We provided theoretical proofs for both steps. Additionally, numerical experiments and theoretical derivations demonstrate that our method's false discovery proportion (FDP) estimations converge to the oracle values that utilize information from the unobserved variables, guaranteeing FDR control.

Taiyu Ye
University of Wisconsin-Madison
tye33@stat.wisc.edu

PP5

Finite Volume Property: Binary Compressed Sensing With Fourier Sampling Matrix

One-bit quantization with random dithering has recently found significant utilization potential in statistical signal processing applications due to its relatively low power consumption and low implementation cost. In addition to such advantages, an attractive feature of one-bit analog-to-digital converters (ADCs) is their superior sampling rates as compared to their conventional multi-bit counterparts. This characteristic endows one-bit signal processing frameworks with what one may refer to as sample abundance. Previous literature has demonstrated that theoretical guarantees for one-bit compressed sensing (CS) are attainable under certain statistical assumptions concerning the sensing matrix, such as the restricted isometry-type assumption. However, such guarantees are absent when considering the Fourier measurement matrix, which is more relevant in practical applications. In this paper, for the first time in the literature, we provide the theoretical guarantee for one-bit CS in the case of the Fourier measurement matrix by introducing the concept of Finite-volume property (FVP). As demonstrated, within the FVP framework, we can establish universal convergence for Fourier one-bit CS with a high probability.

Farhang Yeganegi, Arian Eamaz, Mojtaba Soltanalian
University of Illinois - Chicago, IL
University of Illinois - Chicago, IL
fyegan2@uic.edu, aeamaz2@uic.edu, msol@uic.edu

PP5

Neural Networks with Kernel-Weighted Corrective Residuals: A Mesh-Independent Approach for Inverse Design

Topology optimization is a challenging inverse problem as it is high-dimensional and usually constrained by partial differential equations (PDEs) and additional inequalities. Recently, Physics-Informed Neural Networks (PINNs) have been employed to simplify this process, yet they struggle to meet all design requirements and their effectiveness largely depends on the network configuration. To address these challenges, we leverage neural networks (NNs) with kernel-weighted Corrective Residuals (CoRes) for topology optimization. We have recently developed NN-CoRes to integrate the strengths of kernel methods and deep neural networks. Our method is inherently mesh-independent, completely avoiding domain discretization. In this presentation, we demonstrate how it significantly helps in (1) satisfying equality constraints in the design problem, (2) minimizing gray areas that are not favorable in real-world applications, and (3) simplifying the inverse design by reducing the sensitivity of neural networks to factors such as random initialization, architecture type, and choice of optimizer.

Amin Yousefpour, Carlos Mora, Shirin Hosseinmardi, Ramin Bostanabad
University of California, Irvine
yousefpo@uci.edu, cmorasar@uci.edu, shirinh1@uci.edu, raminb@uci.edu

PP5

A Convergent Interacting Particle Method for Computing KPP Front Speeds in Random Flows

This work aims to efficiently compute the spreading speeds of reaction-diffusion-advection fronts in divergence-free random flows under the Kolmogorov-Petrovsky-Piskunov nonlinearity. We develop a stochastic interacting particle method (IPM) for the reduced principal eigenvalue (Lyapunov exponent) problem of an associated linear advection-diffusion operator with spatially random coefficients. The Fourier representation of the random advection field and the Feynman-Kac formula of the principal eigenvalue (Lyapunov exponent) form the foundation of our method, which is implemented as a genetic evolution algorithm. The particles undergo advection-diffusion and mutation/selection through a fitness function originated in the Feynman-Kac semigroup. We analyze the convergence of the algorithm based on operator splitting and present numerical results on representative flows such as 2D cellular flow and 3D Arnold-Beltrami-Childress (ABC) flow under random perturbations. The 2D examples serve as a consistency check with semi-Lagrangian computation. The 3D results demonstrate that IPM, being mesh-free and self-adaptive, is easy to implement and efficient for computing front spreading speeds in the advection-dominated regime for high-dimensional random flows on unbounded domains where no truncation is needed.

Tan Zhang
Department of Mathematics, The University of Hong Kong
thta@connect.hku.hk

Zhongjian Wang
Nanyang Technological University
zhongjian.wang@ntu.edu.sg

Jack Xin
Department of Mathematics
UCI
jxin@math.uci.edu

Zhiwen Zhang

Department of Mathematics,
The University of Hong Kong,
zhangzw@hku.hk

PP5

Deep Neural Networks Learn Cellular Automaton Rules in Many-Valued Logic

We develop a theory characterizing the fundamental capability of deep neural networks to learn, from evolution traces, the logical rules governing the behavior of cellular automata (CA). This is accomplished by first establishing a novel connection between CA and Lukasiewicz propositional logic. While binary CA have been known for decades to essentially perform operations in Boolean logic, no such relationship exists for general CA. We demonstrate that many-valued (MV) logic, specifically Lukasiewicz propositional logic, constitutes a suitable language for characterizing general CA as logical machines. This is done by interpolating CA transition functions to continuous piecewise linear functions, which, by virtue of the McNaughton theorem, yield formulae in MV logic characterizing the CA. Recognizing that deep rectified linear unit (ReLU) networks realize continuous piecewise linear functions, it follows that these formulae are naturally extracted from CA evolution traces by deep ReLU networks. A corresponding algorithm together with a software implementation is provided. As the algorithm applies to networks with general, in particular also real-valued, weights, it can be used to extract logical formulae from deep ReLU networks trained on data.

Yani Zhang, Helmut Bölcskei
ETH Zurich
yanizhang@mins.ee.ethz.ch, hboelcskei@ethz.ch

PP5

On Consistency of Signatures Using Lasso

Signature transforms are iterated path integrals of continuous and discrete-time time series data, and their universal nonlinearity linearizes the problem of feature selection. This paper revisits some statistical properties of signature transform under stochastic integrals with a Lasso regression framework, both theoretically and numerically. Our study shows that, for processes and time series that are closer to Brownian motion or random walk with weaker inter-dimensional correlations, the Lasso regression is more consistent for their signatures defined by It integrals; for mean reverting processes and time series, their signatures defined by Stratonovich integrals have more consistency in the Lasso regression. We provide applications in option pricing. Our findings highlight the importance of choosing appropriate definitions of signatures and stochastic models in statistical inference and machine learning.

Ruixun Zhang
Peking University
zhangruixun@pku.edu.cn

Xin Guo
UC Berkeley
xinguo@ieor.berkeley.edu

Binnan Wang, Chaoyi Zhao
Peking University
3086185493@qq.com, zhaochaoyi@pku.edu.cn

PP5

MS91: Topological Analysis Reveals Organizational Routines Recurrence in Outpatient Medical Clinics During Covid-19

Organizational routines in outpatient medical clinics are repetitive, recognizable patterns of action that change over time. To visualize routine dynamics, we use a novel topological data analysis tool called the Temporal Mapper. We use time-stamped digital trace data to model routines as directed graphs that describe the state of a clinic on a particular day. Over the course of the COVID-19 pandemic, clinical routines varied dramatically. We use Temporal Mapper to identify and visualize stable states and recurrence of organizational routines during different phases of the COVID-19 pandemic, from January 2020 to December 2021. We found that some clinics bounced back to their pre-pandemic routines, while other clinics did not. These results demonstrate that topological data analysis (TDA) has the potential to enrich our understanding of routine dynamics and other recurrent processes of social organization.

Li Zhang
Michigan State University
zhan2070@msu.edu

PP5

Polynomial Lower Approximation for Two-Stage Stochastic Optimization

We introduce a two-phase approach to find global optimal solutions of two-stage stochastic programs with continuous decision variables and nonconvex recourse functions. The first phase involves the construction of a polynomial lower bound for the recourse function through a linear optimization problem over a nonnegative polynomial cone. Given the complex structure of this cone, we employ semidefinite relaxations with quadratic modules to facilitate our computations. In the second phase, we solve a surrogate first-stage problem by substituting the original recourse function with the polynomial lower approximation obtained in the first phase. Our method is particularly advantageous for two reasons: it not only generates global lower bounds for the nonconvex stochastic program, aiding in the verification of global optimality for prospective solutions like stationary solutions computed from other methods, but it also simplifies the computation of the expected value of the recourse function by using moments of random vectors. This makes our overall algorithm particularly suitable for the case where the random vector follows a continuous distribution or when dealing with a large number of scenarios. Numerical experiments are given to demonstrate the effectiveness of our proposed approach.

Suhan Zhong
University of California, San Diego
suzhong@tamu.edu

Ying Cui
UC Berkeley
yingcui@berkeley.edu

Jiawang Nie
University of California, San Diego

njw@math.ucsd.edu

zhu.3440@osu.edu

PP5

Sequential Quadratic Programming for Optimal Transport

The Monge optimal transport (OT) problem seeks to optimize transportation cost from a source probability measure to a target probability measure. The optimization is over a space of transport maps and the transportation cost is defined by a cost functional of the maps. As recent Interest arises in OT, many works focus on tackling the OT problem computationally using finite dimensional approximations. In this work, we present the infinite-dimensional formulation of the OT problem over a Banach space. We provide explicit expressions of the first and second-order variation of the objective functional, and of the function-form constraint with respect to the transport map. We propose a Sequential Quadratic Programming (SQP) framework for this infinite-dimensional problem. We show that subject to reasonable regularity assumptions, our framework satisfies known conditions for local convergence. Moreover, we demonstrate that a merit functional with sufficient constraint penalization serves as an effective step-size monitor within SQP iterations, leading to global convergence towards critical points. To the best of our knowledge, this is the first attempt at a globally convergent SQP operator recursion over infinite-dimensional spaces.

Zihe Zhou
Purdue University
School of Industrial Engineering
zhou408@purdue.edu

Harsha Honnappa, Raghu Pasupathy
Purdue University
honnappa@purdue.edu, pasupath@purdue.edu

PP5

MS89: Similarity of Layer-Wise Representations Within Transformers

Analyzing the similarity of internal representations within and across different models has been an important technique for understanding the behavior of deep neural networks. Most existing methods for analyzing the similarity between representations of high dimensions, such as the widely used centered kernel alignment (CKA) and those based on canonical correlation analysis (CCA), rely on statistical properties of the representations for a set of data points. In this paper, we focus on transformer models and study the similarity of representations between the hidden layers of individual transformers. In this context, we show that a simple sample-wise similarity metric is sufficient and aligns with complicated ones such as the CKA. Our experimental results on common transformers for both vision and NLP tasks also reveal that representations across layers are positively correlated, albeit the similarity decreases when layers are far apart. We then propose a simple training method to enhance the similarity between internal representations, with trained models that enjoy the following properties: (1) the last-layer classifier can be directly applied right after any hidden layers, yielding layer-wise accuracies much higher than those under standard training, (2) the layer-wise accuracies monotonically increase and reveal the minimal depth needed for the given task.

Zhihui Zhu
Ohio State University