siam 2022

Searchable Abstracts Document

Conference on Mathematics of Data Science (MDS22)

September 26–30, 2022

Town and Country Resort, San Diego, California, U.S.

This document was current as of September 14, 2022. 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

Welcome Remarks and Presentation: Learning to Optimize

"Learning to Optimize" or L2O is a method of generating or improving optimization algorithms. The resulting algorithms are often able to efficiently solve a set of target optimization problems. L2O has made gratifying progress in applications of signal processing, image processing, and other inverse problems, integer and combinatorial optimization, and optimal control, and it has achieved remarkable success in many individual types of problems. This talk introduces the background and motivation of L2O, and briefly outlines the different types of L2O approaches that have emerged recently, including deep-neural-network models, models based on traditional optimization methods, and various approaches to combining them. We will discuss how to train parameters and ensure correct convergence results.

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

IP2

Mathematics of Neural Networks in the Billion-parameter Age

The pace of progress of large-scale machine learning keeps increasing, towards even bigger models and datasets, producing astonishing results along the way in data-heavy domains such as text or images. Such rapid progress also leaves our mathematical understanding further behind, to the extent that one wonders whether it will ever catch up? In this talk, we will raise salient questions about this trend while zooming-in on technical snippets, covering approximation properties of transformers, mathematical aspects of score-based diffusion generative models, and optimization aspects of learning semi-parametric models.

Joan Bruna
New York University
Courant Institute
bruna@cims.nyu.edu

IP3

Remarks and Presentation: Over-parameterized Neural Networks and Causality in the Context of Drug Discovery and Single-cell Data Integration

Massive data collection holds the promise of a better understanding of complex phenomena and ultimately, of better decisions. An exciting opportunity in this regard stems from the growing availability of perturbation / intervention data (for example from drug/knockout screens in biology, advertisement, online education, etc.). In order to obtain mechanistic insights from such data, a major challenge is the development of a framework that integrates observational and interventional data and allows causal transportability, i.e., predicting the effect of unseen interventions or transporting the effect of interventions observed in one context to another. I will discuss how over-parameterized neural networks can be used for these problems. In particular, I will characterize the implicit bias of over-parameterized autoencoders and link this to causal transportability in the context of virtual drug screening.

Caroline Uhler
Massachusetts Institute of Technology
cuhler@mit.edu

IP4

A Mathematicians Perspective on How to Address Bias and Ethics in Data Science

Data science impacts many facets of our lives: from marketing to finance to voting to facial recognition to medical care. What happens if data science develops technology that amplifies societal biases and blatant racism? Talitha Washington will share her vision on how the mathematical community can collaboratively build industry relationships, develop diverse data science talent, and amplify ethics in data science.

Talitha Washington
Clark Atlanta University & Atlanta University Center
twashington@aucenter.edu

IP5

Remarks and Presentation - The Challenge of Understanding What Users Want: Inconsistent Preferences and Engagement Optimization

Online platforms have a wealth of data, run countless experiments and use industrial-scale algorithms to optimize user experience. Despite this, many users seem to regret the time they spend on these platforms. One possible explanation is that incentives are misaligned: platforms are not optimizing for user happiness. We suggest the problem runs deeper, transcending the specific incentives of any particular platform, and instead stems from a mistaken foundational assumption. To understand what users want, platforms look at what users do. This is a kind of revealed-preference assumption that is ubiquitous in user models. Yet research has demonstrated, and personal experience affirms, that we often make choices in the moment that are inconsistent with what we actually want: we can choose mindlessly or myopically, behaviors that feel entirely familiar on online platforms. In this work, we develop a model of media consumption where users have inconsistent preferences. We consider what happens when a platform that simply wants to maximize user utility is only able to observe behavioral data in the form of user engagement. Our framework is based on a stochastic model of user behavior, in which users are guided by two conflicting sets of preferences – one that operates impulsively in the moment, and the other of which makes plans over longer time-scales. By linking the behavior of this model to abstractions of platform design choices, we can develop a theoretical framework and vocabulary in which to explore interactions between design, behavioral science, and social media.

Jon M. Kleinberg
Cornell University
kleinber@cs.cornell.edu

IP6

Steins Method, Learning, and Inference

Steins method is a powerful tool from probability theory for bounding the distance between probability distributions. In this talk, Ill describe how this tool designed to prove central limit theorems can be adapted to assess and improve the quality of practical inference procedures. Along the way, Ill highlight applications to Markov chain Monte Carlo sampler selection, goodness-of-fit testing, generative modeling, de novo sampling, post-selection inference, dis-

tribution compression, bias correction, and nonconvex optimization, and Ill close with opportunities for future work.

Lester Mackey
Microsoft Research New England
lmackey@stanford.edu

IP7

Remarks and Presentation: Topological Methods to Analyze Complex Prediction Functions

It is now standard practice across science to study models that have been trained, fit, or learned based on a set of data. Many of these models involve a large number of parameters that make direct interpretation of the model challenging and a near black-box model view appropriate. We explore the possibilities of using ideas based on topological analysis methods to understand and evaluate these complex prediction functions. These show a surprising ability to generate easy to understand insights into these black boxes.

David F. Gleich
Purdue University
dgleich@purdue.edu

IP8

Geospatial Technologies for Ride-Sharing and Delivery Platforms

Ride-sharing and delivery platforms require complex geospatial inputs in order to generate their user experiences, match demand with drivers, and calculate fares. For example, route planning for meal deliveries uses predictions of the travel time between any two locations in the road network, and platform efficiency heavily depends on the accuracy of these predictions. I will describe the ML and optimization technologies, including those for travel time prediction, route optimization, and map error detection, that form the foundation of such multi-sided platforms. I will detail the challenges, such as data sparsity on parts of the road network, and show that highly accurate predictions need to take into account the granular dynamics of the physical system (traffic patterns in the road network). I will also illustrate machine learning architectures that incorporate this contextual information.

Dawn Woodard
LinkedIn
dawn@woodardscience.com

IP9

Closing Remarks and Presentation: Supervised Learning In Function Space

Supervised learning has shown enormous success as a data-driven methodology for function approximation between (finite dimensional) Euclidean spaces. However, for many problems arising in the sciences and engineering, it is often desirable to learn maps between (infinite dimensional) spaces of functions. The talk highlights recent progress in this area, explaining different approaches taken, describing numerical and theoretical results which study the complexity (cost versus accuracy) of these approaches, and showcasing their use in applications.

Andrew Stuart
Computing + Mathematical Sciences
California Institute of Technology
astuart@caltech.edu

SP1

SIAM Activity Group on Data Science Early Career Prize - When Will You Become the Best Reviewer of Your Own Papers? A Mechanism-Design-Based Approach to Estimation

Alice submits a number of papers to a machine learning conference and has knowledge of the quality of her papers. Given noisy grades provided by independent reviewers, can Bob obtain accurate estimates of the ground-truth quality of the papers by asking Alice a question about the ground truth? In this talk, we address this when the payoff of Alice is additive convex utility over all her papers. First, if Alice would truthfully answer the question because by doing so her payoff is maximized, we show that the questions must be formulated as pairwise comparisons between her papers. Moreover, if Alice is required to provide a ranking of her papers, which is the most fine-grained question via pairwise comparisons, we prove that she would be truth-telling. By incorporating the ground-truth ranking, we show that Bob can obtain an estimator with the optimal squared error in certain regimes based on any possible ways of truthful information elicitation. Moreover, the estimated grades are substantially more accurate than the raw grades when the number of papers is large and the raw grades are very noisy. Finally, we conclude the talk with several extensions and some refinements for practical considerations. This is based on arXiv:2206.08149 and arXiv:2110.14802.

Weijie Su
University of Pennsylvania
suw@wharton.upenn.edu

CP1

Graph Geneos for Tda

The relevance of concepts of graph and equivariant operator is difficult to overestimate in data analysis and geometric deep learning. How could we establish a link between these two concepts? In recent years, a topological and geometric study of group equivariant non-expansive operators (GENEOs) has been initiated, leading to theoretical results on their geometric properties and new methods for constructing such operators in practice [https://rdcu.be/bP6HV]. The main contribution of this research is to show that the mathematical machinery developed for GENEOs between signal spaces associated with invariance groups can be extended to graphs. One model considers a fixed graph, and data are represented as functions on the vertex set; another model starts from the complete graph on a fixed set of vertices, and sees data as functions on the edge set. The latter has the advantage to enable us to consider different graphs (with that set of vertices) as data. In addition to introducing the definition of GENEO between weighted graphs, we show how we can concretely construct this new type of GENEOs by adapting the definition of permutant, so extending previous results about classical GENEOs. We also describe several examples that show the potential of our approach and raise the hope that GENEOs can lead to discrimination of graph-theoretical features through Machine Learning algorithms, providing a new mathematical tool for data analysis.

Faraz Ahmad
University of Bologna
faraz.ahmad2@unibo.it

Massimo Ferri
University of Bologna
Italy
massimo.ferri@unibo.it

Patrizio Frosini
University of Bologna
patrizio.frosini@unibo.it

CP1

Geometric Data Analysis Through Quantum Dynamics

We propose that, much like how quantum mechanics models nature at ?ne scales, the ?ne-scaled resolution of the organization and structure of data is also best characterized using quantum mechanical processes. We realize this through two novel facets: on the one hand, we reveal a connection between spectral graph theory of data arising from finite Euclidean point cloud samples on Riemannian submanifolds and semiclassical analysis. On the other hand, we establish a consistency between wave dynamics on the finite data and on the underlying manifold with exponential convergence rates. Together, these yield a quantum-classical correspondence between matrix dynamics on data and the geodesic flow on the submanifold. Conceptually, these results connect the notion of discretization imposed by data sampling to the concept of quantization in physics. We demonstrate this quantum data analysis framework with a convergent manifold learning algorithm, inspired by properties of quantum dynamics of free particles on curved spaces and grounded in this novel interplay between spectral data analysis and the framework of microlocal/semiclassical analysis, for recovering geodesics from the sampled data. We illustrate this new algorithm on data sampled from model manifolds and moreover, we demonstrate its dimensionality reduction and anomaly-detection capabilities with a real-world COVID-19 mobility dataset.

Mohan Sarovar
Sandia National Laboratories
Livermore, CA
mnsarov@sandia.gov

Akshat Kumar
Clarkson University
akumar@clarkson.edu

CP1

Machine Learning Methods for Vessel Segmentation in Organs

The vascular system plays a crucial role in diagnostics, treatment, and surgical planning in a wide array of diseases. Recently, there has been a growing interest in automating the manual vessel segmentation process to save time. We aim to efficiently and effectively segment the vascular system in the liver organ using deep learning techniques in order to improve on current manual methods. To do this, we propose a 3D DenseNet using PocketNet paradigm with binary and ternary classifications that has less parameters to train than the state of the art methods. We explore the impact of various preprocessing techniques on the accuracy of the neural network using the dice score coefficient. The proposed model reduces training times and increases accuracy per training parameter in medical imaging segmentation of the liver vessels. We find that successful preprocessing and post-processing filters and neural network parameters are necessary for consistently high dice scores.

Bilyana Tzolova
Rice University
bmt3@rice.edu

David Fuentes
University of Texas MD Anderson Cancer Center
dtfuentes@mdanderson.org

Beatrice Riviere
Rice University
Houston, Texas, USA
riviere@caam.rice.edu

CP1

Cost Function for Assessing the Quality of Low-Dimensional Manifolds

Dimensionality reduction and manifold learning techniques are used in numerous disciplines to find low-dimensional manifolds in complex systems with many degrees of freedom. This approach allows for a substantial reduction in the number of parameters needed to visualize, describe and predict complex systems, but some topological properties of low-dimensional manifolds can hinder their practical application. Here, we present a quantitative metric for characterizing the quality of low-dimensional manifolds. The metric collects information about variance in dependent variable values happening across multiple spatial scales on a manifold. It can characterize two topological aspects in particular: non-uniqueness and feature sizes in low-dimensional parameterizations. Our metric is scalable with manifold dimensionality and can work across different linear and nonlinear dimensionality reduction techniques. Moreover, dependent variables that are most relevant in modeling can be selected a priori and the manifold topology can be assessed for those variables specifically. Using the metric as a cost function in optimization algorithms, we show that optimized low-dimensional manifolds can be found. The metric can be also used to tune the hyperparameters of reduction techniques and to select appropriate data preprocessing strategies to obtain improved manifolds. Our approach provides a way to quantify and automate decisions that need to be made prior to applying a reduction technique.

Kamila Zdybal
Université Libre de Bruxelles
kamilazdybal@gmail.com

Elizabeth Armstrong, James Sutherland
University of Utah
elizabeth.armstrong@chemeng.utah.edu,
james.sutherland@utah.edu

Alessandro Parente
Université Libre de Bruxelles
alessandro.parente@ulb.be

CP1

Unsupervised Homological Manifold Learning for Non-Euclidean Data

Homology is a powerful machinery to decode the global structures of a topological space in terms of homology groups. This algebraic topology characterization naturally

enables learning, process, and analysis of non-Euclidean data, which are essential to various emerging applications in topological data science, including manifold learning, shape recognition, and scene reconstruction. In recent years, extensive research has been devoted to developing algorithms for computing homology over $\mathbb{Z}_2$, e.g., persistent homology. Although the simplicity on $\mathbb{Z}_2$ offers computational advantages, it leads to dramatic loss of topological information. In this work, to strike a balance between computational efficiency and information retention, we propose a learning-based method to facilitate the computation of homology groups over $\mathbb{R}$. Specifically, we develop an iterative unsupervised learning algorithm to construct simplicial complexes and extract topological features from data, by which computing $\mathbb{R}$-homology groups is transformed to a sparse linear inverse problem (SLIP). By integrating the techniques of singular value decomposition and rank optimization, we design an effective kernel-based algorithm to find the solution of the associated SLIP in terms of Betti numbers of the constructed simplex. We show that this solution adequately reflects the homology groups guaranteed by universal coefficient theorem.

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

CP2

Matrix-Free Interior Point Methods for Point Set Matching Problems

Point sets matching problems can be handled by optimal transport. The mechanism behind it is that optimal transport recovers the point-to-point correspondence associated with the least curl deformation. Optimal transport is a special form of linear programming with dense constraints. Linear programming can be handled by interior point methods effectively, provided that the involved ill-conditioned Hessians can be computed accurately. Recently, matrix balancing has been employed to compute optimal transport. In this work, we incorporate sparse support constraints to matrix-balancing based interior point methods, in which the sparse constraint set fulfilling total support is iteratively updated according to the multiplier vectors. Total support condition ensures the existence of matrix balancing. Experiments justify the effectiveness of the proposed algorithm on point-sets with large cardinality.

Pengwen Chen
National Chung Hsing University
pengwen@nchu.edu.tw

CP2

High Dimensional Covariance Estimation in Cryo-EM

Cryogenic electron-microscopy (cryo-EM) is an imaging technique able to recover the 3D structures of proteins at near-atomic resolution. A unique characteristic of cryo-EM is the possibility of recovering the structure of flexible proteins in different conformations from a single electron microscopy image dataset. One way to estimate these conformations relies on estimating the covariance matrix of the scattering potential directly from the electron data. From that matrix, one can perform principal component analysis to recover the distribution of conformations of a protein. While theoretically attractive, this method has been constrained to low resolutions because of high storage and computational complexity; indeed, the covariance matrix contains $\mathcal{O}(N^6)$ entries where N is the size of the grid in one dimension. In this talk, we discuss and present a new estimator for the covariance matrix and show that we can compute it in a rank k-approximate covariance in $\mathcal{O}(kN^3)$. Finally, we demonstrate that we can recover the conformation of structures at high resolution using a sparsified PCA.

Marc Aurele T. Gilles
Cornell University, Center for Applied Mathematics
mg6942@princeton.edu

CP2

Probabilistic PCA Algorithms for Heterogeneous-Quality Data

Principal component analysis (PCA) is a workhorse data science technique for reducing data dimensionality, but it is suboptimal when the samples have heterogeneous quality, as is common in modern applications. PCA treats all samples uniformly so degrades when the noise is heteroscedastic across samples, as occurs, e.g., when samples come from different sources that are of different quality. This talk presents a new probabilistic PCA variant that estimates and accounts for this heterogeneity by incorporating it in the statistical model. Unlike the usual (homoscedastic) setting, the resulting nonconvex optimization problem is seemingly not solved by the singular value decomposition. We develop a heteroscedastic probabilistic PCA technique (HePPCAT) algorithm that uses efficient alternating maximization to jointly estimate both the underlying factors and the unknown noise variances. Simulation experiments illustrate the comparative speed of various algorithmic choices, the benefits of accounting for the heterogenous quality, and the seemingly favorable optimization landscape of this problem. We illustrate the technique on real environmental air quality data. For details see: https://doi.org/10.1109/TSP.2021.3104979

David Hong
University of Pennsylvania, U.S.
dahong67@wharton.upenn.edu

Kyle Gilman
University of Michigan
kgilman@umich.edu

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

Jeffrey Fessler
University of Michigan
fessler@umich.edu

CP2

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 net-

work. 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
hji5@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

CP2

Multi-Task fMRI Data Fusion Using Independent Vector Analysis and the PARAFAC2 Tensor Decomposition

Data fusion—the joint analysis of multiple related datasets—through coupled factorizations has the promise to enable enhanced knowledge discovery, and hence is an active area. Various formulations of coupled matrix factorizations have been proposed, each with its own modeling assumptions. In this presentation, we study two such methods, namely Independent Vector Analysis (IVA), i.e., extension of Independent Component Analysis (ICA) to multiple datasets, and PARAFAC2, a tensor factorization approach. We demonstrate the modeling assumptions of IVA and PARAFAC2 using simulations motivated by multi-task functional Magnetic Resonance Imaging (fMRI) data fusion, where multiple fMRI data sets in the form of *subjects* by *voxels* matrices corresponding to different tasks are jointly analyzed. Our simulation study reveals that while both methods can accurately capture the latent components, PARAFAC2 captures group differences more reliably, providing a compact representation when subject scores differ only up to a scaling across different datasets/tasks, and IVA performs better when different subject scores are expected in different datasets. Furthermore, on real multi-task fMRI data, we show how the two methods can be used for achieving two important goals at once, namely capturing group differences between patients with schizophrenia and healthy controls with interpretable components, as well as understanding the relationship across multiple tasks.

Isabell Lehmann
Signal and System Theory Group
Paderborn University
isabell.lehmann@sst.upb.de

Evrim Acar
Simula Metropolitan Center for Digital Engineering
evrim@simula.no

Tanuj Hasija
Signal and System Theory Group
Paderborn University
tanuj.hasija@sst.upb.de

M.A.B.S. Akhonda
University of Maryland
Baltimore County
mo32@umbc.edu

Vince Calhoun
Georgia state University
vcalhoun@gsu.edu

Peter Schreier
Signal and System Theory Group
Paderborn University
peter.schreier@sst.upb.de

Tülay Adali
University of Maryland
Baltimore County
adali@umbc.edu

CP2

Variance-Reduced Randomized Kaczmarz Algorithm in Xfel Single-Particle Imaging Phase Retrieval

In this work, we propose the Variance Reduced Randomized Kaczmarz (VR-RK) algorithm for XFEL signal particle imaging phase retrieval. The VR-RK algorithm is inspired by the randomized Kaczmarz algorithm and the variance reduction in stochastic gradient methods. The formulations of the VR-RK algorithm under the L_1 and L_2 constraints are also presented. Numerical simulations demonstrate that the VR-RK method has a faster convergence rate compared with the randomized Kaczmarz method. Tests on the synthetic signal particle imaging data and the PR772 XFEL real imaging data show that the VR-RK algorithm can recover information with higher accuracy. It is useful for biological data processing.

Poline Y. Xian
Hong Kong University of Science and Technology
Beijing Computational Science Research Center
poline3939@gmail.com

Haiguang Liu
Beijing Computational Science Research Center (CSRC)
hgliu@csrc.ac.cn

Xuecheng Tai
Hong Kong Baptist University
xuechengtai@hkbu.edu.hk

Yang Wang
Hong Kong University of Science and Technology
yangwang@ust.hk

CP3

Data-Driven Distributionally Robust Model Predictive Control for Nonlinear Systems

The optimization of nonlinear stochastic dynamic systems is a key challenge both in industry and in academia. Model predictive control (MPC) is the go-to method to address these problems. However, traditional MPC does not di-

rectly account for uncertainty, and in the presence of significant stochasticity, it can lead to poor performance and infeasibility issues. Stochastic model predictive control (SMPC) has risen as an alternative to deal with the presence of stochastic behaviour. SMPC methods assume that the true probability distribution of uncertainties is provided in advance. However, in real-world systems, only partial distribution information can be acquired for SMPC. The discrepancy between the true distribution and the distribution assumed can result in sub-optimality or even infeasibility. To address this, we present a new data-driven distributionally robust data-driven MPC scheme for stochastic nonlinear systems. We first propose a data-driven MPC scheme to control constrained stochastic linear systems using distributionally robust optimization. The resulting distributionally robust MPC framework is computationally tractable, efficient, and recursively feasible. Additionally, the information about the uncertainty can be determined empirically from the data. Subsequently we present alternatives to adapt this strategy to nonlinear systems in a tractable way. We show examples while examining the benefits and shortcomings of the proposed approach.

Ehecatl Antonio Del Rio Chanona
Imperial College London
a.del-rio-chanona@imperial.ac.uk

Panagiotis Petsagkourakis
Imperial college london
p.petsagkourakis@imperial.ac.uk

CP3

Combining Dynamic Mode Decomposition with Ensemble Kalman Filtering for Tracking and Forecasting

Data assimilation techniques, such as Ensemble Kalman Filtering, have been proven to be a highly effective and efficient way to combine noisy data with a mechanistic model to track and forecast systems. However, when dealing with high-dimensional data it is often difficult to derive a mechanistic model, making data assimilation difficult to apply. In this talk, we use Dynamic Mode Decomposition to generate a low-dimensional, linear model of a dynamical system directly from high-dimensional data which is defined by temporal and spatial modes. We then combine Dynamic Mode Decomposition with the Ensemble Kalman Filter (which we call the DMDEnKF) to iteratively update the current state and temporal modes as new data becomes available. We demonstrate that this approach has the advantage of being able to track time varying dynamical systems in synthetic examples. We then apply the DMDEnKF to influenza-like illness data from the USA Centers for Disease Control and Prevention.

Stephen Falconer
University of Surrey
s.falconer@surrey.ac.uk

CP3

A Structure Preserving Algorithm for Data Driven Koopman Analysis

Operator theoretic approaches to analyzing complex phenomena have garnered much attention of late as they allow interpretability backed by strong theoretical guarantees. Here, we propose an algorithm for approximating the Koopman operator that leverages known invariant solutions and eigen-functions induced by the underlying dynamical system. Linearity of the Koopman operator allows these priors to be encoded as linear constraints on a quadratic program, which can be easily solved. This approach recovers conventional linear stability analysis in the presence of a single invariant solution. Furthermore, it can always compute non-trivial invariant sets that empirically correlate with the basin of attraction of the specified fixed points/ limit cycles. Prototypical dynamical systems illustrate the applicability of this technique.

Gowtham S Seenivasaharagavan
University of California Santa Barbara
gowthamsan@ucsb.edu

Milan Korda
LAAS-CNRS
korda@laas.fr

Hassan Arbabi
MIT
arbabiha@gmail.com

Igor Mezic
Department of Mechanical Engineering
University of California, Santa Barbara
mezic@ucsb.edu

CP3

Numerical Methods for Approximately Optimal Equilibrium in Matching for Teams and Wasserstein Barycenter Beyond Discrete Measures

We propose numerical algorithms for computing approximately optimal solutions of the matching for teams problem involving measures that are not necessarily discrete. This problem was introduced by Carlier and Ekeland and contains the well-known Wasserstein barycenter problem as a special case. We develop two relaxation schemes in which marginal constraints are replaced by finitely many linear constraints: the first scheme relaxes the multi-marginal optimal transport formulation and the second scheme parametrizes the transfer functions in the original problem. Through proving tailored duality results, we develop two numerical algorithms for computing feasible and approximately optimal primal and dual solutions. These solutions give upper and lower bounds on the optimal value of the matching for teams problem, and their difference provides a direct estimate of the sub-optimality of these solutions. Moreover, we are able to control an upper bound on the sub-optimality to be arbitrarily close to 0. We subsequently prove that these approximately optimal primal and dual solutions converge when their sub-optimality goes to 0, and their limits constitute a true matching equilibrium. Thus, they can be regarded as an approximate matching equilibrium. Through numerical examples, we showcase that the proposed algorithms can produce high-quality approximate matching equilibria, and that the computed sub-optimality estimates are much less conservative than their theoretical upper bounds.

Ariel Neufeld
Nanyang Technological University
Singapore
ariel.neufeld@ntu.edu.sg

Qikun Xiang
School of Physical and Mathematical Sciences
Nanyang Technological University

qikun001@e.ntu.edu.sg

CP3

Model Selection of Nonlinear PDEs for Pattern Formation

Partial differential equations (PDE) have been widely used to reproduce patterns in nature and to give insight into the mechanism underlying pattern formation. Although many PDE models have been proposed, they rely on the prerequest knowledge of physical laws and symmetries, and developing a model to reproduce a given desired pattern remains difficult. We propose a method to estimate the best dynamical PDE for one snapshot of a target pattern under the stationary state without ground truth. We apply our method to nontrivial patterns,such as quasi-crystals (QCs), a double gyroid and Frank Kasper structures. By using the estimated parameters for the approximant of QCs, we successfully generate three-dimensional dodecagonal QCs from a PDE model. Our method works for noisy patterns and the pattern synthesised without the ground truth parameters, which are required for the application toward experimental data.

Natsuhiko Yoshinaga
WPI- Advanced Institute for Materials Research
yoshinaga@tohoku.ac.jp

CP3

Sensitivity Analysis for Quasi-Stationary Distributions

When a stochastic process has an absorbing state, the stationary conditional probability distribution conditioning on survival is called the quasi-stationary distribution (QSD). Many chemical reaction models, epidemiological models, and ecological models have QSDs instead of invariant probability measures. In this talk, I will introduce a data-driven approach to quantitatively estimate the sensitivity of QSDs against the change of noise types, boundary conditions, and diffusion approximations. The main idea is to use coupling technique to estimate the speed of mixing and the finite time sensitivity, which leads to the sensitivity of the QSD.

Yaping Yuan
Department of Mathematics and Statistics
University of Massachusetts, Amherst
yapingyuan@umass.edu

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

CP4

Agent-Based AI for Gaming Simulations in Unity

Reinforcement Learning (RL) is a novel machine learning methodology based on training Artificial Intelligence (AI) agents that can optimally make decisions within a simulated environment. In recent years, RL has gained much attention due to the success of companies such as DeepMind in developing and training software agents that can (1) navigate highly complex strategy games, (2) learn the rules of the game, (3) beat the highest ranked players in the world at that game, and (4) develop innovative never-before seen strategies to win. Utilizing this type of AI to solve puzzles or develop strategy within games can translate to the production of AI that can perform similarly or achieve similar goals in a real world setting. For example, the development of AI capable of making decisions within a warship game could eventually lead to the development of manless warships, preventing potential casualties. In this work, we examine the use of reinforcement learning to make optimal decisions in complex maze gaming environments including enemy AI and a points-based system. Additionally, we compare the performance to other popular methodologies such as A*. This custom-built maze environment inside of Unity will be assessing computational expense, accumulating a higher score via the point system, and evasion of the enemy, as well as comparison to a human player. Finally, we discuss potential application and future directions of this work.

Joseph E. Jabour, Quyen Brannon
USACE ERDC
joseph.e.jabour@erdc.dren.mil,
quyen.t.dong@erdc.dren.mil

Thomas L. Arnold
U.S. Army Corps of Engineers
Engineer Research and Development Center
thomas.l.arnold@erdc.dren.mil

Haley R. Dozier
U.S. Army Engineer Research and Development Center
Information Technology Laboratory
Haley.R.Dozier@erdc.dren.mil

CP4

Resilient Agent-Based Decision-Making

In recent years, the concept of system resilience has become a topic of interest within Government, Academia, and Industry. A system's ability to adapt and recover from risks is critical to maintaining reliable systems, especially in the event of unforeseen disruptions such as those caused by climate change, cyber-attack, or pandemics. Due to the unpredictability of these disruptions along with the lack of disruption data, traditional AI-based control methods often fail to anticipate unknown risks. Therefore, alternative methods are required to capture the complexities of a dynamic system. Recent advances in machine learning and artificial intelligence capabilities have led to the development of agent-based algorithms capable of modeling and predicting optimal sequential decision-making in a virtual environment. These programs are built to exploit AI systems' ability to learn and improve through repetition in addition to generalizing from past experience. This presentation will explore the use of reinforcement learning as a tool for resilient decision-making in a toy problem. The advantages and disadvantages of this methodology along with how to move forward in this field will also be discussed.

Haley R. Dozier
U.S. Army Engineer Research and Development Center
Information Technology Laboratory
Haley.R.Dozier@erdc.dren.mil

Alexandre Ligo
The University of Virginia
aligo@virginia.edu

Althea Henslee
U.S. Army Engineer Research and Development Center
Information Technology Laboratory

althea.c.henslee@erdc.dren.mil

Ashley Abraham
U.S. Army Engineer Research and Development Center
ashley.n.abraham@erdc.dren.mil

CP4

Offline Reinforcement Learning for Logistics Applications

Reinforcement Learning (RL) methods provide a mathematical framework for learning-based control and decision problems by mimicking the ways humans learn through interactions with their environment. While typical RL methods require continuous interaction with an environment or simulation resulting in an "online" experience, new avenues of offline reinforcement learning seek to utilize previously collected data rather than simulation. By combining the strengths of existing RL techniques with large datasets, offline RL creates a powerful decision-making tool for many avenues previously unexplored using RL. In this presentation, we propose a framework for turning large, real-world traffic datasets into a discrete-time stochastic control process for modeling optimal decision-making for routing problems. Additionally, we discuss the advantages and disadvantages of offline and online RL algorithms for routing applications.

Althea Henslee
U.S. Army Engineer Research and Development Center
Information Technology Laboratory
althea.c.henslee@erdc.dren.mil

Jules White
Vanderbilt University
jules.white@gmail.com

Haley R. Dozier, Ashley Abraham
U.S. Army Engineer Research and Development Center
Information Technology Laboratory
Haley.R.Dozier@erdc.dren.mil,
ashely.n.abraham@erdc.dren.mil

CP4

Efficient Performance Bounds for Primal-Dual Learning from Demonstrations

We consider large-scale Markov decision processes with an unknown cost function and address the problem of learning a policy from a finite set of expert demonstrations. We assume that the learner is not allowed to interact with the expert and has no access to reinforcement signal of any kind. Existing inverse reinforcement learning methods come with strong theoretical guarantees, but are computationally expensive, while state-of-the-art policy optimization algorithms achieve significant empirical success, but are hampered by limited theoretical understanding. To bridge the gap between theory and practice, we introduce a novel bilinear saddle-point framework using Lagrangian duality. The proposed primal-dual viewpoint allows us to develop a model-free provably efficient algorithm through the lens of stochastic convex optimization. The method enjoys the advantages of simplicity of implementation, low memory requirements, and computational and sample complexities independent of the number of states. We further present an equivalent no-regret online-learning interpretation.

Angeliki Kamoutsi
ETH Zurich
ETH Zürich
kamoutsa@control.ee.ethz.ch

Goran Banjac, John Lygeros
ETH Zurich
gbanjac@ethz.ch, jlygeros@ethz.ch

CP4

Model-Based and Model-Free Reinforcement Learning Comparison in Custom Game Environment

Mission engineering and analysis plays a critical role in meeting the Defense goal of protecting the security of our country. Mission engineering (ME) includes, but is not limited to, logistics and sustainment, identifying optimal force mixtures and concepts of employment. etc. The complexities of mission simulation tools, operator biases, and other factors can severely limit the outcomes for ME analyses. Despite this, there is a renewed need to estimate the impact on current and future combat scenarios of future technology as well as changes to the current force. Recent advances in Artificial Intelligence (AI) and Machine Learning can allow the Mission Engineering community to explore operations and requirements for weapon systems. Reinforcement learning (RL) is a field of AI in which you build an intelligent system that learns from its environment, such as a combat or mission environment, through interaction and evaluates what it learns in real-time. RL learning algorithms are divided in two main paradigms: model-free (MF) and model-based (MB). The objective for this presentation is to demonstrate the effectiveness of popular MB and MF algorithms on a custom game environment, then discuss potential algorithm applicability to mission engineering scenarios that vary in complexity.

Indu Shukla, Haley R. Dozier
U.S. Army Engineer Research and Development Center
Information Technology Laboratory
indu.shukla@erdc.dren.mil, Haley.R.Dozier@erdc.dren.mil

CP4

A Deep Reinforcement Learning Approach Towards Human-Like Multi-Agent Dynamics

Controlling the motion of virtual agents with Reinforcement Learning (RL) is an area of continuing interest in multiple research communities. In this work, we focus on simulating the dynamics of multiple virtually embodied agents moving in a crowd-like manner. Crowd dynamics are important in multiple domains, such as computer vision, robotics and entertainment. When virtually capturing such dynamics, most existing RL work is concerned with directing agents towards locomotion goals in a robot-like motion. While such approaches may be effective for imitating crowds for potential robotic and entertainment applications, the resulting dynamics do not resemble human-like crowds. In this work, we propose a novel RL model which rewards agents that move in a human-like manner. Our approach is based on a novel reward function, which includes components that reward movement towards goals, a penalty for collision with other agents, and a novel human-like navigation reward, which penalizes the agent's acceleration from increasing beyond a bio-mechanically allowed threshold. We demonstrate our method's robustness on multiple crowd locomotion scenarios with varying random conditions in terms of environment, number

of crowd agents, and navigation complexity, and show that our method generalizes well from one learned policy.

Tomer Weiss, Bilas Talukdar
New Jersey Institute of Technology
tweiss@njit.edu, bt26@njit.edu

CP5

Optimizing Penalization Hyperparameters in NMF Problems

Hyperparameters Optimization in learning algorithms represents an open problem of great interest in several fields. It has a direct impact on the performances of the algorithms as well as on the extraction of knowledge from datasets. Matrix Decompositions (MD) are gaining attention in Data Science as mathematical techniques able to capture latent information embedded in large datasets. Among low rank MD, Nonnegative Matrix Factorization (NMF) is one of the most effective methods for analyzing real-life nonnegative data. It can be seen as optimization problem often penalized to emphasize useful properties of the data matrix. How to choose optimal penalization hyperparameters (HPs) in this context represents an open issue. To the best of our knowledge, the literature panorama lacks of general framework addressing this problem. In this work, we consider HPs selection problem using bi-level optimization: the choice of the HPs is incorporated directly in the algorithm as part of the updating process. This problem was approached from two perspectives: theorems of existence and convergence of numerical solutions, under appropriate assumptions, are presented together with the proposal of a new algorithm to tune the HPs in NMF problems. The proposed approach provides competitive results to control sparsity on synthetic and real datasets. This is a joint work with Nicoletta Del Buono, Laura Selicato (University of Bari, Italy) and Rafal Zdunek(Politechnika Wroclawska, Poland).

Flavia Esposito
Department of Mathematics, University of Bari Aldo Moro
flavia.esposito@uniba.it

CP5

Comparison of First Order Optimization Algorithms for Neural Networks

In this day and age more and more applications tend to be based in some form of machine learning algorithms. One of the most widely used machine learning algorithms is neural networks. Even though neural networks tend to become more and more common, some problems still preserve. For example, the training of a neural network is a cumbersome optimization task that still requires fine tuning. The aim of this work is to compare various optimization methods used in neural network training. The algorithms that will be included in the computational comparison to train these neural networks are mainly first order optimization algorithms; these are Stochastic Gradient Descent, Adagrad, Adadelta, RMSprop, Adam, Adamax, Nadam, and FTRL. The experimental process will be composed of typical neural network structures like convolutional and recurrent ones. The experiments will be also performed for shallow and wide, deep and narrow as well as balanced neural networks. Moreover, another parameter that we will take into consideration is the density of the neural networks, meaning that the experiments will be performed for both dense and sparse neural networks. The optimization methods will be compared according to their execution time required to train the neural network at an accuracy threshold, the accuracy of the model after a certain number of epochs, and finally, the generalization of each model.

Xristos Krallis
University of Western Macedonia
x.krallis@uowm.gr

Nikolaos Ploskas
University of Western Macedonia
Greece
nploskas@uowm.gr

CP5

Federated Learning with Heterogeneous Data: A Superquantile Optimization Approach

We present a federated learning framework that is designed to robustly deliver good predictive performance across individual clients with heterogeneous data. The proposed approach hinges upon a superquantile-based learning objective that captures the tail statistics of the error distribution over heterogeneous clients. We present a stochastic training algorithm which interleaves differentially private client reweighting steps with federated averaging steps. The proposed algorithm is supported with finite time convergence guarantees that cover both convex and non-convex settings. Experimental results on benchmark datasets for federated learning demonstrate that our approach is competitive with classical ones in terms of average error and outperforms them in terms of tail statistics of the error.

Krishna Pillutla
University of Washington
pillutla@cs.washington.edu

Yassine Laguel
Rutgers Business School
laguel.yassine@gmail.com

Jérôme Malick
University of Grenoble
jerome.malick@univ-grenoble-alpes.fr

Zaid Harchaoui
University of Washington
zaid@uw.edu

CP5

A Cutting Plane Approach to Constrained Data Fitting with Splines

Throughout aerospace engineering, data fitting often requires function approximations to meet a variety of constraints, e.g. the function is monotonic in a specified input or the graph of the function is contained within a certain region. We present a method for constrained approximation to observed data using tensor product spline functions. We consider certain classes of shape constraints that result in a convex semi-infinite programming problem where the constraint system is indexed by an infinite set, e.g. every point in the domain. In the univariate case, this feasible set is semidefinite representable, and semidefinite programming may be leveraged to produce a constrained approximation. However, no such representation exists in higher dimensions. Our approach utilizes some basic prop-

erties of splines to develop an efficient cutting-plane algorithm for constrained multivariate approximation. Producing the cutting planes for the algorithm involves identifying the global extrema of a secondary spline function, which is done using an evaluation-free branch-and-bound approach. Our algorithm is contrasted against a common relaxation technique for obtaining an inner approximation to the semi-infinite problem. We conclude by demonstrating the performance of the algorithm in a number of examples.

Jeffrey D. Poskin
The Boeing Company
jdposkin@gmail.com

Thomas A. Hogan
The Boeing Math Group
thomas.a.hogan@boeing.com

CP5

Constrained Optimization on Manifolds

Many applications involve non-Euclidean data, such as graphs, strings, or matrices. In such cases, exploiting Riemannian geometry can deliver algorithms that are computationally superior to standard nonlinear programming approaches. This observation has resulted in an increasing interest in Riemannian methods in the optimization and machine learning community. In this talk, we consider the problem of optimizing a convex function on a (Riemannian) manifold subject to convex constraints. We will discuss several instances of this problem that arise in Machine Learning applications. Those include matrix-valued subroutines, such as barycenter problems, and optimization tasks on the space of probability distributions, endowed with the Wasserstein distance. In addition, we present algorithms for efficiently solving constrained tasks on manifolds. Specifically, we introduce Riemannian Frank-Wolfe (RFW) methods, a class of projection-free algorithms for constrained geodesically convex optimization and give guarantees for its efficiency.

Melanie Weber
Harvard University
mweber@seas.harvard.edu

CP6

Physics Informed Neural Network: Dynamics of Granular Matter Flow

We propose physics informed neural network that are trained to solve supervised learning tasks while respecting physical properties of the dynamics of growing sandpiles generated by a vertical source on a flat rectangular table, modeled by a 2×2 system of non-linear partial differential equations, for which classical/approximate Riemann solvers are not possible. The model is given by:

$$u_t + N_u[u,v](t,x) = 0 \quad \text{in} \quad \Omega \times (0,T], \tag{1}$$
$$v_t + N_v[u,v](t,x) = 0 \quad \text{in} \quad \Omega \times (0,T], \tag{2}$$
$$u(x,0) = u_0(x) = 0, \;\; v(x,0) = v_0(x) = 0 \;\; \text{in} \;\; \Omega \tag{3}$$

and with the boundary condition

$$u_b(.,t) = 0 \;\; \text{in} \;\; \partial\Omega \times [0,T] \tag{4}$$

where $N_u[u,v](t,x) = -(1 - |\nabla u|)v, N_v[u,v](t,x) = -\nabla.(v\nabla u) + (1 - |\nabla u|)v - f$,$u(\mathbf{x},t)$ is the local height of the pile containing the grains at rest and is called as the *standing* layer, and $v(\mathbf{x},t)$ is the rolling layer representing the sand above the *standing* layer, which is formed only by the grains that roll on the surface of the pile until they are captured by the standing layer. Simulations are presented to illustrate that the proposed surrogate model detects the solutions efficiently and comparisons are made with the existing studies.

Aekta Aggarwal
Indian Institute of Management, Indore
aektaaggarwal@iimidr.ac.in

Luv Verma
Indian Institute of Technology, Madras
luvverma2011@gmail.com

CP6

Efficient Multiscale PDE Solvers via Signal Decomposition and Neural Networks

Multiscale problems have long been challenging due to the computational cost for resolving the details. One general idea is to explore the intrinsic structure between scales which makes numerical homogenization possible. In the first half of this talk, we present a seamless solver for multiscale boundary value problems based on signal processing and numerical averaging techniques. The observed parameters are decomposed into different frequency domains based on scale detection and separation techniques, which are applicable to several seamless numerical solvers. The second half of the talk focuses on the possibility of neural-network-based solvers. The multiscale elliptic problem is translated into the equivalent Ritz form which enables parametrization by a suitable network structure. We also explore a unified way of describing different microstructures.

Ziheng Chen
Department of Mathematics
the University of Texas at Austin
ziheng.chen@math.utexas.edu

Bjorn Engquist
UT Austin
Department of Mathematics
engquist@oden.utexas.edu

CP6

Surrogate Models for Computational Fluid Dynamics Simulations Using Convolutional Autoencoder Neural Networks and Physical Constraints

Simulations of fluid flow are generally very costly because high grid resolutions are not only required to obtain quantitatively accurate results, but too low grid resolutions may also lead to qualitatively incorrect results. In applications, however, one is often not interested in accurate approximations of the complete flow field but only in the qualitative behavior of the flow or in individual quantities. In this talk, the use of convolutional autoencoder neural networks to construct efficient reduced order surrogate models for high fidelity computational fluid dynamics (CFD) simulations is discussed. In particular, the geometry of the computational domain is the input of the neural network, and the flow and pressure fields are the output. In order to construct accurate surrogate models, U-Net type convolutional neural networks are employed and the architecture and hyper parameters are optimized to this application. As a first step, a fully supervised approach, which requires the

availability of simulation results as the training data, is presented. After that, a novel approach is introduced, which does not require CFD simulation results but is based on introducing physical constraints via the loss function. As a testbed for the surrogate models, the flow around obstacles with varying shape and size within a channel is considered. Moreover, results for the application to geometries of arteries with aneurysms are presented.

Alexander Heinlein
Delft University of Technology
Delft Institute of Applied Mathematics
a.heinlein@tudelft.nl

Matthias Eichinger, Viktor Grimm
Department of Mathematics and Computer Science,
University o
eichingm@smail.uni-koeln.de, viktor.grimm@uni-koeln.de

Axel Klawonn
Universität zu Köln
axel.klawonn@uni-koeln.de

CP6

Approximation and Robust Preconditioning for the Low-Rank Tensor-Structured Solution of Elliptic PDEs

Multilevel low-rank approximation in the form of matrix-product states (MPS), or tensor train (TT) decomposition, has been rigorously analyzed for certain classes of functions solving elliptic second-order PDEs. Analytic functions, solutions with algebraic corner singularities and highly-oscillatory solutions to multiscale diffusion problems have been shown to admit exponentially convergent approximations of this type. The approximation power of MPS/TT tensor methods is based on the successive adaptive approximation in suitable low-dimensional subspaces. In the case of multilevel representation, that amounts to the adaptive computation of effective, low-parametric discretizations within extravagantly large but generic finite-element spaces. That leads to data-driven computations, in which the effective discretizations are adapted to the data and are constructed in the course of computation. In this contribution, we revisit recent results on multilevel approximation and BPX-type preconditioning in the MPS/TT decomposition and present a novel algorithm for preconditioning iterative tensor-structured solvers. Exploiting an hierarchy of nested finite-element spaces, it dramatically speeds up the computation of approximate low-rank solutions. Our numerical experiments demonstrate the efficiency of the new approach in two and three dimensions, including the setting of multiscale diffusion with solutions exhibiting low regularity and high-frequency oscillations.

Vladimir Kazeev
University of Vienna
vladimir.kazeev@univie.ac.at

CP6

Compressive Fourier Collocation Methods for High-Dimensional Diffusion Equations with Periodic Boundary Conditions

High-dimensional Partial Differential Equations (PDEs) are a popular mathematical modeling tool. However, standard numerical techniques for solving these PDEs are typically affected by the curse of dimensionality. In this work, we tackle this challenge while focusing on stationary diffusion equations defined over a high-dimensional domain with periodic boundary conditions. Inspired by recent progress in sparse function approximation in high dimensions, we propose a new compressive Fourier collocation method. Combining ideas from compressive sensing and spectral collocation, our method replaces the use of structured collocation grids with Monte Carlo sampling and employs sparse recovery techniques, such as orthogonal matching pursuit and ℓ^1 minimization, to approximate the Fourier coefficients of the PDE solution. We conduct a rigorous theoretical analysis showing that the approximation error of the proposed method is comparable with the best s-term approximation (with respect to the Fourier basis) to the solution. Using the framework of random sampling in bounded Riesz systems, our analysis shows that the compressive Fourier collocation method mitigates the curse of dimensionality with respect to the number of collocation points under sufficient conditions on the regularity of the diffusion coefficient. We also present numerical experiments that illustrate the accuracy and stability of the method for the approximation of sparse and compressible solutions.

Weiqi Wang
Concordia University
Department of Mathematics and Statistics
weiqi.wang@concordia.ca

Simone Brugiapaglia
Concordia University
simone.brugiapaglia@concordia.ca

CP6

Predicting the Geometric Location of Adaptive Coarse Basis Functions of Feti-Dp

Domain decomposition methods (DDMs) are highly scalable, iterative solvers for the solution of high-dimensional systems of linear equations, e.g., arising from the discretization of partial differential equations. The convergence rate of classic DDMs in general deteriorates severly for coefficient distributions with large discontinuities or contrasts in the coefficient function. To retain the robustness for such problems, the coarse space of the DDM can be enriched by additional coarse basis functions, often obtained by solving local generalized eigenvalue problems. However, the set-up and the solution of these eigenvalue problems typically takes up a significant part of the total time to solution. Additionally, for many realistic model problems, only the solution of a small number of the eigenvalue problems is necessary to design a robust coarse space. In general, it is difficult to predict a priori which of the eigenvalue problems are needed. Here, we train a supervised machine learning model to predict where eigenvalue problems have to be solved, often reducing its number significantly. To obtain such an a priori classification, we use dense feedforward neural networks and a mesh-independent sampling strategy which is comparable to an image recognition problem of the given coefficient function. We present numerical results for both, two and three spatial dimensions, for linear diffusion and elasticity problems and different realistic coefficient distributions.

Janine Weber
Universität zu Köln
Department of Mathematics and Computer Science
janine.weber@uni-koeln.de

Alexander Heinlein
Delft University of Technology

Delft Institute of Applied Mathematics
a.heinlein@tudelft.nl

Axel Klawonn
Universität zu Köln
axel.klawonn@uni-koeln.de

Martin Lanser
Universitaet zu Koeln
Mathematisches Institut
martin.lanser@uni-koeln.de

CP7

Convex Minimization with Nonlinear Compositions

We first investigate the duality properties of a minimization problem involving the sum of a nonlinearly composed convex function and linearly composed inf-convolved convex functions. A Kuhn–Tucker operator is constructed for this problem as an extension of that arising in classical Fenchel–Rockafellar duality theory. We show that the composite formulation and its dual can be solved by finding a zero of this Kuhn–Tucker operator via monotone operator splitting. The proximal algorithm we propose fully split all the constituents of the problem and operate in a block-iterative fashion. Data science applications are discussed.

Patrick L. Combettes
Department of Mathematics
North Carolina State University
plc@math.ncsu.edu

Luis Briceno-Arias
Univ. F. Santa Maria, Santiago, Chile
luis.briceno@usm.cl

CP7

A Principled Reinforcement Learning Strategy for Chaotic Dynamical Systems

Predicting the future state of a complex physical system, and possibly controlling it with respect to a given objective, is a cornerstone in a wide range of situations. This aim however often remains difficult to achieve. In particular, for a high-dimensional multi-scale nonlinear system, a model approximating its evolution is not necessarily available or usable due to real-time control constraints which drastically limit the affordable computational complexity. Data-driven control circumvents the need for an a priori model and constitutes an attractive approach in these situations. In this talk, we will focus on a reinforcement learning (RL) strategy fully leveraging its theoretical foundation in optimal control theory while accounting for challenges of practical situations such as weak observability of the state through the measurements, delay between action and its impact on the measurements and the aleatoric nature of the system due to random perturbations and non-stationarity. To thoroughly address these challenges, our approach relies on tools from Information Theory, the Mori-Zwanzig formalism, and the theory of dynamical systems, and allows a solid and consistent performance. Our framework will be illustrated on simplified models of turbulent flows such as the Kuramoto-Sivashinsky equations. The results will be put in perspective by comparing the RL policies with standard optimal controllers such as the full information Linear Quadratic Regulator (LQR).

Lionel Mathelin
LIMSI - CNRS, France
lionel.mathelin@lisn.upsaclay.fr

Remy Hosseinkhan Boucher, Amine Saibi, Onofrio Semeraro
LISN-CNRS, Univ. Paris-Saclay, France
remy.hosseinkhan@universite-paris-saclay.fr,
amine.saibi@universite-paris-saclay.fr,
onofrio.semeraro@universite-paris-saclay.fr

CP7

Communication-Efficient and Scalable Algorithms for Decentralized Consensus, Stochastic Optimization, Inference

We study the decentralized consensus, stochastic optimization, and hypothesis testing problems under compressed communications. Our main objective is to optimize the sum of n functions, each accessible to a single node, where local communications are subject to a fixed communication network. First, we propose a new decentralized consensus algorithm with quantized communications that scales linearly with the network size n. We prove that the proposed method converges to the average of the initial values held locally by the agents. Moreover, we consider the decentralized consensus and stochastic optimization problems under arbitrary compressed message sharing over a directed graph. We present an iterative push-sum algorithm with arbitrary compressed communications and provide convergence rates for the stochastic optimization problem on smooth functions that are either (i) strongly convex, (ii) convex, or (iii) non-convex. We finally study the problem of distributed hypothesis testing (non-Bayesian learning) and propose a Byzantine-resilient algorithm with compressed belief sharing. We prove an almost sure asymptotic and a probabilistic non-asymptotic linear convergence of the beliefs around the set of optimal hypotheses. We present numerical experiments that confirm our theoretical results and illustrate the scalability, communication efficiency, and fault tolerance of our algorithms.

Mohammad Taha Toghani, César A. Uribe
Rice University
mttoghani@rice.edu, cauribe@rice.edu

CP7

Gamma Ray Tracking with Integer Programs

State-of-the-art gamma ray spectroscopy as done by modern spectrometers such as AGATA and GRETINA relies on the ability to determine entire gamma ray trajectories from observed gamma ray interactions, a procedure known as tracking. Current methods typically depend on a simple clustering of interactions into gamma rays followed by a brute-force optimization of the "Figure of Merit", a measure of how well observed properties align with predicted physical properties, for determining the order. This two-step procedure often has the drawback of failing to distinguish gamma rays which have very similar trajectories as well as disallowing gamma rays which make large jumps between interactions. We propose a variety of mixed integer programs to address gamma ray tracking which address clustering and ordering of gamma rays simultaneously. We consider first the task of minimizing the Figure of Merit directly. We then consider the easier problem of where we have estimates of the number of clusters and the en-

ergies of the gamma rays and propose an integer program for this scenario. We demonstrate modest improvements in the amount of gamma rays correctly recovered and peak-to-total ratio on simulated data over the current methods.

Dominic Yang
University of California, Los Angeles
domyang@g.ucla.edu

Sven Leyffer, Torben Lauritsen, Hanqi Guo
Argonne National Laboratory
leyffer@anl.gov, torben@anl.gov, hguo@anl.gov

CP8

Machine Learning Extensions of Linear Estimation Theory

Recent years have seen active efforts within the geophysical community to merge traditional Data Assimilation (DA) methods with rapid advances in the field of Machine Learning (ML). Most research so far has been focused on variational DA approaches due to their apparent similarity in terms of how the underlying optimization problem is solved. During the upcoming conference, I will present a novel, ML-inspired DA theory which combines the flexibility of advanced particle-based methods and the analytical tractability of linear estimation algorithms. In particular, a new nonlinear state-space model will be constructed whose filtering and smoothing distributions are available in closed form and remain arbitrarily non-Gaussian. There are two additional characteristics which make the proposed framework especially appealing for large-scale applications. First, linear estimation results follow immediately as a special case of the obtained solutions. Second, the prior and posterior distributions belong to the same parametric family, implying that the correlation structure and associated dynamical balances in the state are preserved following the update step.

Hristo G. Chipilski
National Center for Atmospheric Research (NCAR)
hristochipilski@gmail.com

CP8

High-Dimensional Uncertain Field Prediction Using Nonlinear Reduced Order Modeling and Polynomial Chaos Expansion

This study proposes a non-intrusive reduced order modeling method using nonlinear dimensionality reduction (DR) and polynomial chaos expansion for uncertainty quantification in fields with complex, nonlinear features. Uncertainty propagation in simulations with high-dimensional outputs, such as computational fluid dynamics, is challenging due to limited training data and prohibitive computational costs. Proper Orthogonal Decomposition (POD) is a popular method for DR, but achieving parametric robustness is particularly challenging in problems with strong nonlinearity, discontinuity, and gradients. The proposed approach uses manifold learning, specifically Isometric Mapping (ISOMAP) to identify a low-dimensional latent space that better captures nonlinear features of the problem than existing POD-based methods. A sparse polynomial chaos expansion (PCE) is then used to construct a mapping between the uncertain input parameters and the latent space coordinates. The performance of the proposed nonlinear ROM is compared to the existing POD-PCE method on three numerical examples: supersonic flow past a wedge with inflow variation and two analytical test cases with discontinuities. These methods are benchmarked against Monte Carlo simulations to quantify the impact of polynomial order and sample count on the predicted mean, standard deviation, and uncertainty distributions.

Nikhil Iyengar
Georgia Institute of Technology
niyengar8@gatech.edu

Kenneth Decker
Georgia Intitute of Technology
kdecker8@gatech.edu

Dimitri Mavris
Georgia Institute of Technology
dm17@gatech.edu

CP8

Discovery of Interpretable Structural Model Errors by Combining Bayesian Sparse Regression and Data Assimilation

The discrepancy between the mathematical representations of a true physical system and its imperfect model is called the model error. These model errors can lead to differences between the solution of the model and the state of the system. Thus, there is increasing interest in reducing model errors, particularly by leveraging the rapidly growing observational data to understand the underlying mechanism. Here we introduce a framework named MEDIDA: Model Error Discovery with Interpretability and Data Assimilation. MEDIDA only requires a working numerical solver of the model and a small number of noise-free or noisy sporadic observations of the system. First the model error is estimated from differences between the observed states and model-predicted states (the latter are obtained from a number of one-time-step numerical integrations from the previous observed states). If observations are noisy, a data assimilation (DA) technique such as ensemble Kalman filter (EnKF) is employed to provide the analysis state of the system, which is then used to estimate the model error. Finally, a sparsity-promoting Bayesian method, is used to identify an interpretable, parsimonious, closed-form representation of the model error. We demonstrate the performance of MEDIDA in discovering different types of structural/parametric model errors, representing different types of missing physics, using noise-free and noisy observations.

Rambod Mojgani, Ashesh K. Chattopadhyay
Rice University
rambod.mojgani@rice.edu, akc6@rice.edu

Pedram Hassanzadeh
Mechanical Engineering and Earth Science
Rice University
pedram@rice.edu

CP8

Extreme Value and Large Deviation Theory Based Analysis of Anomalous Behavior in Streaming Data

Motivated by the need for anomaly detection in complex streaming data we present a probabilistic framework to explicitly model the normal and anomalous behaviors and reason about the data. We explore two algorithms – one based on extreme value theory and one on large deviation theory to robustly classify anomalies in such data. In particular, we show application of these ideas to COVID-19

data.

Abani Patra
Tufts University
abani.patra@tufts.edu

CP8

Discontinuous Neural Networks and Discontinuity Learning

In several practical situations, discontinuous functions may arise; while being by far a non-trivial task, the ability to detect discontinuity interfaces can be an important issue. This is the case of uncertainty quantification settings in which the Quantity of Interest (QoI) can exhibit jumps or gradient discontinuities, whose location cannot be predicted or easily detected; this situation can occur in problems characterized by a complex geometry, affected by uncertainty: a small variation of the stochastic parameters may induce an abrupt change in the topological structure of the domain, which is reflected by a jump in the values of an associated QoI. Non-smoothness of the QoI prevents effective use of standard stochastic collocation approaches, as the error decay would be exceedingly slow. To avoid this drawback, multi-element techniques can be used, leveraging the knowledge of discontinuity interfaces. Aiming at approximating discontinuous function, and simultaneously detecting discontinuity interfaces, we consider an approach involving Neural Networks. In particular, we define a novel typology of layers endowed with new learnable parameters and discontinuities in the space of the activation functions. These layers allow the creation of new Neural Networks, able both to approximate discontinuous functions and to learn and detect the discontinuity interfaces. Preliminary tests on discontinuous functions are proposed, in order to assess the potential of such instruments.

Sandra Pieraccini
Dipartimento di Scienze Matematiche
Politecnico di Torino, Italy
sandra.pieraccini@polito.it

Francesco Della Santa
Politecnico di Torino
Dipartimento di Scienze Matematiche
francesco.dellasanta@polito.it

CP9

Motion Correction for Cardiac Mri Images Using Generative Neural Networks with Disentangled Latent Space Variables

Patient motion poses a great challenge in magnetic resonance imaging (MRI), both for reconstruction as well as interpretation of images. While in static MRI, patient motion is considered an artifact, dynamic MRI, and in particular cardiac MRI, aims to image physiological dynamics such as heart motion. Nevertheless, also dynamic images might be distorted by additional motion artifacts, e.g., from breathing motion. The focus of this work is to remove such distortions while maintaining the desired dynamics. To this aim, a method based on generative neural networks is proposed, that represents an image sequence as the forward mapping via a generative neural network of some latent space variables. Assuming that the given sequence contains both wanted and unwanted motion, and the latent variables are split accordingly. Using some 1D prior information of one of the motion types, the proposed method is able to represent the given image sequence via the split latent variables. More importantly, this learned representation allows to freeze one (unwanted) type of motion while maintaining the other one. In dynamic MRI, as our numerical examples show, this allows for instance to freeze breathing motion while maintaining heart motion. The proposed method is fully unsupervised as all network parameters are learned directly from the given data. Furthermore, while we currently deal with reconstructed images only, an extension to image reconstruction is easily possible.

Abdullah Abdullah
University of Graz, Graz, Austria
+ The Chinese University of Hong Kong
abdullahuom1@yahoo.com

Martin Holler
University of Graz
Institute for Mathematics
martin.holler@uni-graz.at

Karl Kunisch
University of Graz and Radon Institute, Austrian Academy of
Science, Austria
karl.kunisch@uni-graz.at

Malena Landman
Department of Applied Mathematics and Theoretical Physics
University of Cambridge
ms2845@cam.ac.uk

CP9

Towards a Better Evaluation of Football Players Using Data Science

The game of football is undergoing a significant shift towards the quantitative. Much of the progress made in the analytics space can be attributed to play-by-play data and charting data. However, recent years have given rise to tracking data, which has opened the door for innovation that was not possible before. In this talk I will describe how to gain an edge in player evaluation by building off of traditional charting data with state-of-the-art player tracking data, and foreshadow how such methods will revolutionize the sport of football in the future.

Eric A. Eager
University of Nebraska - Lincoln
eric.eager@pff.com

CP9

Drift Matters: Improving Unsupervised Anomaly Detection in National Security Applications

In this presentation, we characterize the problem of unsupervised anomaly detection under concept drift for two very different fields of interest in national security: maritime trajectory surveillance and additive manufacturing. In the real world, the definition of what is normaland what is anomalous with respect to that normchanges over time. This is concept drift. Most methods for unsupervised anomaly detection under concept drift have been developed for cyber and network security application spaces. Yet many other applications exhibit different drift properties. In this presentation, we focus on three kinds of concept drift: recurrent, gradual, and abrupt. We first review how these kinds of drift affect unsupervised anomaly

detection in our case studies on maritime trajectory surveillance and additive manufacturing. We then demonstrate how incorporating the three kinds of concept drift into the underlying nominal models for anomaly detection improves analysis. We conclude by discussing how the three definitions of concept drift can and should be applied to other fields. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525

Amelia Henriksen
University of Texas at Austin
aahenri@sandia.gov

CP9

Parameter Identification for a Model for Multi-Functional Materials with Hysteresis and Thermodynamic Compatibility

Multifunctional materials such as magnetostrictive, electrostrictive, and piezolectric materials have tremendous potential for engineering applications as they are able to convert mechanical to electro-magnetic energy and vice-versa. In recent years, they have found a niche in the area of energy harvesting. One of the features of this class of materials is that they show significant hysteresis which needs to be modeled correctly in order to maximize the energy harvesting potential. A method of modeling multi-functional materials that exhibit the phenomenon hysteresis and is compatible with the laws of thermodynamics was developed recently. The model is based on the Preisach hysteresis operator and its storage function. The difficulty is that the parameters in the model appear in a non-linear fashion, and there are several constraints that must be satisfied by the parameters for thermodynamic compatibility. In this article, we present a novel methodology that uses the rate independent memory evolution properties of the Preisach operator to split the parameter estimation problem into two linear least squares problems with constraints. Then we used alternative direction methods of multipliers(ADMM) algorithm and accelerated proximal gradient method to estimate the Preisach weights. Numerical results are presented for Galfenol.

Md Sakhawat Hossain
Texas Tech University
md-sakhawat.hossain@ttu.edu

Ram V. Iyer
Dept. of Mathematics and Statistics
Texas Tech University
ram.iyer@ttu.edu

CP9

New Surrogate Marker of Chronic Kidney Disease Progression and Mortality in Medical-Word Virtual Space: Prospective Cohort Study

Chronic kidney disease (CKD) leads to end-stage renal disease (ESRD) and death. A new surrogate marker reflecting its pathophysiology has been needed for CKD therapy. In this study, we developed a virtual space unifying data in the medical literature and that of actual CKD patients and created a surrogate marker of CKD progression and mortality using natural language processing. A virtual space of medical words was constructed from the CKD-related literature ($n = 165,271$) using natural language processing, in which CKD-related words ($n = 106,612$) composed a network. The data of CKD patients of a prospective cohort study for three years ($n = 26,433$) were transformed into the space and linked with the network on the basis of category theory. We let the relationship between a patient and the outcome (ESRD or death) in the network be a surrogate marker of the outcome. The network satisfied the definitions of vector keeping their medical meanings. The inner products highly accurately predicted the primary outcomes; C-statistics, 0.911. Cox proportional hazards models with spline showed that the high inner products were associated with high hazard ratio of the primary outcomes (p¡0.0001). Moreover, the risk of the primary outcome in high-inner-product group was 21.92 (95% CI: 14.77, 32.51) times higher than that in the low-inner-product group. The relationship between medical words is a new surrogate marker for CKD therapy.

Eiichiro Kanda
Medical Science, Kawasaki Medical School
kms.cds.kanda@gmail.com

Bogdan I. Epureanu
Department of Mechanical Engineering
University of MIchigan - Ann Arbor
epureanu@umich.edu

Taiji Adachi
Institute for Frontier Life and Medical Sciences,
Kyoto University, Japan
adachi@infront.kyoto-u.ac.jp

Naoki Kashihara
Department of Nephrology and Hypertension,
Kawasaki Medical School, Japan
kashinao@med.kawasaki-m.ac.jp

CP9

Improved Characterization of Membrane Transport Properties Through Data Analytics

Membrane characterization provides essential information for the scale-up, design, and optimization of new separation systems. We recently proposed the Diafiltration Apparatus for high-Throughput Analysis (DATA), which enables a 10-times reduction in the time, energy, and the number of experiments necessary to characterize membrane transport properties. In this talk, we apply Fisher information matrix (FIM) analysis and model-based design of experiments (MBDoE) techniques to further improve DATA. We highlight two non-ideal phenomena, namely lag and overflow, which occur when changing the operating pressure of the system. Guided by the tools of data science, we show that modeling these phenomena can leverage the additional data within the start-up process to elucidate the underlying physics, improve the parameter precision, and brings insights to design a time-varying applied pressure in DATA. A time correction for permeate product collected is also introduced to improve the model predictions. Moreover, our framework, which integrates data analytics and instrumentation design can be applied to investigate concentration-dependent membrane performance to further accelerate the development of materials. For example, we apply the improved DATA to explore the dependency of membrane transport parameters on feed conditions of a surface-charged membrane by ranking candidate models using information criteria.

Xinhong Liu, Jonathan Ouimet, Laurianne Lair, William Phillip
Department of Chemical and Biomolecular Engineering
University of Notre Dame

xliu27@nd.edu, jouimet@nd.edu, llair@nd.edu, wphillip@nd.edu

Alexander Dowling
University of Notre Dame
adowling@nd.edu

CP10

Stochastic Correlation Modelling with Von Mises Process

Multi-asset financial derivatives can carry an intra-asset correlation risk. Previous empirical studies show that the correlation between assets is not stable, and there is a correlation risk premium. Pricing models for these derivatives should account for the time-varying nature of the correlation. Assuming constant correlation may lead to pricing and hedging risks. We propose a continuous-time model for correlation as a random variable on the circle. We consider a model based on the von Mises process whose stationary distribution is the von Mises distribution, the maximum entropy distribution on the circle. We discuss several estimation methods and results for the model. We apply the model to real-life financial data to study the correlation between equity and currency exchange rates. We also consider the application of the model to the pricing of options.

Sourav Majumdar, Arnab Kumar Laha
Indian Institute of Management Ahmedabad
phd18souravm@iima.ac.in, arnab@iima.ac.in

CP10

Chaotic Hedging with Iterated Integrals

In this paper, we extend the Wiener-Ito chaos decomposition to the class of continuous semimartingale diffusion processes, whose drift and diffusion coefficient are of linear growth. By omitting the orthogonality in the chaos expansion, we are able to show that every p-integrable functional, for $1 < p < \infty$, can be represented as sum of iterated integrals of the underlying process. Using a truncated sum of this expansion and neural networks for the integrands, whose parameters are learned in a machine learning setting, we show that every financial option can be approximated arbitrarily well. In this case, the hedging strategy of the approximating option can be calculated in closed form.

Ariel Neufeld
Nanyang Technological University
Singapore
ariel.neufeld@ntu.edu.sg

Philipp Schmocker
Nanyang Technological University
philippt001@e.ntu.edu.sg

CP10

Some Data-Science-Based Refinements of Financial Models with Applications

In this presentation, at first, we discuss the Barndorff-Nielsen and Shephard (BN-S) model, a stochastic volatility model that is useful for both derivative and commodity markets. Though this model is very efficient and analytically tractable, it is well known that it suffers from the absence of long-range dependence and many other issues. It will be shown that a refinement of this model is possible with the implementation of various data-science-based algorithms. An application of this data-science-based improved model will be presented to find an optimal hedging strategy for the oil commodity from the Bakken, a new region of oil extraction that is benefiting from fracking technology.

Indranil Sengupta
Department of Mathematics
North Dakota State University
indranil.sengupta@ndsu.edu

CP10

Model Invariants and Functional Regularization

When modeling data, we would like to know that our models are extracting facts about the data itself, and not about something arbitrary, like the order of the factors used in the modeling. Formally speaking, this means we want the model to be invariant with respect to certain transformations. Here we look at different models and the nature of their invariants. We find that regression, MLE and Bayesian estimation all are invariant with respect to linear transformations, whereas regularized regressions have a far more limited set of invariants. As a result, regularized regressions produce results that are less about the data itself and more about how it is parameterized. To correct this, we propose an alternative expression of regularization which we call functional regularization. Ridge regression and lasso are special cases of functional regularization, as is Bayesian estimation. But functional regularization gives a framework under which the models become invariant with respect to linear transformations. It is also more flexible, and easier to understand, thus yielding a number of advantages over ridge regression and lasso.

Harvey Stein
Bloomberg LP
hjstein@gmail.com

CP10

Mathematics of Big Data in Finance

We present a new approach to study big data in finance (specifically, in limit order books), based on stochastic and mathematical modelling of price changes associated with high-frequency and algorithmic trading. We first introduce a big data in finance, namely, limit order books (LOB), and describes them by Lobster data-academic data for studying LOB. Numerical results, associated with Lobster and other data, are presented, and explanation and justification of our method of studying of big data in finance are considered. We also describe various stochastic models for mid-price changes in the market, and explain how to use these models in practice, highlighting the methodological aspects of using the models.

Anatoliy Swishchuk
University of Calgary
aswish@ucalgary.ca

CP11

Optimizing Shoe Midsoles for Running Performance

Additive manufacturing has become a viable and trendy production method for the midsoles of running shoes. While traditional midsoles are made from uniform foam materials, shoes made with 3D printing technology can

be designed to have spatially varying material properties, potentially allowing for the creation of higher performing running shoes than exist on the market today. Exploring this expanded design space, however, is difficult because the process of fabricating and testing new designs is slow and expensive. Therefore, to identify shoe designs that may be high-performing and thus worthy of pursuing further through fabrication and testing, a new method for predicting the performance of midsole designs was created. This method is based on both a dynamical model of running and on the existing biomechanical running data available in the field. This new model treats muscle activation as an optimization problem that a runners body is trying to solve. The optimization problem has an initially unknown objective function which is learned from data. The resulting model successfully recreates the relationship between shoe properties and running performance for well-established shoe designs, and it allows the flexibility for evaluation of shoes which have yet to be made. Specific 3D-printed midsole geometries were then evaluated using this method, and the predicted highest performing shoe design was fabricated for further evaluation.

Sarah C. Fay, Anette "Peko" Hosoi
Massachusetts Institute of Technology
scfay@mit.edu, peko@mit.edu

CP11

Combining Agglomerative Clustering Models in an Ensemble Learning Algorithm for Intrusion Detection

Anomaly-based endpoint detection and response (EDR) and intrusion detection systems (IDS) help protect enterprise networks, cyber-physical systems, and Internet of Things, from unauthorized activities (e.g., botnets, distributed denial of service, web attacks) which can degrade and disrupt communications. In this talk, we develop an ensemble learning algorithm which combines unsupervised and ensemble learning to detect cyberattacks with a variety of characteristics. With this modeling approach, we implement hierarchical agglomerative clustering algorithms with different linkages and pairwise distances between observations combined with an ensemble learning algorithm. It is an attractive approach for two primary reasons: unsupervised learning is useful for monitoring network traffic for zero-day attacks; and ensemble learning tends to improve prediction performance results more than can be achieved with a single algorithm. In addition, we present prediction results for network traffic datasets to demonstrate the effectiveness of this algorithm for EDR and IDS.

Nandi Leslie
Raytheon Technologies
nandi.o.leslie@raytheon.com

Frederica Nelson, Charles Kamhoua
Army Research Laboratory
frederica.f.nelson.civ@army.mil,
charles.a.kamhoua.civ@army.mil

CP11

Longitudinal Issue Analysis of Wikibots

Wikibots are software bots made specifically to maintain Wikipedia, one of the most notable works of the nonprofit foundation Wikimedia. They automatically maintain over 50 million pages of Wikipedia by rapidly performing repetitive and mundane tasks to edit those pages, making them much more efficient than manpower. Now, there are over 2,500 different types of Wikibots, and more will continue to be created. However, as a result of limited time and experience, the volunteer bot designers working on Wikibots generally prioritize the functionality of the bots with respect to its ability to carry out automated tasks over enhanced security. Such security vulnerabilities can be abused by malicious users to manipulate informative content on the platform. Using a longitudinal assessment strategy based on Pylint scores, our work analyzes the evolution of 21 Wikibots. Through the analysis, six major security defects are identified with their evolutions. It is demonstrated that our method may help Wikibots developers avoid certain potential defects. From a software engineering point of view, our work provides a practical tool to help maintain the Wikipedia community.

Cindy X. Lin
Mira Costa High School
cindylinpacific@gmail.com

Yang Cao
Virginia Tech
ycao.vt@gmail.com

CP11

Machine Learning in Ultra-Intense Laser Physics

As a professor of physics at Ohio State University, I lead NSF funded project to use machine learning for the control and optimization of ultra-intense laser systems, especially as it relates to accelerating protons to significant energies. In this talk I will review various efforts to leverage machine learning in ultra-intense laser science. I will also summarize some of the unique challenges for machine learning in this field and ways that the community could potentially help improve methods to overcome these challenges.

Chris Orban
Ohio State University
orban@physics.osu.edu

CP11

Efficient and Robust Nonlinear High-Dimensional Maximum Entropy Estimation via Nonlinear Primal-Dual Hybrid Gradient Algorithms

Maximum entropy models are a widely-used class of statistical models for estimating probability distributions from data. These models use the maximum entropy principle to provide a mean for constructing probability distributions that reproduce as closely as possible key statistics of data sets. Large-scale Maxent models require estimating probability distributions from massive data sets comprising billions of features. Due to this enormous number, large-scale Maxent models need efficient and robust algorithms to scale well. State-of-the-art algorithms for Maxent models, however, were not originally designed to handle massive data sets; these algorithms either rely on technical devices that may yield unreliable numerical results or lack scalable parallelism or scale poorly in size or depend on assumptions (e.g., smoothness) that are not satisfied by many Maxent models in practice. These limitations make it practically impossible to construct large-scale Maxent models for applications without adequate and costly computational resources. In this talk, we present a novel optimization algorithm that overcomes the shortcomings of state-of-the-art algorithms used for constructing large-scale Maxent models. Our proposed algorithm is an accelerated nonlinear

variant of the classical primal-dual hybrid gradient algorithm. We describe our method in detail and present some preliminary numerical results on large data sets.

Gabriel Provencher Langlois, Jérôme Darbon
Brown University
gabriel_provencher_langlois@brown.edu,
jerome_darbon@brown.edu

CP11

Predicting Hurricane Trajectories: A Comparison Between Classical Statistical Modelling and Machine Learning Methods

Natural disasters such as storms, cyclones and hurricanes are detrimental to many countries across the globe due to their destructive nature. National and International Hurricane Centres are tasked with tracking these trajectories for timely projections. In addition to casualties, infrastructural damage has incurred high costs and environmental degradation. In this regard, the accurate prediction of hurricane pathways is critical to mitigate the risks. Both classical statistical modelling and machine learning have specific merits for timely disaster management. Classical statistics often attempt to determine relationships between variables for inferential tests. Machine learning models, while largely built on statistical theory, tend to compromise interpretability to gain in predictive prowess. This work provides a comparison between traditional and machine learning methods with an overview of aspects of accuracy, variable importance, interpretability and computation time. The likelihood of impact and assumed trajectories based on empirical evidence are supported by predictions, using either method. These metrics are crucial for more robust models in predicting future hurricane trajectories.

Letetia M. Addison, Kevan Rajaram
The University of the West Indies
letetia.addison@gmail.com, kevan.rajaram@gmail.com

CP12

Fast and Provable Tensor Robust Principal Component Analysis via Scaled Gradient Descent

An increasing number of data science and machine learning problems rely on computation with tensors, which better capture the multi-way relationships and interactions of data than matrices. When tapping into this critical advantage, a key challenge is to develop computationally efficient and provably correct algorithms for tensor processing that are simultaneously robust to data corruptions and ill-conditioning. This lecture tackles tensor robust principal component analysis (RPCA), which aims to recover a low-rank tensor from its observations contaminated by sparse corruptions, under the Tucker decomposition. To minimize the computation and memory footprints, we propose to directly recover the low-dimensional tensor factors—starting from a tailored spectral initialization—via scaled gradient descent (ScaledGD), coupled with an iteration-varying thresholding operation to adaptively remove the impact of corruptions. Theoretically, we establish that the proposed algorithm converges linearly to the true low-rank tensor at a constant rate that is independent with its condition number, as long as the level of corruptions is not too large. Empirically, we demonstrate that the proposed algorithm achieves better and more scalable performance than the state-of-the-art matrix and tensor RPCA algorithms through synthetic experiments and real-world applications.

Harry Dong, Tian Tong
Carnegie Mellon University
harryd@andrew.cmu.edu, ttong1@andrew.cmu.edu

Cong Ma
University of Chicago
congm@uchicago.edu

Yuejie Chi
Carnegie Mellon University
yuejiec@andrew.cmu.edu

CP12

Theory and Computation of Persistence Diagram Bundles: A Multidimensional Generalization of Vineyards

In topological data analysis, we use algebraic topology to analyze the "shape" and "global structure" of a data set. In particular, we use "persistent homology" (PH) to study how the topology of a data set changes as a scale parameter changes. Developing new methods for analyzing how the topology changes as multiple parameters vary is an active area of research. For example, if a point cloud evolves over time, then one might be interested in using time as a second parameter. In nature, common examples of time-evolving point clouds include swarming or flocking animals whose positions and/or velocities are represented by points. In many of these dynamical systems, distinct topological features arise in different parameter regimes. When there are only two parameters (e.g., a scale parameter and time), one can efficiently compute a "vineyard", which encodes how PH changes over time. In this work, I generalize the concept of a vineyard to an n-parameter vineyard, in which n is arbitrarily large. I construct an algorithm for efficiently computing an n-parameter piecewise-linear vineyard, and I implement it for the case $n = 3$. I also show that n-parameter vineyards for $n > 2$ have an intrinsic limitation that ordinary (2-parameter) vineyards do not have: namely, it is not generally possible to track a single homology class over the entire parameter domain (i.e., there are not individual "vines" in an n-parameter vineyard when $n > 2$).

Abigail Hickok
UCLA
abigailhickok2@gmail.com

CP12

Scalable Symmetric Tucker Tensor Decomposition

Motivated by practical usages, we study the best low-rank Tucker decompositions of symmetric tensors. We advocate the straightforward projected gradient descent (PGD) framework for the underlying computation. Higher-order multivariate moment, an exemplar of symmetric tensors, is the primary interest of this paper. We develop scalable adaptations of the basic PGD and HOEVD to particularly decompose sample moment tensors. With implicit and streaming techniques, we evade the heavy overhead of building the moment tensor and caching it into limited memory. Such reductions make the practical computation realizable when involving vast data instances and exceedingly high dimensions. Numerical assessments demonstrate the remarkable efficiency of the algorithm and the applicability of moment tensor decompositions in scientific domains, such as anomaly detections. At last, we

study the convergence analysis on Riemannian manifold and show that the update sequence derived by the PGD solver achieves first and second-order criticality.

Ruhui Jin
University of Wisconsin
rjin@math.wisc.edu

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

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

Rachel Ward
Department of Mathematics
University of Texas at Austin
rward@math.utexas.edu

CP12

Lasso-Inspired Variants of Weighted Orthogonal Matching Pursuit with Applications to Sparse High-Dimensional Approximation

Motivated by recent developments in sparse high-dimensional approximation from Monte Carlo sampling, we propose new weighted generalizations of the Orthogonal Matching Pursuit (OMP) algorithm. Greedy algorithms of this type are more computationally efficient than convex optimization-based methods for small values of the target sparsity and offer a promising way to mitigate the curse of dimensionality. In this work, we propose new theoretically-justified greedy selection criteria that are inspired by variants of the LASSO optimization program. A key issue is the robustness of the optimal choice of the tuning parameter with respect to the measurement noise, which is realized by the square-root LASSO program in the context of convex optimization. We investigate how this property is carried over into the context of LASSO-based OMP methods. Conducting numerical experiments in high-dimensional polynomial approximation, we show the efficacy of the proposed algorithms by studying the recovery error as a function of the algorithm iterations. Moreover, we illustrate settings where the optimal choice of the tuning parameter is more robust against the noise.

Sina Mohammadtaheri
Concordia University
Concordia University
sina.mohammadtaheri@concordia.ca

Simone Brugiapaglia
Concordia University
simone.brugiapaglia@concordia.ca

CP12

Scalable Algorithms for Statistical Hypothesis Testing

Statistical testing of hypothesis on data is central to the scientific method. Practitioners often choose which test to use on the basis of *computational* considerations, rather than purely *statistical* considerations: for example, they may choose the χ^2-test over Fisher's exact test because it is computationally cheaper, and the latter over Barnard's, for the same reason. These tests though make different *statistical assumptions* on the data generation process, but these discrepancies are often overlooked by scientists who often chose the *computational cheaper test* without considering the impact of the different assumptions. In this work we develop *scalable* algorithms for statistical hypothesis testing, so tests can be chosen purely on statistical grounds. We consider *scalability along multiple axes*, including: the size of the data, the number of hypotheses to test (our algorithms control the *Family-Wise Error Rate*, but we also discuss controlling the *(marginal) False Discovery Rate*), the richness of the data (e.g., for tests on time series or graphs), the assumptions made on the data generating process (e.g., by considering more representative null models). We achieve these results by combining smart hypothesis space exploration techniques adopted from pattern mining with concepts from statistical learning theory, such as Rademacher averages and pseudodimension.

Matteo Riondato
Amherst College
Department of Computer Science
mriondato@amherst.edu

CP12

A Lagged Particle Filter for Stable Filtering of Certain High-Dimensional State-Space Models

The process of stochastic filtering is a very challenging task as analytical solutions are typically not available and many numerical approximation methods can have a cost that scales exponentially with the dimension of the state of the system (the curse of dimensionality). In this talk, I focus on the problem of high-dimensional filtering of state-space models (SSMs) at discrete times. Inspired by lag-approximation methods for the smoothing problem, we introduce a lagged approximation of the smoothing distribution that is necessarily biased. We develop a sequential Monte Carlo (SMC) method to recursively estimate expectations with respect to our biased smoothing distributions. Moreover, we prove for a class of SSMs that can contain dependencies amongst coordinates that as the dimension of the state $d \to \infty$ the cost to achieve a stable mean square error in estimation, for classes of expectations, is of $\mathcal{O}(Nd^2)$ per-unit time, where N is the number of simulated samples in the SMC algorithm. Our methodology is implemented on several challenging high-dimensional examples including a shallow-water model and is found to be superior to commonly used methods like e.g. the ensemble Kalman Filter.

Hamza Ruzayqat, Aimad Er-Raiy
King Abdullah University of Science and Technology
hamza.ruzayqat@kaust.edu.sa, aimad.erraiy@kaust.edu.sa

Alexandros Beskos
University College London
a.beskos@ucl.ac.uk

Dan Crisan
Imperial College
d.crisan@imperial.ac.uk

Ajay Jasra
King Abdullah University of Science and Technology
ajay.jasra@kaust.edu.sa

Nikolas Kantas
Imperial College London

n.kantas@imperial.ac.uk

CP13

Bounding Wasserstein Distance with Couplings

Markov chain Monte Carlo (MCMC) provides asymptotically consistent estimates of intractable posterior expectations as the number of iterations tends to infinity. However, in large data applications, MCMC can be computationally expensive per iteration. This has catalyzed interest in sampling methods such as approximate MCMC, which trade off asymptotic consistency for improved computational speed. In this article, we propose estimators based on couplings of Markov chains to assess the quality of such asymptotically biased sampling methods. The estimators give empirical upper bounds of the Wassertein distance between the limiting distribution of the asymptotically biased sampling method and the original target distribution of interest. We establish theoretical guarantees for our upper bounds and show that our estimators can remain effective in high dimensions. We apply our quality measures to stochastic gradient MCMC, variational Bayes, and Laplace approximations for tall data and to approximate MCMC for Bayesian logistic regression in 4500 dimensions and Bayesian linear regression in 50000 dimensions.

Niloy Biswas
Harvard University
niloy_biswas@g.harvard.edu

Lester Mackey
Microsoft Research New England
lmackey@stanford.edu

CP13

Response-Aided Score-Matching Approaches for Big Data Analysis under Generalized Linear Models

Big data analysis has brought a lot of challenges to traditional statistical methods. We propose an efficient method called Response-Aided Score Matching Representative (RASMR) approach to facilitate the big data analysis under generalized linear models. This method utilizes representatives of natural or algorithmically generated data blocks to produce estimates of parameters, which approximate the estimate from full data set very well. In additional, a major advantage of RASMR is that the representatives of the data set can be used for further analysis. The accurate estimation and high quality representatives enable promising application of RAMSR in a variety of statistical analysis problems, such as link function selection, variable selection and cross-validation in big data analysis.

Keren Li
Northwestern University
li.keren.cn@gmail.com

Jie Yang, Duo Zheng
University of Illinois at Chicago
jyang06@uic.edu, dzheng22@uic.edu

CP13

Divergence Frontiers for Generative Models: Sample Complexity, Quantization Effects, and Frontier Integrals

The spectacular success of deep generative models calls for quantitative tools to measure their statistical performance. Divergence frontiers have recently been proposed as an evaluation framework for generative models, due to their ability to measure the quality-diversity trade-off inherent to deep generative modeling. We establish non-asymptotic bounds on the sample complexity of divergence frontiers. We also introduce frontier integrals which provide summary statistics of divergence frontiers. We show how smoothed estimators such as Good-Turing or Krichevsky-Trofimov can overcome the missing mass problem and lead to faster rates of convergence. We illustrate the theoretical results with numerical examples from natural language processing and computer vision.

Lang Liu
Department of Statistics
University of Washington
liu16@uw.edu

Krishna Pillutla
University of Washington
pillutla@cs.washington.edu

Sean Welleck
University of Washington
Allen Institute for Artificial Intelligence
wellecks@uw.edu

Sewoong Oh
University of Washington
sewoong@cs.washington.edu

Yejin Choi
University of Washington
Allen Institute for Artificial Intelligence
yejin@cs.washington.edu

Zaid Harchaoui
University of Washington
zaid@uw.edu

CP13

Logical Analysis of Data: Estimation of Cause-Effect Relationship under Noise. A New Approach to Classification

Logical analysis of data (LAD) is a methodology introduced by Peter L. Hammer in 1986, aimed at discovering hidden structural information in data sets by using the theory of partially defined Boolean functions. A modification of the approach involving estimation of the cause-effect relationship under noise was proposed by Prekopa in 1994 and turned out to be efficient for data sets for which the sets of positive and negative observations can be separated with a good accuracy by a hyperplane. This paper develops a technique that combines Prekopa's method with initial data clustering. We illustrate the method on water quality monitoring data set in Perm, Russia.

Mariya Naumova, Andras Prekopa
Rutgers University
mnaumova@business.rutgers.edu,

prekopa@rutcor.rutgers.edu

CP13

Data-Driven Learning Based Algorithms for Multiscale Problems

In this talk, we present some recent progress on the development of data-driven, machine learning algorithms for solving a class of forward and inverse problems with multiple scale features. The algorithms make use of existing multiscale solvers to build the internal structures of the whole simulator. The first part of the talk focuses on a class of parametrized and multiscale time-dependent problems. For the forward problem, we developed a multi-stage neural network architecture, containing a front-end reduction module formed by the multiscale solvers, for approximating the solutions of the problem. The second part of this talk considers a class of multiscale inverse problems. We formulate the problem into the framework of reinforcement learning and solve it with the aid of multi-level Monte Carlo Markov Chain sampling methods. Numerical results will be presented to demonstrate the efficiency and accuracy of the proposed computational methods.

Sai Mang Pun
Texas A&M University
simonpun123@tamu.edu

Eric Chung
The Chinese University of Hong Kong
Department of Mathematics
tschung@math.cuhk.edu.hk

Yalchin Efendiev
Dept of Mathematics
Texas A&M University
efendiev@math.tamu.edu

Wing Tat Leung
Department of Mathematics
University of California, Irvine
wtleung@uci.edu

Zecheng Zhang
Purdue University
zecheng.zhang.math@gmail.com

CP13

Sampling with Mirrored Stein Operators

Accurately approximating an unnormalized distribution with a discrete sample is a fundamental challenge in machine learning and probabilistic inference. Particle evolution methods like Stein variational gradient descent tackle this challenge by applying deterministic updates to particles using operators based on Steins method and reproducing kernels to sequentially minimize Kullback-Leibler divergence. However, these methods break down for constrained targets and fails to exploit informative non-Euclidean geometry. In this talk, I will introduce a new family of particle evolution samplers suitable for constrained domains and non-Euclidean geometries. These samplers are derived from a new class of Stein operators and have deep connections with Riemannian Langevin diffusion, mirror descent, and natural gradient descent. We demonstrate that these new samplers yield accurate approximations to distributions on the simplex, deliver valid confidence intervals in post-selection inference, and converge more rapidly than prior methods in large-scale unconstrained posterior inference. Finally, we establish the convergence of our new procedures under verifiable conditions on the target distribution.

Jiaxin Shi
Microsoft Research New England
ishijiaxin@gmail.com

Chang Liu
Microsoft Research Asia
chang.liu@microsoft.com

Lester Mackey
Microsoft Research New England
lmackey@stanford.edu

CP14

Partial Identifiability for Nonnegative Matrix Factorization

Given a nonnegative input matrix X and a factorization rank r, nonnegative matrix factorization (NMF) aims to compute two nonnegative matrices, W with r columns and H with r rows, such that $WH = X$. In this talk, we discuss the uniqueness, also known as the identifiability, of the factors, W and H, in nonnegative matrix factorization (NMF). Most works so far have focused on the identifiability of the two factors, W and H. In this talk, we present new identifiability results regarding the partial uniqueness for a subset of the columns of W and rows of H. We illustrate our results in multivariate curve resolution that aims to unmix the components in a chemical reaction.

Nicolas Gillis
Université de Mons
Department of Mathematics and Operational Research
nicolas.gillis@umons.ac.be

Róbert Rajkó
University of Pécs
rajko@gamma.ttk.pte.hu

CP14

Big Data Minimum Sum-of-Squares Clustering

The purpose of this talk is to define key concepts of the minimum sum-of-squares (MSSC) clustering for big data, analyse the big data phenomenon and provide a theoretical as well as empirical review of the most prominent existing MSSC algorithms for large and big data. We will outline the required as well as desirable properties of the clustering algorithms for big data. In practice, analysis of the available amount of data by traditional methods on an average computer leads to one of the following situations: there are no difficulties with their processing; there are some technical difficulties, but processing is still possible; processing is not possible due to lack of computing resources or unacceptable time costs. According to the situations listed above, we introduce the corresponding classification of the analyzed datasets: small, large, and big, respectively. Thus, big data is a dataset of such a huge volume that its processing causes significant technical difficulties or is impossible when using traditional methods and average computing resources. Big data is a modern phenomenon that, on one hand, presents specific technical difficulties for an effective processing. On the other hand, big data provides substantial benefits that eliminate the problem of data shortage. The following algorithms will be reviewed

and compared to each other: Forgy K-means, K-means++, K-means‖, Ward's method, LMBM-Clust, and their other variations.

Ravil Mussabayev
University of Washington
ravmus@uw.edu

Rustam Mussabayev
Institute of Information and Computational Technologies
Almaty, Kazakhstan
rustam@iict.kz

CP14

Some Theory on Generalized Bayesian Additive Regression Trees (BART)

Bayesian Additive Regression Trees (BART) are a powerful ensemble learning technique for modeling nonlinear regression functions. Although initially BART was proposed for predicting only continuous and binary response variables, over the years multiple extensions have emerged that are suitable for estimating a wider class of response variables (e.g. categorical and count data) in a multitude of application areas. In this work we describe a Generalized framework for Bayesian trees and their additive ensembles where the response variable comes from an exponential family distribution and hence encompasses a majority of these variants of BART. We provide sufficient conditions on the response distribution, under which the posterior concentrates at a minimax rate, up to a logarithmic factor.

Enakshi Saha
Postdoctoral Researcher
Harvard T.H. Chan School of Public Health
esaha@hsph.harvard.edu

CP14

Principal Tradeoff Analysis and Disc Game Embedding of Functional Form Games

Functional form games model a wide range of competitive systems studied in biology, game theory, and artificial intelligence. A functional form game is a two-player game whose outcome is determined by the player's attributes via a performance (payout) function. Player attributes model physical characteristics, parameters controlling their behavior, or hyper-parameters determining training protocols. The performance function f accepts a pair of attribute vectors x, y and returns the payout to the first agent, $f(x, y)$. Performance is always skew-symmetric. We illustrate that, under loose assumptions on f, performance can be decomposed into a linear combination of low dimensional embeddings. Each embedding is planar and is equipped with simple game dynamics, namely, disc game dynamics. A disc game, a continuous extension of rock-paper-scissors, is the simplest cyclic functional form game. Thus the decomposition into disc games represents a decomposition into cyclic modes associated with principal strategic tradeoffs. We introduce a sequence of embedding methods for specific function classes and develop an associated approximation theory. The decomposition into disc games is closely related to low-rank matrix decomposition, where the principal tradeoffs are analogous to feature directions in principal component analysis. We conclude by showing that the sequence of embeddings enable informative low-dimensional visualization and present examples for a variety of games.

Alexander Strang
University of Chicago
alexstrang@uchicago.edu

David Rosenbluth, David Sewell, Kevin Alcedo
Lockheed Martin Artificial Intelligence Center, U.S.
david.rosenbluth@lmco.com, david.r.sewell@lmco.com, kevin.alcedo@lmco.com

CP14

Reproducing Activation Function for Deep Learning

The speaker will propose reproducing activation functions (RAFs) motivated by applied and computational harmonic analysis to improve deep learning accuracy for various applications ranging from computer vision to scientific computing. The idea is to employ several basic functions and their learnable linear combination to construct neuron-wise data-driven activation functions for each neuron. Armed with RAFs, neural networks (NNs) can reproduce traditional approximation tools and, therefore, approximate target functions with a smaller number of parameters than traditional NNs. In NN training, RAFs can generate neural tangent kernels with a better condition number than traditional activation functions lessening the spectral bias of deep learning. As demonstrated by extensive numerical tests, the proposed RAFs can facilitate the convergence of deep learning optimization for a solution with higher accuracy than existing deep learning solvers for audio/image/video reconstruction, PDEs, and eigenvalue problems. With RAFs, the errors of audio/video reconstruction, PDEs, and eigenvalue problems are decreased by over 14%, 73%, 99%, respectively, compared with baseline, while the performance of image reconstruction increases by 58%.

Chunmei Wang
University of Florida
chunmei.wan@ufl.edu

CP14

Fair Data Representation for Machine Learning at the Pareto Frontier

As machine learning powered decision-making becomes important in our daily lives, it is imperative to strive for fairness of the underlying data processing. We propose a pre-processing algorithm for fair data representation via which L^2-objective supervised learning algorithms result in estimations of the Pareto frontier between prediction error and statistical disparity. Particularly, we apply the optimal affine transport to approach the post-processing Wasserstein-2 (W_2) barycenter characterization of the optimal fair L^2-objective supervised learning via a pre-processing data deformation. We call it W_2 pseudo-barycenter. Furthermore, we show that the W_2 geodesics from learning outcome marginals to their barycenter characterizes the Pareto frontier between L^2-loss and total W_2 distance among the marginals. Thereby, McCann interpolation generalizes the pseudo-barycenter to a family of data representations via which L^2-objective supervised learning algorithms estimate the Pareto frontier. Numerical simulations underscore the advantages: (1) the pre-processing step is compositive with arbitrary L^2-objective learning methods and unseen data; (2) the representation protects data privacy by preventing access to sensitive knowledge;

(3) the optimal affine maps are computationally efficient on high-dimensional data; (4) experiments shed light on the fairness of L^2-objective unsupervised learning.

Shizhou Xu
University of California Davis, Department of Mathematics,
TETRAPODS Institute of Data Science
shzxu@ucdavis.edu

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

CP15

A Unified Theory of Error Feedback and Variance Reduction Mechanisms for Biased and Unbiased Compression in Distributed Optimization

In the big data era, it is necessary to rely on distributed computing settings. We consider the client-server model, in which several parallel workers communicate back and forth with a distant server. Communication, which can be costly and slow, is the main bottleneck. A popular idea is to make use of compression, to reduce the number of bits sent within each communication round of iterative methods. There are two classes of compression operators and separate algorithms making use of them. In the case of unbiased random compressors, the DIANA algorithm of Mishchenko et al. (2019), which implements a variance-reduction technique, is the current state of the art. In the case of biased and contractive compressors, the EF21 algorithm of Richtrik et al. (2021), which implements an error-feedback mechanism, is the current state of the art. These two classes are distinct and rely on different analyses. We unify them into a single framework: we propose a general class of compressors with two parameters, the bias and the variance, as well as a new algorithm exploiting them, recovering DIANA and EF21 as particular cases, with proved linear convergence. Our general approach inherits the best of the two worlds: biased compressors can be used, and independent randomness at the compressors allows to mitigate the effects of compression, with the convergence rate improving when the number of workers is large. This is the first time that an algorithm with all these features is proposed.

Laurent Condat
KAUST
laurent.condat@kaust.edu.sa

Kai Yi, Peter Richtárik
King Abdullah University of Science and Technology (KAUST)
kai.yi@kaust.edu.sa, richtarik@gmail.com

CP15

Efficient Distributionally Robust Learning via Unbiased Gradient Estimation

Seeking to improve generalization of mathematical learning models, we consider a new statistical learning approach called distributionally robust learning (DRL) that utilizes difficult-to-solve minimax formulations. We devise a new algorithm to solve these formulations that applies stochastic gradient descent (SGD) to the outer minimization problem. Our algorithm efficiently estimates the gradient of the inner maximization problem through multi-level Monte Carlo randomization. We establish theoretical results that shed light on why standard gradient estimators fail, and then leverage these theoretical results to establish the optimal parameterization of the gradient estimators of our approach that balances a fundamental tradeoff between computation time and stochastic error. We rigorously establish convergence to a near-optimal solution under standard regularity assumptions and, for strongly convex losses, match the best known $O(\epsilon^{-1})$ rate of convergence of SGD for standard stochastic optimization. Computational experiments demonstrate that our DRL approach yields significant benefits over previous work.

Soumyadip Ghosh, Mark Squillante
IBM Research
ghoshs@us.ibm.com, mss@us.ibm.com

CP15

Stochastic Gradient Descent for Inverse Problems in Banach Spaces

Stochastic gradient descent (SGD) and its variants are among the most promising modern optimisation techniques. At each iteration they use a small subset of data to stochastically estimate the gradient of the objective. Resulting methods scale well, making them a popular choice for a large number of problems, often outperforming traditional methods. SGD methods have seen growing use for inverse problems, and are mostly used in inner product spaces. Convergence for linear inverse problems is in these settings well understood. However, Banach spaces require different techniques and are the native domain in many problems, such as parameter identification in elliptic PDEs. Full batch gradient descent methods for inverse problems in Banach spaces typically follow Landweber type iterations, with several differences from Hilbert space settings. Namely, since the gradient lies in the dual space, duality mappings are needed to conduct the algebra. The iterates are thus non-linear and Bregman divergence is used for the convergence analysis. In this work we present a mathematical framework and analysis for SGD in Banach spaces for linear and non-linear inverse problems. Analysis in the Banach space setting presents unique challanges, requiring novel mathematical tools. This is achieved by combining insights from Hilbert space theory with approaches from modern optimisation. The developed theory and algorithms open doors for a wide range of methods that we aim to study in future work.

Zeljko Kereta
University College London
z.kereta@ucl.ac.uk

CP15

On the Inversion-Free Newton's Method and Its Applications

In this paper, we survey the recent development of inversion-free Newtons method and its applications in estimating parameters of models such as linear and logistic regression. A detailed review of the existing methodology is provided, along with comparisons of various competing algorithms. We provide numerical examples that highlight some deficiencies of existing approaches and demonstrate how the inversion-free methods can improve performance. Motivated by the works of [CGBP20], [WZM18], and [WYS19], we provide a unified subsampling framework that can be combined with the inversion-free Newtons method to estimate model parameters including those

of linear and logistic regression. Numerical examples are provided for illustration.

Thai H. Nguyen
Université Laval
thaibopy@gmail.com

Huy Chau
University of Manchester
huy.chau@manchester.ac.uk

Justin Kirkby
Georgia Institute of Technology
jkirkby3@gatech.edu

Duy Nguyen
Department of mathematics
Marist College
NDUCDUY@GMAIL.COM

Nhu N. Nguyen
University of Connecticut, Storrs
nguyen.nhu@uconn.edu

Dang H. Nguyen
University of Alabama
dangnh.maths@gmail.com

CP15

Theory of Stochastic Gradient Descent

Stochastic Gradient Descent (SGD) is a foundational algorithm in machine learning and other data-driven disciplines, which has resulted in an extensive effort to understand SGD's behavior in a variety of contexts such as nonconvex objectives, non-smooth objectives, and unbounded noise of the stochastic gradients. In this talk, we begin by presenting canonical examples in machine learning that motivate the need for more sophisticated SGD theory. Then, we will state recent results that address the gaps in SGD theory.

Vivak Patel
Department of Statistics, University of Wisconsin-Madison
vivak.patel@wisc.edu

CP15

Hierarchical Hessian Representations for Scalable Second-Order Deep NN Training Methods

The main computational obstacles to the successful development of Newton-type methods in deep neural network training, with large data and parameter dimensions, are the operations on the Hessian. While subsampling the Hessian can reduce the cost of managing large data dimensions, the storage of the formally dense matrix and the solution of the Newton system are prohibitive for large parameter dimensions. The key to successful Newton methods is to approximate the Hessian in a way to make its computation, storage, and inversion manageable. Hessian-free Newton methods are a natural way of implementing second-order methods in a scalable fashion but the preconditioning of the linear system remains the bottleneck. In this work, we present a hierarchically low rank matrix representation of the Hessian as a tunable preconditioner representation that balances accuracy on one hand and storage and arithmetic complexity on the other. The setting of our training problem is distributed multi-GPU systems, with the data distributed among the GPUs, and the hierarchical Hessian is updated incrementally by low rank updates from the individual GPUs. We show the favorable scaling results of generalized Gauss-Newton methods for training ResNet-32 and ResNet-50 models on CIFAR data, and PINN models on a variety of PDEs and inverse PDEs. We also describe the role of the Hessian in identifying and quantifying uncertainties in trained model features and parameters.

George M. Turkiyyah
KAUST
George.Turkiyyah@kaust.edu.sa

Stefano Zampini
KAUST, Saudia Arabia
KAUST, Saudia Arabia
stefano.zampini@kaust.edu.sa

Bernard Ghanem
KAUST
bernard.ghanem@kaust.edu.sa

David E. Keyes
King Abdullah University of Science and Technology (KAUST)
david.keyes@kaust.edu.sa

CP16

Optimality and Complexity of Classification by Random Projection

The generalization error of a classifier is related to the complexity of the set of functions among which the classifier is chosen. We study a family of low-complexity classifiers consisting of thresholding a random one-dimensional feature. The feature is obtained by projecting the data on a random line after embedding it into a higher-dimensional space parametrized by monomials of order up to k. More specifically, the extended data is projected n-times and the best classifier among those n, based on its performance on training data, is chosen. We show that this type of classifier is extremely flexible and given full knowledge of the class conditional densities, the error of these classifiers would converge to the optimal (Bayes) error as k and n go to infinity. On the other hand, if only a training dataset is given, we show that the classifiers will perfectly classify all the training points as k and n go to infinity. We also bound the generalization error of our random classifiers. In general, our bounds are better than those for any classifier with VC dimension greater than O(ln n). In particular, our bounds imply that, unless the number of projections n is extremely large, there is a significant advantageous gap between the generalization error of the random projection approach and that of a linear classifier in the extended space. Thus, there is a potentially large gain in generalization properties by selecting parameters at random, rather than optimization.

Evzenie Coupkova
Purdue University
ecoupkov@purdue.edu

Mireille Boutin
Purdue University
U.S.

mboutin@purdue.edu

CP16

Robust Ensemble Estimation

The theory of optimally weighted ensemble estimation is a general theory that can be applied to any ensemble of estimators to achieve improved convergence rates as long as the bias and variance of the base estimator can be expressed in a specific way. In particular, this theory has been applied to simple plug-in estimators of information theoretic quantities such as entropy, mutual information, and entropy to achieve the parametric rate when the densities are sufficiently smooth. However, the optimal weights for this ensemble estimation approach currently depend on knowing explicit expressions for the bias and variance. We empirically show that naively learning the weights by assuming the form of the bias and variance results in a bias reduction even when the bias differs from the assumed form, greatly expanding the potential field of applications. We also perform extensive experiments to demonstrate the robustness of the approach to the choice of tuning parameters.

Mina Mahbub Hossain
Utah State University
mahbub.hossain@usu.edu

CP16

Learning with Multidimensional and Multiscale Scientific Data: Combining Probabilistic Graphical Models and Deep Learning Architectures

The growth in data size and complexity coming from experiments across the U.S. Department of Energy overwhelms current statistical and learning approaches for analysis and understanding. Consequently, scientific discoveries are falling behind the pace of data generation. In several scientific fields, researchers combine simulations and experiments to analyze data. Data acquired from different sources are often complementary and provide different types or levels of information. These levels of information may come from different spatial dimensions, scales, or time steps. In this case, even though multimodal, multiscale, and temporal correlated information may exist theoretically, finding that feature correlation automatically and using it for learning and analysis is challenging. The use of learning algorithms to analyze scientific data becomes even more challenging due to the low data regime (lack of training data). In this lecture, I will talk about my recent work on combining probabilistic graphical models (PGM) and deep learning models to overcome these challenges. Specifically, I will discuss current developments on new Markov random fields (MRF) maximum a posteriori (MAP) inference based on alternating direction method of multipliers (ADMM) for arbitrary (higher-order multilabel) models. Finally, I will give a general overview on the challenges and opportunities around developing machine learning algorithms for multidimensional and multiscale scientific data.

Talita Perciano, Matthew Avaylon, Robbie Sadre, Zhe Bai
Lawrence Berkeley National Laboratory
tperciano@lbl.gov, mavaylon@lbl.gov, rssadre@ucdavis.edu, zhebai@lbl.gov

CP16

Self-Supervised Metric Learning in Multi-View Data: A Downstream Task Perspective

Self-supervised metric learning has been a successful approach for learning a distance from an unlabeled dataset. The resulting distance is broadly useful for improving various distance-based downstream tasks, even when no information from downstream tasks is utilized in the metric learning stage. To gain insights into this approach, we develop a statistical framework to theoretically study how self-supervised metric learning can benefit downstream tasks in the context of multi-view data. Under this framework, we show that the target distance of metric learning satisfies several desired properties for the downstream tasks. On the other hand, our investigation suggests the target distance can be further improved by moderating each directions weights. In addition, our analysis precisely characterizes the improvement by self-supervised metric learning on four commonly used downstream tasks: sample identification, two-sample testing, k-means clustering, and k-nearest neighbor classification. When the distance is estimated from an unlabeled dataset, we establish the upper bound on distance estimations accuracy and the number of samples sufficient for downstream task improvement. Finally, numerical experiments are presented to support the theoretical results in the paper.

Shulei Wang
University of Illinois at Urbana-Champaign
shuleiw@illinois.edu

CP16

Spectral Machine Learning Theory

We investigate the spectral properties during training of linear-width feed-forward neural networks on both synthetic data and real-world data. Under this linear-width regime, we show empirically that the spectrum of the weight matrix is invariant when trained by gradient descent with a small constant learning rate, and that the changes in the weight matrix in both operator and Frobenius norms are of constant order, implying that the variance for each parameter vanishes in the limit. Our implementation also empirically shows invariance for both the conjugate and neural tangent kernels. In addition, we find the same empirical characteristics for models trained with moderately sized mini-batch stochastic gradient descent with a small learning rate. This suggests that these trained neural networks in the linear-width regime still reside in the kernel regime. We further show an escape from the kernel regime by a combination of choices in the weight initialization matrix coupled with tuning the batch size and learning rate. Indeed, spectral properties such as the bulk, spikes, and heavy-tailed distributions directly correlate with tuning such choices, and this tuning permits us to generate each spectrum on shallow networks. With a two-layer teacher-student model on synthetic data, we can derive different kinds of spectra using differing training strategies, and we show a close relationship between different spectral behaviors and the generalization of the trained networks.

Tony Chiang
Pacific Northwest National Laboratory
tony.chiang@pnnl.gov

Zhichao Wang
UCSD
zhw036@ucsd.edu

Andrew Engel
PNNL

andrew.engel@pnnl.gov

Anand Sarwate
Rutgers University
ads221@soe.rutgers.edu

Ioana Dumitriu
UCSD
idumitriu@ucsd.edu

CP17

Fast Computation of Matrix Function-Based Centrality Measures for Layer-Coupled Multiplex Networks

Centrality measures identify and rank the most influential entities of complex networks. In this talk, we generalize matrix function-based centrality measures, which consider subgraphs of all sizes and interpolate between local degree and global eigenvector centrality, to layer-coupled multiplex networks. The layers of these networks can reflect different relationships and interactions between entities or changing interactions over time. We use the supra-adjacency matrix as network representation, which has already been used to generalize eigenvector centrality to temporal and multiplex networks. As the explicit evaluation of the involved matrix function expressions becomes infeasible for medium to large-scale networks, we present highly efficient approximation techniques, which rely on Krylov subspace methods, Gauss quadrature, and stochastic trace estimation. We present extensive numerical studies on synthetic and real-world multiplex networks with a focus on urban public transport systems. Our experiments corroborate the linear computational complexity of the employed numerical methods in terms of the network size that is theoretically indicated under the assumption of sparsity in the supra-adjacency matrix. This excellent scalability allows the efficient treatment of large-scale networks with the number of node-layer pairs of order 10^7 or higher.

Kai Bergermann, Martin Stoll
Department of Mathematics,
Technische Universitaet Chemnitz
kai.bergermann@math.tu-chemnitz.de,
martin.stoll@math.tu-chemnitz.de

CP17

Nonbacktracking Spectral Clustering of Nonuniform Hypergraphs

Spectral methods offer a tractable, global framework for clustering in graphs via eigenvector computations on graph matrices. Hypergraph data, in which entities interact on edges of arbitrary size, poses challenges for matrix representations and therefore for spectral clustering. We study spectral clustering for nonuniform hypergraphs based on the hypergraph nonbacktracking operator. After reviewing the definition of this operator and its basic properties, we prove a theorem of Ihara-Bass type to enable faster computation of eigenpairs. We then propose an alternating algorithm for inference in a hypergraph stochastic blockmodel via linearized belief-propagation, offering proofs that both formalize and extend several previous results. We perform experiments in real and synthetic data that underscore the benefits of hypergraph methods over graph-based ones when interactions of different sizes carry different information about cluster structure. Through an analysis of our algorithm, we pose several conjectures about the limits of spectral methods and detectability in hypergraph stochastic blockmodels writ large.

Philip S. Chodrow
University of California, Los Angeles
phil@math.ucla.edu

Jamie Haddock
Harvey Mudd College
320 E Foothill Blvd, Claremont, CA 91711
jhaddock@g.hmc.edu

Nicole Eikmeier
Grinnell College
eikmeier@grinnell.edu

CP17

Regression on Graphs via Graph-Informed Neural Networks

Graphs are used in network analysis to study many different phenomena, such as transportation systems, epidemic spread, and social interactions. Recently, new key contributions have been proposed by the neural network (NN) community, extending deep learning (DL) approaches to graphs via the so-called graph neural networks (GNNs) and the more recent graph convolutional networks (GCNs). Despite the good results of GCNs in many applications, some challenges still exist: (i) building deep GCNs with good performances; (ii) building GCNs scalable for large graphs. In this talk, we present a new type of layer designed for regression tasks on graphs, a framework for which GCNs are not well suited and the use of multi-layer perceptrons (MLPs) is preferred. Our new graph layer exploits the graph structure to improve the NN training (compared to an MLP). Moreover, it allows the building of deep NNs and it is scalable for large graphs. We call this layer a graph-informed (GI) layer [Berrone et al., Graph-Informed Neural Networks for Regressions on Graph-Structured Data, Mathematics, 2022, https://doi.org/10.3390/math10050786]. We report the results of numerical experiments to show the potentiality of the graph-informed NNs (GINNs). In particular, we analyze the regression abilities of the GINNs on maximum-flow regression problems, comparing the results with MLPs performances on the same problem.

Francesco Della Santa, Stefano Berrone, Antonio Matropietro
Politecnico di Torino
Dipartimento di Scienze Matematiche
francesco.dellasanta@polito.it, stefano.berrone@polito.it,
antonio.mastropietro@polito.it

Sandra Pieraccini
Dipartimento di Scienze Matematiche
Politecnico di Torino, Italy
sandra.pieraccini@polito.it

Francesco Vaccarino
Politecnico di Torino
francesco.vaccarino@polito.it

CP17

ExSpliNet: How Kolmogorov Would Design an Interpretable and Expressive Neural Network, Walking in a B-Spline Forest

Solving problems that require the approximation of data in

high-dimensional spaces is computationally extremely challenging. In this talk we present ExSpliNet, an interpretable and expressive neural network model that combines ideas of Kolmogorov neural networks, ensembles of probabilistic trees, and multivariate B-spline representations. The new model is a feasible generalization of the B-spline neural network (BSNN) model towards high-dimensional data relying on a Kolmogorov-like structure to avoid the use of high-variate tensor-product B-splines while still maintaining expressive power. Furthermore, we carry out a theoretical study of the universal approximation properties of ExSpliNet. The main ingredients of the proof are the Kolmogorov superposition theorem and classical approximation estimates for multivariate splines. Finally, we illustrate the suitability of the proposed model to address data-driven function approximation and to face differential problems, in the spirit of physics-informed neural networks (PINNs). We also show the general applicability of the model for classical machine learning tasks like image classification and regression.

Daniele Fakhoury
University Rome, Tor Vergata
daniele.fak@gmail.com

Hendrik Speleers
Department of Mathematics
University of Rome
speleers@mat.uniroma2.it

Emanuele Fakhoury
University Rome, Tor Vergata
emanuelefak@gmail.com

CP17

Topological Analysis of Temporal Hypergraph Data

Hypergraphs extend graphs to include high-order edges recording multi-way relationships. Graphs are hypergraphs with uniform edge size two, but hypergraphs can have edges of arbitrary size. They are closely related to topological structures like abstract simplicial complexes, which are themselves hypergraphs including all sub-edges, and can have non-trivial topological properties. Simplicial homology identifies the topological structure of hypergraph data by measuring the dimensions of their homology groups (Betti sequences), and thus their open cycles of different dimension k, including connected components as a special case for $k = 0$. Hypergraph structures commonly carry additional data, including time intervals on edges, so the topological properties of such temporal hypergraphs can change over time. A sliding window over the temporal intervals of a THG identify a temporal sequence of sub-hypergraphs, and thus a changing sequence of Betti sequences. While useful, a more sophisticated approach called zigzag persistent homology links the sequential sub-hypergraphs to not only identify common structures between sample windows, but how much they persist over all the data. We demonstrate our method first on synthetic data, and then on a collection of Reddit threads, tracking users against threads in the context of their time stamps. A complex temporal pattern is revealed for dimensions $k = 0, 1, 2$. Comparison to non-topological hypergraph properties is provided.

Cliff Joslyn
Pacific Northwest National Laboratory
cliff.joslyn@pnnl.gov

Audun D. Myers
Michigan State University
Department of Mechanical Engineering
myersau3@msu.edu

Purvine Emilie, Shapiro Madelyn
Pacific Northwest National Laboratory
emilie.purvine@pnnl.gov, madelyn.shapiro@pnnl.gov

CP18

Nonlocal Models for Deep Neural Networks

Recently, continuous-depth models of neural networks have shown a series of advantages in machine learning, which include memory costs that do not increase with the number of layers (a major bottleneck in training deep models), parameters efficiency, as well as obtaining continuous time series models that can naturally incorporate data at arbitrary times. In this talk I will present some continuous-depth models that also incorporate nonlocality with respect to the layer and discuss their performance and advantages. The results are backed by theoretical studies into well-posedness of solutions for the integro-differential models, as well as stability results which are important in designing robust DNNs that can withstand adversarial attacks

Animesh Biswas
University of Nebraska-Lincoln
abiswas2@unl.edu

CP18

Examining Stiffness in ResNets through Interpretation as Discretized Neural ODEs

Neural Ordinary Differential Equations (NODEs) emerge as the limit of Residual Neural Networks (ResNets) as the number of layers tends to infinity. While NODEs are successfully applied in practice to a range of machine learning problems, studying their behavior can provide insights into the role of depth in network architecture. We expect that mathematical insights and legacy knowledge of ODE analysis can be applied to NODEs to better understand the behavior of neural networks, improve their generalization, and develop more efficient training algorithms. Drawing from ODE theory, we focus on the concept of stiffness and consider the analogue of stiffness for ResNets. We first develop a heuristic for quantifying the stiffness of a ResNet given an input training set. We then study the use of this stiffness measure as a penalizing regularizer during training, and examine its effects on generalization.

Joshua Hudson
Johns Hopkins University Applied Physics Lab
jlhudso@sandia.gov

Khachik Sargsyan, Marta D'Elia
Sandia National Laboratories
ksargsy@sandia.gov, marti.delia@gmail.com

Habib N. Najm
Sandia National Laboratories
Livermore, CA, USA
hnnajm@sandia.gov

CP18

Differentiable Computational Fluid Dynamics As a

Layer

With ever accelerating advances in machine learning, fluid dynamicists have begun to intersperse machine learning techniques in fluid dynamics simulations to achieve in-depth control of the physical process, acceleration by dimensionality reduction, or more enhanced active learning. In adjacent fields, such as optimization, researchers have gone a step further, and made their optimization problems differentiable, and started to embed them as layers in their deep architectures. This has not only enabled them to improve computational performance, but also made it easier to solve previously hard-to-solve optimization problems. In fluid dynamics such approaches have so far been out of reach, as differentiability implied the need for hand-written adjoints, or specialized software which is often unable to differentiate hybrid parallelism, and GPU-computations while incurring significant performance degradation. In this lecture we will present how, enabled by compiler-based approaches to automatic differentiation, we are able to overcome these constraints and seamlessly embed simulation layers inside of deep networks with simulation structures abstracted away as layers in our end-to-end learned architecture. A presented example of this are Neural PDEs, a fully end-to-end learned system where the simulation's PDE solver is used on a coarse-grained latent space. All presented approaches are directly generalizable to other simulation-based fields.

Ludger Paehler, Josef Winter, Julian Stang, Tim Gymnich, Nikolaus Adams
Technical University of Munich
ludger.paehler@tum.de, josef.winter@tum.de, julian.stang@iwb.tum.de, tim.gymnich@tum.de, nikolaus.adams@tum.de

CP18

Memory-Efficient Neural ODE Based on High-Level Adjoint Differentiation

We present a memory-efficient neural ODE framework PNODE based on high-level adjoint algorithmic differentiation. It is implemented using PyTorch and PETSc, one of the most commonly used portable, scalable scientific computing libraries. By leveraging discrete adjoint time integrators and advanced checkpointing strategies tailored for these integrators, PNODE can provide a balance between memory and computational costs while computing the gradients consistently and accurately. We demonstrate the performance through numerical experiments on image classification, continuous normalizing flow, and time series regression. We show that PNODE achieves the highest memory efficiency when compared with other reverse-accurate methods. We also show PNODE enables the application of implicit time integration methods that are desired for stiff dynamical systems.

Hong Zhang
Argonne National Laboratory
hongzhang@anl.gov

CP19

New Block Generalized Averaged Gauss Quadrature Rules

Golub and Meurant presented the way of using the symmetric block Lanczos algorithm for calculating the block Gauss quadrature rules for approximating of matrix functions $w^T f(A)u$, where A is a large square matrix, u and v are block-vectors and f is a function. We describe a new block quadrature rules . These rules can be calculated using the symmetric or nonsymmetric block Lanczos algorithms for calculating error estimates for computed quantities, and shows how to achieve more accuracy than standard block Gauss rules for the same computational effort. Our methods are based on block generalizations of the generalized averaged Gauss quadrature rules that were recently proposed by Spalevic.The new representation suggested by Spalevic is a new averaged Gauss quadrature rule that has higher degree of exactness and the same number of nodes as the averaged rule proposed by Laurie. Numerical experiments reported in this paper show the latter averaged rule to yield higher accuracy than Lauries averaged rule for smooth integrals and, therefore, also can be used to estimate the error in Gauss rules and to approximate integrals. In addition, We describe how to use a symmetric or nonsymmetric adjacency matrix for a network to evaluate functions as applications.

Hessah Alqahtani
Department of Mathematics, Faculty of Science and Arts, King
hfq1981@gmail.com

Lothar Reichel
Kent State University
reichel@math.kent.edu

Sara Balkhram
Department of Mathematics, Faculty of Science and Arts, King
sbalakhram0002@kau.edu.sa

CP19

Special Matrices in Data Smoothing and Analytics for Machine Learning

This lecture will focus on special matrices of multidimensional data analytics and machine learning algorithms. Interesting examples of mathematical challenges will be addressed with emphasis on some special domain area applications.

Tor A. Kwembe
Chicago State University, Department of Mathematics and Computer Science
tor.a.kwembe@jsums.edu

CP19

Interpretability of Convolutional Neural Networks Using a Singular Value Decomposition of its Weight Tensors

The ability of convolutional neural networks to classify images quickly and accurately has revolutionized computer vision, but the complexity and size of these networks can make their decision process seem impenetrable. This lack of transparency limits the confidence we can comfortably place in the network. In this talk we approach interpretability as a decomposition problem. We perform a simple matrization on the weight tensors of the convolutional layers and apply the singular value decomposition to the matrix representations of each fully connected and convolutional layer. This process exposes the dynamics of each linear layer. Latent features are discoverable by examining the projection or signal of the input onto the singular directions, and their significance is quantified by the corresponding rescaling by the singular values. We

then apply the techniques of optimized feature visualization, exemplary image identification, and saliency masks to uncover the semantic meaning of the features at each layer. Using benchmark datasets and models, we illustrate our approach by explaining their classification process. We then apply our methodology to nuclear forensics data and discuss more general applications.

Brenda Praggastis, Davis Brown, Emilie Purvine, Madelyn Shapiro
Pacific Northwest National Laboratory
brenda.praggastis@pnnl.gov, davis.brown@pnnl.gov, emilie.purvine@pnnl.gov, madelyn.shapiro@pnnl.gov

Bei Wang
University of Utah
beiwang@sci.utah.edu

CP19

Statistical Leverage Scores: Complexity, Algorithms and Applications

Statistical leverage scores are important quantities in statistics, data science, and quantum mechanical applications. In the last two decades, they have also made an impact in the design of numerical algorithms. Given a data matrix A with size $n \times d$, we discuss algorithms for computing the leverage scores of A, focusing on the "tall-and-skinny' case, that is, when the number of data points n is much larger than the number of features d ($n \gg d$). We first revisit a straightforward algorithm, which directly computes the Gramian $A^\top A$ and its pseudoinverse $(A^\top A)^+$, and then returns the squared Euclidean row norms of the matrix $A((A^\top A)^+)^{\frac{1}{2}}$. In the very tall and very sparse regime, i.e. when $d = o(n^{1/3})$ and A has $O(1)$ nonzeros per row, this simple algorithm is optimal, as it runs in $O(n)$ time. For matrices with $O(\log n)$ non-zeros per row, it still optimal up to log factors. With these observation in place, we then revisit the state-of-the-art in different size and sparsity regimes, including random sampling and random projection based approaches. We then describe some improved algorithms which are especially effective for low-rank and approximately low-rank matrices. We also discuss connections of leverage scores computations with other problems in Numerical Linear Algebra, which can be interesting for a Fine-Grained Complexity analysis.

Aleksandros Sobczyk
IBM Research Europe
ETH Zürich
obc@zurich.ibm.com

MS1

Alternating the Population and Agent Control via Two Neural Networks to Solve High-Dimensional Stochastic Mean Field Games

We present APAC-Net, an alternating population and agent control neural network for solving stochastic mean field games (MFGs). Our algorithm is geared toward high-dimensional instances of MFGs that are beyond reach with existing solution methods. We achieve this in two steps. First, we take advantage of the underlying variational primal-dual structure that MFGs exhibit and phrase it as a convex-concave saddle point problem. Second, we parameterize the value and density functions by two neural networks, respectively. By phrasing the problem in this manner, solving the MFG can be interpreted as similar to training a generative adversarial network (GAN). We show the potential of our method on up to 100-dimensional MFG problems.

Alex Lin
UCLA
atlin@math.ucla.edu

MS1

Parametric Fokker-Planck Equations

Based on the theory of Wasserstein gradient flows, we develop and analyze a numerical method proposed for solving high-dimensional Fokker-Planck equations (FPE). The gradient flow structure of FPE allows us to derive a finite-dimensional ODE by projecting the dynamics of FPE onto a finite-dimensional parameter space whose parameters are inherited from certain generative model such as normalizing flow. We design a bi-level minimization scheme for time discretization of the proposed ODE. Such algorithm is sampling-based, which can readily handle the computations in high-dimensional space. Moreover, we establish theoretical bounds for the asymptotic convergence analysis as well as the error analysis for our proposed method. Several numerical examples are provided to illustrate the performance of the proposed algorithm and analysis.

Shu Liu
University of California, Los Angeles, U.S.
shuliu@math.ucla.edu

MS1

Efficient Natural Gradient Method for Large-scale Optimization Problems

Large-scale optimization is at the forefront of modern data science, scientific computing, and applied mathematics with areas of interest, including high-dimensional PDE, inverse problems, machine learning, etc. First-order methods are workhorses for large-scale optimization due to modest computational cost and simplicity of implementation. Nevertheless, these methods are often agnostic to the structural properties of the problem under consideration and suffer from slow convergence, being trapped in bad local minima, etc. Natural gradient descent is an acceleration technique in optimization that takes advantage of the problem's geometric structure and preconditions the objective function's gradient by a suitable "natural" metric. Hence parameter update directions correspond to the steepest descent on a corresponding "natural" manifold instead of the Euclidean parameter space rendering a parametrization invariant descent direction on that manifold. Despite its success in statistical inference and machine learning, the natural gradient descent method is far from a mainstream computational technique due to the computational complexity of calculating and inverting the preconditioning matrix. This work aims at a unified computational framework and streamlining the computation of a general natural gradient flow via the systematic application of efficient tools from numerical linear algebra. We obtain efficient and robust numerical methods for natural gradient flows without directly calculating, storing, or inverting the dense preconditioning matrix. We treat Euclidean, Wasserstein, Sobolev, and FisherRao natural gradients in a single framework for a general loss function.

Levon Nurbekyan
UCLA

levonnurbekian@gmail.com

MS1

Dynamic Behavior for a Gradient Algorithm with Energy and Momentum

We investigate a novel gradient algorithm, using both energy and momentum (called AGEM), for solving general non-convex optimization problems. The solution properties of the AGEM algorithm, including uniformly boundedness and convergence to critical points, are examined. The dynamic behavior is studied through analysis of a high-resolution ODE system. Such ODE system is nonlinear and obtained by taking the limit of the discrete scheme while keeping the momentum effect through a rescale of the momentum parameter. In particular, we show global well-posedness of the ODE system, time-asymptotic convergence of solution trajectories, and further establish a linear convergence rate for objective functions satisfying the Polyak-Lojasiewicz condition.

Xuping Tian
Iowa State University
xupingt@iastate.edu

MS2

A Randomized Algorithm for Computing the Action of a Matrix Function on a Vector

This talk is concerned with the computation of the action of a matrix function $f(A)$, such as the matrix exponential or the matrix logarithm, on a vector b. For a general matrix A, this can be done by computing the projection of A onto a suitable Krylov subspace. Such projection is usually computed by forming an orthonormal basis of the Krylov subspace using the Arnoldi method. In this talk, we propose to compute (non-orthonormal) bases in a faster way and to use a fast randomized algorithm for least squares problems to compute the projection of A onto the Krylov subspace. We present some numerical examples which show that our proposed algorithms can be faster than the standard Arnoldi method while achieving comparable accuracy.

Alice Cortinovis
EPFL
alicecrvs@gmail.com

Daniel Kressner
EPFL Lausanne
Mathicse
daniel.kressner@epfl.ch

Yuji Nakatsukasa
University of Oxford
nakatsukasa@maths.ox.ac.uk

MS3

Life Beyond Orthogonality: Sparse Recovery in Bounded Riesz Systems - Theory and Applications

In compressive sensing and sparse recovery, a wide class of popular measurement schemes can be analyzed within the framework of randomly sampled bounded orthonormal systems. This setting includes measurement matrices such as subsampled isometries, partial Fourier matrices, and sampling matrices associated with orthogonal polynomials. Yet, the orthogonality assumption is too restrictive in applications where the sampling matrix does not have a trivial covariance. In this talk, we will discuss how to address this issue by working with randomly sampled bounded Riesz systems. The main theoretical tool of our analysis is a new upper bound for the expectation of the supremum of a Bernoulli process associated with the restricted isometry constant of the random matrix of interest. Using this bound, we will illustrate sparse recovery guarantees in bounded Riesz systems based on the restricted isometry and the robust null space property, an application to local coherence-based sampling schemes, and the extension to the weighted sparsity setting. Going beyond orthogonality, the additional flexibility of bounded Riesz systems allows one to analyze a wider class of sparse recovery problems. Here, we will illustrate applications in scientific computing such as numerical methods for high-dimensional partial differential equations.

Simone Brugiapaglia
Concordia University
simone.brugiapaglia@concordia.ca

MS3

Geometric Scattering and Applications

The scattering transform is a wavelet-based model of convolutional neural networks (CNNs) introduced for functions defined on Euclidean space by S. Mallat. It differs from traditional CNNs by using predesigned filters rather than filters learned from training data, which leads to a network with provable stability and invariance guarantees. In my talk, I will focus on variations of the scattering transforms for data sets with non-Euclidean geometric structure such as graphs or manifolds. These networks use wavelets constructed via an appropriate Laplacian or diffusion operator and have stability and invariance guarantees analogous to their Euclidean predecessor. After introducing these networks, I will then talk about modifications of the graph scattering transform which can increase numerical performance and also work using the graph scattering transform as part of a molecular generation network.

Michael Perlmutter
University of California, Los Angeles
perlmutter@g.ucla.edu

Yimeng Min
Cornell University
min@cs.cornell.edu

Frederik Wenkel
University of Montreal
frederik.wenkel@umontreal.ca

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

Mathew Hirn
Michigan State University
mhirn@msu.edu

MS3

Sparse Random Features - Theoretical Results, Algorithms, and Applications

Signal decomposition and multiscale signal analysis provide many useful tools for time-frequency analysis. We propose

a signal representation algorithm using a sparse and randomized short-time Fourier transform. This builds from the recent sparse random feature expansion for approximating high-dimensional functions. The randomization is both in the time window location and the frequency sampling, which lowers the overall sampling and computational cost. Sparse optimization extracts a sparse time-frequency representation, which has the added benefit of forming a simple decomposition due to the sharpening of the spectrograph. Experiments on synthetic data show that the decomposition closely resembles the multiscale properties of the data and tests on real datasets show robustness. Comparisons show that our proposed approach performs better or comparable to other state-of-the-art or popular methods.

Giang Tran
University of Waterloo
giang.tran@uwaterloo.ca

Rachel Ward
Department of Mathematics
University of Texas at Austin
rward@math.utexas.edu

Hayden Schaeffer
Carnegie Mellon University
schaeffer@cmu.edu

Nicholas Richardson
University of Waterloo, Canada
njerichardson@uwaterloo.ca

MS4

Analysis of Signal Tracking in Partially Observed Systems

With large scale availability of precise real time data, their incorporation into physically based predictive models, became increasingly important. This procedure of combining the prediction and observation is called data assimilation. One especially popular algorithm of the class of Bayesian sequential data assimilation methods is the ensemble Kalman filter which successfully extends the ideas of the Kalman filter to the non-linear situation. However, in case of spatio-temporal models one regularly relies on some version of localization, to avoid spurious oscillations. In this work we develop a-priori error estimates for the time continuous variant of the ensemble Kalman filter, known as localized ensemble KalmanBucy filter. More specifically we aim for the scenario of sparse observations applied to models from fluid dynamics.

Gottfried Hasterman
University of Potsdam
gottfried.hastermann@uni-potsdam.de

MS4

Progression and Pathophysiology: Quantifying the Clinical Evolution of the Human Lung under Mechanical Ventilation

Mechanical ventilation to support patient respiration is an important life-support technology in clinical and hospital environments. It may also have deleterious effects on patient lungs, such as ventilator-induced lung injury (VILI). Quantifying the progressive effects of ventilation on lungs requires representing the lung-ventilator system (LVS) in a way suitable for such investigation. However, the physical system comprises interacting human and mechanically-controlled components that evade a purely physiological description. Likewise, clinical LVS data reflects nonstationary heterogeneity of unique patient lung dynamics under therapeutic applications. Together, these considerations confound traditional data assimilation as well as machine learning approaches. We present an informed-model data assimilation framework robust enough to provide individualization across a wide variety of LVS behaviors, and this method may be generalizable to applications where periodic data-generating processes are not easily modelable. In the LVS context considered, we analyze individualized parameter timeseries for a cohort of patients with acute respiratory distress syndrome (ARDS) to identify temporal trajectory patterns that may help inform injury progression and lessen the risk of VILI.

J N. Stroh
University of Colorado, Anschutz Medical Campus
jn.stroh@cuanschutz.edu

Yanran Wang
University of Colorado
yanran.wang@cuanschutz.edu

Bradford Smith, Peter Sottile
University of Colorado Anschutz Medical Campus
bradford.smith@cuanschutz.edu,
peter.sottile@cuanschutz.edu

George Hripcsak
Columbia University
gh13@cumc.columbia.edu

David Albers
University of Colorado Anschutz Medical Campus
david.albers@cuanschutz.edu

MS4

Reconstructing the Dynamics of the Outer Electron Radiation Belt by Means of the Standard and Ensemble Kalman Filter with the Verb-3D Code

Reconstruction and prediction of the state of the near-Earth space environment is important for anomaly analysis, development of empirical models, and understanding of physical processes. Accurate reanalysis or predictions that account for uncertainties in the associated model and the observations, can be obtained by means of data assimilation. The ensemble Kalman filter (EnKF) is one of the most promising filtering tools for nonlinear and high dimensional systems for terrestrial weather prediction. In this study, we adapt traditional ensemble-based filtering methods to perform data assimilation in the radiation belts. By performing a fraternal twin experiment, we assess the convergence of the EnKF to the standard Kalman filter (KF). Furthermore, with the split-operator technique, we develop two new 3-dimensional EnKF approaches for electron phase space density that account for radial and local processes, and allow for reconstruction of the full 3D radiation belt space. The capabilities and properties of the proposed filter approximations are verified using Van Allen Probe and GOES data. Additionally, we validate the two 3D split-operator Ensemble Kalman filters against the 3D split-operator KF. The use of the split-operator technique allows us to include more physical processes in our simulations and is a computationally efficient data assimilation tool that delivers an accurate approximation of the optimal

KF solution, and is suitable for real-time forecasting.

Angelica Tibocha
German Research Centre for Geosciencs
angelica@gfz-potsdam.de

MS4

Aspects of Data Assimilation for Emerging Applications

Nowadays it is hard to find an area in science or industry where data assimilation is not used in one form or other, although it might appear under a different name. That does not mean that all these fields use the same methods, and that a universal best data-assimilation method exists. Instead, different applications ask for different approaches, and in this talk I will discuss which method to choose for which application area, and why.

Peter Jan van Leeuwen
Colorado State University and University of Reading
peter.vanleeuwen@colostate.edu

MS5

Partial Optimal Transport with Applications on Positive-Unlabeled Learning

Classical optimal transport problem seeks a transportation map that preserves the total mass between two probability distributions, requiring their masses to be equal. This may be too restrictive in some applications, since the distributions may have arbitrary masses and/or only a fraction of the total mass has to be transported. In this talk, we address the partial Wasserstein and Gromov-Wasserstein problems and propose exact algorithms to solve them. We showcase the new formulation in a positive-unlabeled (PU) learning application and highlight that partial Wasserstein-based metrics prove effective in usual PU learning settings. This is joint work with Laetitia Chapel (Rennes University) and Gilles Gasso (INSA Rouen Normandie).

Mokhtar Z. Alaya
University of Technology of Compiègne
elmokhtar.alaya@utc.fr

MS5

Optimal Transport Modeling of Population Dynamics

The ability to understand and predict collective dynamics of a population of cells towards external perturbations or in developmental processes is a core question in molecular biology. As is often the case in challenging scientific applications, notably single-cell genomics, measuring features for these particles requires destroying them. As a result, the population can only be monitored with periodic snapshots, obtained by sampling a few particles that are sacrificed in exchange for measurements. Given only access to these snapshots, can we reconstruct individual trajectories for all other particles? In this talk, I cover two approaches. First, we model these trajectories as collective realizations of a causal Jordan-Kinderlehrer-Otto (JKO) flow of measures: The JKO scheme posits that the new configuration taken by a population at time t+1 is one that trades off an improvement, in the sense that it decreases an energy, while remaining close to the previous configuration observed at t. In order to learn such an energy using only snapshots, we propose JKOnet, a neural architecture that computes (in end-to-end differentiable fashion) the JKO flow given a parametric energy and initial configuration of points. Leveraging the theory of optimal transport again, one can directly learn a coupling describing the response of cell populations upon perturbations such as drugs. Amongst others, this enables us to capture the heterogeneous responses of tumor cells towards cancer drugs.

Charlotte Bunne
ETH Zurich
bunnec@inf.ethz.ch

MS5

A Blob Method for Inhomogeneous Diffusion with Applications to Multi-Agent Control and Sampling

We develop a deterministic particle method for the weighted porous medium equation (WPME) and prove its convergence on bounded time intervals using ideas from optimal transportation. Our method has natural applications to multi-agent coverage algorithms and sampling probability measures. A specific case of our method corresponds exactly to the mean-field dynamics of training a two-layer neural network for a radial basis function activation function. From this perspective, our convergence result shows that, in the over parametrized regime and as the variance of the radial basis functions goes to zero, the continuum limit is given by WPME. This generalizes previous results, which considered the case of a uniform data distribution, to the more general inhomogeneous setting. As a consequence of our convergence result, we identify conditions on the target function and data distribution for which convexity of the energy landscape emerges in the continuum limit.

Karthik Elamvazhuthi
University of California, Riverside
karthikevaz@math.ucla.edu

MS5

Information Geometry and Optimal Transport

Optimal transport and information geometry provide two distinct frameworks for studying the proximity between probability measures. The former is based on the most efficient movement of mass, whereas the latter has its roots in information theory. In recent years, there has been a large body of research which combines insights from these two areas. In this talk, we will discuss some interactions between optimal transport and information geometry, with a focus on topics of interest for data scientists.

Gabriel Khan
Iowa State University
gkhan@iastate.edu

MS6

The Probabilistic Deep Image Prior for Computational Tomography

Existing deep learning based tomographic image reconstruction methods do often not provide accurate estimates of reconstruction uncertainty, hindering their real-world deployment. To address this limitation, we construct a Bayesian prior for tomographic reconstruction, based on the deep image prior (DIP). For inference, we develop an approach based on the linearized Laplace method, which is scalable to high-dimensional settings, with the resulting framework providing pixel-wise uncertainty estimates and

a marginal likelihood objective for hyperparameter optimization. We demonstrate the method on synthetic and real-measured high-resolution CT data and show that it provides superior calibration of uncertainty estimates relative to previous probabilistic formulations of the DIP. We then investigate adaptive design powered by uncertainty for generating effective scanning strategies. Throughout the talk, practical examples will also be shown illustrating the use of the Torch package we released for this work.

Riccardo Barbano
University College London
riccardo.barbano.19@ucl.ac.uk

Simon Arridge
Department of Computer Science
University College London
S.Arridge@cs.ucl.ac.uk

Bangti Jin
University College London
b.jin@ucl.ac.uk

MS6

Monotonically Convergent Regularisation by Denoising

Regularization by denoising (RED) is a widely-used framework for solving inverse problems by leveraging image denoisers as image priors. Recent work has reported the state-of-the-art performance of RED in a number of imaging applications using pre-trained deep neural nets as denoisers. Despite the recent progress, the stable convergence of RED algorithms remains an open problem. The existing RED theory only guarantees stability for convex data-fidelity terms and nonexpansive denoisers. This work addresses this issue by developing a new monotone RED (MRED) algorithm, whose convergence does not require nonexpansiveness of the deep denoising prior. Simulations on image deblurring and compressive sensing recovery from random matrices show the stability of MRED even when the traditional RED algorithm diverges.

Yuyang Hu, Jiaming Liu, Xiaojian Xu, Ulugbek S. Kamilov
Washington University in St. Louis
h.yuyang@wustl.edu, jiaming.liu@wustl.edu, xiaojianxu@wustl.edu, kamilov@wustl.edu

MS6

Depthwise Separable Convolutions Allow for Fast and Memory-Efficient Spectral Normalisation

Deep learning finds its way into more and more domains of our everyday life. The tremendous recent advances in the field of deep learning have the potential to simplify and support our daily routines in applications ranging from virtual personal assistants and autonomous driving to novel medical imaging and diagnosis technologies. Many of the currently available approaches, however, lack in robustness and stability and thus bare a significant safety risk in applications in which incorrect outcomes can be critical for humans. However, most neural networks - e.g. those that consist of fully-connected, convolutional and pooling layers - are Lipschitz continuous. Especially for very large networks, the Lipschitz constant can become arbitrary high, leading to instabilities and non-robustness. To stabilize and accelerate training as well as to make networks more robust, the Lipschitz constant of a network can be restricted to be less or equal to a chosen threshold via spectral normalization. Though, current methods to compute Lipschitz constants of neural networks lack either memory- or timing efficiency, making them unusable for practical problem settings. In this talk, we discuss an approach to memory- and time-efficiently approximate Lipschitz constants of depthwise separable convolutional layers, allowing to restrict Lipschitz constants for practical relevant applications.

Christina Runkel
University of Cambridge, United Kingdom
cr661@cam.ac.uk

Christian Etmann
Deep Render
christian.etmann@deeprender.ai

Michael Moeller
University of Siegen
michael.moeller@uni-siegen.de

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

MS6

Plug-and-Play Algorithm for Uncertainty Quantification in Computational Imaging

Inverse imaging problems aim to find an estimate of an original unknown image from degraded measurements. A classical Bayesian approach finds a Maximum A Posteriori (MAP) estimate, which is defined as a minimiser of the sum of a data-fidelity term and a regularisation term. For high dimensional problems, proximal optimisation algorithms are methods of choice for minimising an objective, owing to their scalability. Nevertheless, they only provide single point estimates, and do not provide any information on the uncertainty of the solution. Recently, it has been shown that to determine whether a structure appearing on the MAP estimate is true or not, simple hypothesis tests can be formulated as minimisation problems, which are then solved using proximal methods. This approach aims to determine, at a given level of confidence, whether or not the structure exists in the ground truth. If not, then the method provides an alternative solution to the MAP estimate, where the structure is absent. This method necessitates a hand-crafted mathematical definition of the "structure", similar to an inpainting process. In this work, we propose to learn the inpainting process of the structure using neural networks (NNs). We then solve the hypothesis test minimisation problem using a plug-and-play algorithm, where the proximity operator related to the inpainting of the structure of interest has been replaced by a NN.

Michael Tang, Audrey Repetti
Heriot-Watt University
mt68@hw.ac.uk, a.repetti@hw.ac.uk

MS7

Novel Methods for Causal Discovery in the Earth and Climate Sciences

In this talk I'll approach the problem of causal discovery from observational data from a pragmatic machine learning point of view. I will introduce several novel methods for causal discovery from time series, with particular fo-

cus on the fields of Earth and climate sciences. We first discuss about new methods that extend Granger causality (GC) to cope with nonlinear and non-stationary processes, based on kernel methods and variational autoencoders able to learn causal feature representations. However, nonlinear GC approaches are not optimal when applied to dynamic systems with weak to moderate variable coupling. To cope with this, we introduce a robust version of convergent cross-mapping (CCM) method which is a nonlinear state-space method. When time is not involved, in bivariate problems, or in i.i.d cases, several methods are available, based on either structural models or kernel mean embeddings. Finally, I will briefly discuss on graph regularization methods that allows to complete causal graphs combining observations and domain knowledge efficiently. We will illustrate performance of all the methods in several challenging problems involving satellite observations, climate models and long-term carbon and water flux records at planetary scale.

Gustau Camps-Valls
Universitat de Valencia, Spain
Image Processing Lab
gustau.camps@uv.es

Gherardo Varando, Emiliano Diaz, Kai-Hendrik Cohrs, Adrian Perez, Jordi Munoz
Image Processing Lab, Universitat de València, Spain
gherardo.varando@uv.es, emiliano.diaz@uv.es, kai.cohrs@uv.es, adrian.perez@uv.es, jordi.munoz@uv.es

MS7

Gaussian Process-Based Learning of Dynamical Graphs

In science and engineering we often deal with problems that are best represented with dynamical graphs. We observe signals on such graphs and the interest is in finding the time-varying topology of the graphs. We propose a Bayesian method which allows for estimating the unknown topology and which does not assume any specific functional relationship among the signals on the graph. The method exploits Gaussian processes and allows for learning the time varying function mappings sequentially. We provide numerical tests that show the performance of the method in various settings.

Petar Djuric
Stony Brook University
petar.djuric@stonybrook.edu

Chen Cui
Department of Electrical and Computer Engineering
Stony Brook University, USA
chen.cui@stonybrook.edu

Yuhao Liu, Marzieh Ajirak
Stony Brook University, USA
yuhao.liu.1@stonybrook.edu,
marzieh.ajirak@stonybrook.edu

Paolo Banelli
Department of Engineering
University of Perugia, Italy
paolo.banelli@unipg.it

MS7

Graph Discovery and Bayesian Filtering in State-Space Models

Modeling and inference in multivariate time series is central in statistics, signal processing, and machine learning, with applications in social network analysis, biomedical, and finance, to name a few. The linear-Gaussian state-space model is a common way to describe a time series through the evolution of a hidden state, with the advantage of presenting a simple inference. A fundamental question when analyzing multivariate sequences is the search for relationships between their entries (or the modeled hidden states). In such context, graphical modeling combined with parsimony constraints allows to limit the proliferation of parameters and enables a compact data representation which is easier to interpret. We propose a novel expectation-maximization algorithm for estimating the linear matrix operator in the state equation. We introduce a novel perspective by relating this matrix to the adjacency matrix of a directed graph, also interpreted as the causal relationship among state dimensions (in the Granger-causality sense). We propose a new method called GraphEM based on the well sounded expectation-maximization (EM). We propose an advanced convex optimization solver relying on a consensus-based implementation of a proximal splitting strategy for solving the M-step. This approach enables an efficient and versatile processing of various sophisticated priors on the graph structure, such as parsimony constraints, while benefiting from convergence guarantees.

Victor Elvira
University of Edinburgh
velvira@ed.ac.uk

Emilie Chouzenoux
Inria Saclay
CVN, CentraleSupélec, Univ. Paris Saclay
emilie.chouzenoux@centralesupelec.fr

MS7

Inferring Graphs from Data: Sparsity, Algorithms and Applications

In this talk, we consider the problem of learning a sparse graph from a Gaussian graphical model whose precision matrix is a Laplacian matrix. Like in the classical graphical lasso problem, recent works made use of the l1-norm with the goal of promoting sparsity in the Laplacian constrained precision matrix estimation. However, through empirical evidence, we observe that the l1-norm is not effective in imposing a sparse solution to this problem. From a theoretical perspective, we prove that a large regularization parameter will surprisingly lead to a solution representing a complete graph, i.e., every pair of vertices is connected by an edge. To address this issue, we propose a nonconvex penalized maximum likelihood estimation method, and establish the order of estimation error and edge recovery consistency. Numerical experiments involving synthetic and real-world data sets demonstrate the effectiveness of the proposed method.

Daniel Palomar
The Hong Kong University of Science and Technology
Department of Electronic and Computer Engineering
palomar@ust.hk

Jiaxi Ying, José Vinícius de Miranda Cardo
Department of Electrical and Computer Engineering
The Hong Kong University of Science and Technology

jx.ying@connect.ust.hk, jvdmc@connect.ust.hk

MS8

Phase Oscillator-Based Analysis and Control of Fluid Flows

Oscillatory fluid flows are encountered in a range of applications in automotive and aviation industries. The major challenge associated with such flows for both analysis and control are their time varying base states. Phase reduction analysis and Floquet theory can support the characterization and control of fluid flows with periodically varying base states. Phase reduction method reduces the high-dimensional flow physics to single scalar phase dynamics. The phase variable is defined based on the limit cycle oscillations of the system. Phase-based analysis focuses on the influence of timing or the phase at which the perturbations are added to the flow field. Spatial analysis of the phase sensitivity study reveals the critical regions that enable phase modification. We implement the phase reduction method to analyze vortex shedding of airfoil wakes where the phase is defined based on the limit cycle oscillations of their lift coefficients. We investigate the influence of chord-based thickness and angle of attack on phase sensitivity fields obtained by an adjoint-based method for flows over symmetric airfoils. We discuss transient lift enhancement techniques through phase advancement or delay by the modification of vortex formation dynamics over the airfoils. Furthermore, we investigate potential pathways to extend phase reduction techniques to fluid flows with multiple dominant frequencies. This work is supported by the US AFOSR (FA9550-21-1-0178) and the US NSF (2129639).

Vedasri Godavarthi
University of California, Los Angeles
vedasrig@g.ucla.edu

Kunihiko Taira
University of California, Los Angeles
Mechanical and Aerospace Engineering
ktaira@seas.ucla.edu

MS8

Data-Driven Modeling and Control of Flexible Wings

From insect wings to wind turbines, aeroelastic structures experience transient unsteady aerodynamic loads that are coupled to their motion. Effective real-time control of these structures relies on accurate and efficient predictions of unsteady aeroelastic forces. Traditional models, such as Theodorsen's model, typically involve quasi-steady or idealized unsteady aerodynamic forces and do not describe transients. For rigid wings, reduced order unsteady aerodynamic models in control state-space form have recently been extended to include viscous effects at low Reynolds numbers. We further extend this rigid wing modeling procedure to a flexible wing, incorporating wing deformation in addition to the quasi-steady forces, added mass forces, and large unsteady transients due to viscous effects. We develop low order linear models based on data from direct numerical simulations of flow past a flexible wing at low Reynolds number, and demonstrate the effectiveness of these models to track a reference lift maneuver while constraining maximum wing deformation. This system identification approach provides an interpretable, accurate, and low-dimensional representation of an aeroelastic system that can aid in system and controller design.

Michelle Hickner, Urban Fasel
University of Washington
mhickner@uw.edu, ufasel@uw.edu

Aditya Nair
University of Nevada, Reno
adityan@uw.edu

Bingni W. Brunton, Steven Brunton
University of Washington
bbrunton@uw.edu, sbrunton@uw.edu

MS8

Learning Nonlinear Projections for Dynamical Systems Using Constrained Autoencoders

Projection onto curved low-dimensional manifolds is an essential tool for constructing simplified reduced-order models of complex systems. Autoencoder neural networks provide natural parameterizations of manifolds via their decoders, and have found increasing applications in model reduction. However, without additional constraints, the process of encoding and decoding fails to define a projection onto the learned manifold. This poses a problem because the direction of projection, and not just the manifold, is important for accurately modeling dynamics of systems with non-normal sensitivity such as high-shear fluid flows. To remedy these problems, we introduce a rich parametric class of nonlinear projection operators defined by autoencoders obeying the constraint that decoding followed by encoding is always the identity. This is accomplished by using invertible activation functions and bi-orthogonal weight matrices in matching layers of the encoder and decoder. We show that the bi-orthogonality constraint defines a smooth matrix manifold on which we carry out optimization during training. We apply the approach to model reduction of systems arising in fluid dynamics, where the autoencoders demonstrate the ability to learn underlying manifolds together with appropriate directions of projection.

Gregory Macchio, Samuel Otto
Princeton University
gm0796@princeton.edu, sotto@princeton.edu

Clarence Rowley
Princeton University
Department of Mechanical and Aerospace Engineering
cwrowley@princeton.edu

MS8

Differentiable Programming-Based Galerkin Models for Unsteady Turbulence

Turbulent flow control has numerous applications and building reduced-order models (ROMs) of the flow and the associated feedback control laws is extremely challenging. Despite the promise of neural networks, they are typically devoid of physical foundations and often lack interpretability. On the other hand, Galerkin projection-based ROMs (GP-ROMs), derived by projecting the Navier Stokes equations on a truncated Proper Orthogonal Decomposition (POD) basis, in spite of their widespread adoption, suffer from instabilities and inaccuracies for predictions over long time horizons. In this work, we propose a differentiable programming approach that blends the strengths of both these strategies, by embedding neural networks explicitly into the GP-ROM, termed Neural Galerkin projection. We

demonstrate this approach on two unsteady flow problems a) Subsonic flow over a cavity at a moderate Mach number. b) Compressible transonic flow over a buffeting NACA0012 airfoil. We show that the Neural Galerkin approach demonstrates significantly longer and more accurate time horizon predictions, when compared to the classical POD-GP (assisted by calibration) and is effective for feedback control. We observe that the key benefits of this approach include increased flexibility in physics-based learning, low computational costs, and a significant increase in interpretability, when compared to purely data-driven neural networks.

Arvind T. Mohan
Los Alamos National Lab
arvindm@lanl.gov

Suryapratim Chakrabarti
The Ohio State University
chakrabarti.23@buckeyemail.osu.edu

Daniel Livescu
Los Alamos National Laboratory
livescu@lanl.gov

Datta Gaitonde
The Ohio State University
gaitonde@osu.edu

MS9

Exploring the Trading Implications of the BHM Framework Using Reinforcement Learning

Building on the information-based asset pricing (or BHM) framework [D C Brody, L P Hughston & A Macrina (2008) Information-Based Asset Pricing. International Journal of Theoretical and Applied Finance **11** (1), 107-142] we present an interactive agent network with several agents, and sequential auctions, where market participants trade futures based on their individual signals about the unknown value of an asset. For various limit order book (LOB) depths, the evolution of the amount of differential information, distribution of overall gains, as well as the dynamics of price discovery are studied numerically. We also formulate an information-based dynamic programming model to harness reinforcement learning algorithms for investigating the existence of a trading policy that maximizes expected terminal gains by exploiting superior information. We find evidence in favor of a policy which involves preserving the superior information up to the known deadline, which seems to support the existence of microstructural market dynamics that are against market efficiency [N Aydin (2022) Reinforcement-learning-based optimal trading in a simulated futures market with heterogeneous agents. Simulation **98** (4) 321-333, first published 3 December 2021].

Nadi Serhan Aydin
Istinye University
serhan.aydin@istinye.edu.tr

MS9

Information Based Asset Pricing: from Signal Processing to Financial Informatics

The Brody-Hughston-Macrina approach to information-based asset pricing introduces a new way of looking at the mechanisms determining price movements in financial markets [D C Brody, L P Hughston & A Macrina (2022) *Financial Informatics*, Singapore: World Scientific Publishing Co]. The resulting theory of financial informatics is applicable across a wide range of asset classes and is distinguished by its emphasis on the explicit modelling of market information data flows. In the BHM theory, each asset is defined by a collection of cash flows and each such cash flow is associated with a family of one or more so-called information processes that provide partial information about the cash flow. The theory is highly intuitive, yet is applicable to trading, investment and risk management, and at the same time leads to interesting mathematics. In this talk we present an overview of the theorys aims and achievements, and highlight some open problems for the future.

Lane P. Hughston
Goldsmiths University of London
l.hughston@gold.ac.uk

MS9

From Pinned Interacting Particle Systems to Interacting Information Flows

In this talk, based in part on [L A Mengtrk, A Family of Interacting Particle Systems Pinned to Their Ensemble Average, Journal of Physics A 54 (2021)], we present a family of pinned stochastic processes that interact with each other through their mean-field, where each particle converges to the ensemble-average of the whole system at some fixed future point in time. The ensemble-average is a Gaussian process whose variance is inversely proportional to the number of particles in the environment, such that each particle converges to zero almost-surely at the fixed future time as the number of particles increases to infinity. As an example, as the size of the system grows asymptotically, each particle becomes an independent α-Wiener bridge in the limit, as opposed to the more classical Ornstein-Uhlenbeck process commonly seen in the interacting particle literature. In a specific setting, we show that the stochastic differential equations describing this system has a connection to the noisy observation processes appearing in the information-based asset pricing framework. More precisely, we will discuss a multi-dimensional financial informatics setup related to the aforementioned interacting particle system, where the ensemble-average is replaced by a conditional expectation. We provide numerical simulations to demonstrate the behaviour of these processes.

Levent Mengütürk
Department of Mathematics
University College London
ucaheng@ucl.ac.uk

MS9

Information Based Trading in Heterogeneous Markets

We consider a pair of traders in a market where the information available to the second trader is a strict subset of the information available to the first trader. The traders make prices based on the information available concerning a security that pays a random cash flow at a fixed time T in the future. Market information is modelled in line with the scheme of [D C Brody, L P Hughston & A Macrina (2008) Int. J. Theor. Appl. Finance **11** (1), 107-142] and [D C Brody, M H A Davis, R L Friedman & L P Hughston (2009) Proc. Roy. Soc. A **365** (2104) 1103-1122]. The risk-neutral distribution of the cash flow is known to the traders, who make prices with a fixed multiplicative bid-offer spread and report their prices to a game master who

declares that a trade has been made when the bid price of one of the traders crosses the offer price of the other. We prove that the value of the first trader's position is strictly greater than that of the second. The results are analyzed by use of simulation studies and generalized to situations where (a) there is a hierarchy of traders, (b) there are multiple successive trades, and (c) there is inventory aversion. Work carried out in collaboration with George Bouzianis and Lane Hughston, Department of Computing, Goldsmiths University of London.

Leandro Sánchez-Betancourt
King's College London
leandro.sanchez-betancourt@kcl.ac.uk

MS10

Physics-Based System Identification with Physics-Informed Dynamic Mode Decomposition (piDMD)

The dynamic mode decomposition (DMD) is the most widely-used algorithm in data-driven dynamical systems research. However, it is well known that the algorithm is sensitive to noise, prone to overfitting, and produces unphysical models. We recently showed that these issues can be ameliorated by incorporating known physical principles into the DMD framework. Our physics-informed dynamic mode decomposition (piDMD) rephrases the DMD optimization as a Procrustes problem where physical principles are interpreted as matrix constraints on the model. The original piDMD algorithm focussed on the five most fundamental physical principles (shift-invariance, energy conservation, self-adjointness, locality and causality) and solved the corresponding optimization problems analytically. In this talk, we will highlight recent improvements to piDMD including modelling the effects of control, incorporating exotic matrix constraints, modelling nonlinear forcings, and weak adherence to physical laws. We apply our framework to a range of physical systems with a focus on applications to fluid mechanics and show that piDMD is a substantial improvement over similar system identification algorithms.

Peter J. Baddoo
Imperial College London
baddoo@mit.edu

Benjamin Herrmann
University of Chile
benjaminh@uchile.cl

Beverley McKeon
California Institute of Technology
mckeon@caltech.edu

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

Steven Brunton
University of Washington
sbrunton@uw.edu

MS10

Updating Digital Twins using Data from a Fleet of Similar Physical Assets

A digital twin is an evolving virtual model of a system or physical asset, that assimilates data during the systems lifecycle so that it becomes a asset-specific model that underpins intelligent automation and drives key decisions. The digital twin concept was devised under the observation that information based on fleet (class) statistics is often not useful for assessing the health of an individual asset, since the condition of assets varies across a fleet due to variability in manufacturing and operating conditions. While it is true that the average health of the asset class cannot be used to reliability certify the health of a single asset, data from multiple assets can be used in combination with local asset data to improve a digital twin. This talk will focus on building networks of Gaussian process-based digital twins to improve inferences about a single asset by utilizing observations of similar assets within an asset class. We will show how to use conditional dependence between the states of each asset in a class to enable the computationally efficient construction of the Gaussian processes with a cost that scales linearly with the number of assets used for inference. Several numerical examples will be used to demonstrate that our novel algorithm significantly improves the quality of digital twin predictions when only limited data can be obtained from each physical asset.

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

Tom Seidl
Sandia National Lab
dtseidl@sandia.gov

MS10

DRIPS: A Framework of Dimension Reduction and Interpolation in Parameter Space

Reduced-order models are often used to describe the behavior of complex systems, whose simulation with a full model is too expensive, or to extract salient features from the full models output. We introduce a new model-reduction framework that combines dynamic mode decomposition (DMD) with reduced-order bases (ROBs). The offline step of this framework relies on DMD to build a low-rank linear surrogate model, endowed with a ROB, for high-dimensional quantities of interest derived from the data computed via a (nonlinear) high-fidelity model. The online step consists of the construction of a parametric reduced-order model for each target/test point in the parameter space, with the interpolation done on a suitable manifold. Through a series of numerical experiments, we demonstrate that our approach outperforms the projection-based proper orthogonal decomposition and Gaussian processes interpolation (Kriging) in terms of both computational costs and accuracy.

Hannah Lu, Daniel M. Tartakovsky
Stanford University
hannahlu@stanford.edu, tartakovsky@stanford.edu

MS10

Deep Neural Network Modeling of Unknown PDEs in Nodal Space

In this talk, we present a numerical framework for deep neural network (DNN) modeling of unknown time-dependent partial differential equations (PDEs) using their trajectory data. Unlike recent work where the learning takes place in modal/Fourier space, the current method conducts the learning and modeling in physical space and

uses measurement data as nodal values. We present a DNN structure that has a direct correspondence to the evolution operator of the underlying PDE, thus establishing the existence of the DNN model. The DNN model also does not require any geometric information of the data nodes. Consequently, a trained DNN defines a predictive model for the underlying unknown PDE over structureless grids. A set of examples, including linear and nonlinear scalar PDE, system of PDEs, in both one dimension and two dimensions, over structured and unstructured grids, are presented to demonstrate the effectiveness of the proposed DNN modeling. Extensions to other equations such as differential-integral equations, are also discussed.

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

Victor Churchill
The Ohio State University
churchill.77@osu.edu

MS11

A Non-Asymptotic Framework for Approximate Message Passing in Spiked Models

Approximate message passing (AMP) emerges as an effective iterative paradigm for solving high-dimensional statistical problems. However, prior AMP theory — which focused mostly on high-dimensional asymptotics — fell short of predicting the AMP dynamics when the number of iterations surpasses $o(\log n/\log\log n)$ (with n the problem dimension). To address this inadequacy, this paper develops a non-asymptotic framework for understanding AMP in spiked matrix estimation. Built upon new decomposition of AMP updates and controllable residual terms, we lay out an analysis recipe to characterize the finite-sample behavior of AMP in the presence of an independent initialization, which is further generalized to allow for spectral initialization. As two concrete consequences of the proposed analysis recipe: (i) when solving Z_2 synchronization, we predict the behavior of spectrally initialized AMP for up to $O(n/\text{poly}\log n)$ iterations, showing that the algorithm succeeds without the need of a subsequent refinement stage (as conjectured recently by Celentano et al.); (ii) we characterize the non-asymptotic behavior of AMP in sparse PCA (in the spiked Wigner model) for a broad range of signal-to-noise ratio.

Yuting Wei
University of Pennsylvania
ytwei@wharton.upenn.edu

MS12

Multiscale Imaging of Tumors on Spatial-temporal Manifolds via Cell Graphs and Fokker-Planck Dynamics

Dysregulation of otherwise physiological pathways contribute to the development and progression of cancer. Examples include tumor hypoxia from disordered vasculature and dysregulation of the immune system, both which are linked to mechanisms of cellular response. New approaches to capture phenotypic changes across length-scale (e.g., tissue, cellular, molecular, etc.) and time-scale (e.g., natural history of disease, response to treatment, etc.) are critical to understanding tumor pathogenesis and therapeutic resistance. Here, we present new data-driven strategies to model the appearance and behavior of cancer across space and time. At a macroscopic scale, we model tumor time dynamics based on assimilation of positron imaging data with Langevin dynamics and Fokker-Planck equations. We demonstrate that individual tumors express dynamic textural changes in tumor metabolism over time, which are linked to several prognostic endpoints. Meanwhile, at a microscopic scale, we study the tumor microenvironment on digitized tissue samples via deep learning, Langevin dynamics, and graph theory. We demonstrate that individual tumors express intrinsic patterns in their cellular architecture, topology, and probability density. These are associated with antibody-specific cell signaling, cell neighborhood clusters, and tumor infiltration by the immune system. Overall, our results suggest that tumors encode multiscale signatures linked to various biological phenomena, such as tumor metabolism and immune dysregulation. These intrinsic expression patterns may lead to a better understanding of the natural history of tumors and optimal selection of personalized therapies.

Kyle Lafata
Duke University
kyle.lafata@duke.edu

MS12

Macroscopic Dynamics for Nonequilibrium Biochemical Reactions from a Hamiltonian Viewpoint

Most biochemical reactions in living cells are open system interacting with environment through chemostats. At a mesoscopic scale, the number of each species in those biochemical reactions can be modeled by the random time-Most biochemical reactions in living cells are open system interacting with environment through chemostats. At a mesoscopic scale, the number of each species in those biochemical reactions can be modeled by the random time-changed Poisson processes. To characterize the macroscopic behaviors in the large volume limit, the law of large number in path space determines a mean-field limit nonlinear reaction rate equation, while the WKB expansion yields a Hamilton-Jacobi equation (HJE) and the convex conjugate L of the corresponding Hamiltonian H gives the good rate function (action functional) in the large deviation principle. We discovered a conservative-dissipative decomposition for any non-equilibrium reaction rate equation by using the stationary solution to HJE. This stationary solution is also the energy landscape of the chemical reactions. We propose a symmetry criteria for a new class of non-equilibrium chemical reactions including many enzyme reactions and this gives rise a new concept of balance within same reaction vector. With this idea, we (i) formulate a Onsager-type gradient flow structure in terms of the energy landscape psi; (ii) find the transition path between multiple non-equilibrium steady states (rare events in biochemical reactions).

Jian-Guo Liu
Duke University
jliu@math.duke.edu

Yuan Gao
Department of Mathematics
Purdue University, USA
gao662@purdue.edu

MS12

Inferring Manifolds from Noisy Data Using Gaus-

sian Processes

In analyzing complex datasets, it is often of interest to infer lower dimensional structure underlying the higher dimensional observations. As a flexible class of nonlinear structures, it is common to focus on Riemannian manifolds. Most existing manifold learning algorithms replace the original data with lower dimensional coordinates without providing an estimate of the manifold in the observation space or using the manifold to denoise the original data. In this talk, we propose a new methodology for addressing these problems, allowing interpolation of the estimated manifold between fitted data points. The proposed approach is motivated by novel theoretical properties of local covariance matrices constructed from noisy samples on a manifold. Our results enable us to turn a global manifold reconstruction problem into a local regression problem, allowing application of Gaussian processes for probabilistic manifold reconstruction. In addition to theory justifying the algorithm, we provide simulated and real data examples to illustrate the performance. This talk is based on the joint work with David B Dunson.

David Dunson
Duke University
dunson@duke.edu

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

MS13

Analysis of Deep PDEs' Solver with Over-parametrization

Recently, numerical analysis for deep PDEs' solver in the perspective of nonparametric estimation has attracted a lot of attention. In this talk, I will present a new analysis results with deep neural networks allowing overparametrization.

Yuling Jiao
Wuhan University
yulingjiao@whu.edu.edu

MS13

Finite Expression Method for Solving High-Dimensional Partial Differential Equations

Designing efficient and accurate numerical solvers for high-dimensional partial differential equations (PDEs) remains a challenging and important topic in computational science and engineering, mainly due to the "curse of dimensionality" in designing numerical schemes that scale in dimension. This paper introduces a new methodology that seeks an approximate PDE solution in the space of functions with finitely many analytic expressions, hence, this methodology is named the finite expression method (FEX). It is proved in approximation theory that FEX can avoid the curse of dimensionality. As a proof of concept, a deep reinforcement learning method is proposed to implement FEX for various high-dimensional PDEs in different dimensions, achieving high and even machine accuracy with a memory complexity polynomial in dimension and an amenable time complexity. An approximate solution with finite analytic expressions also provides interpretable insights into the ground truth PDE solution, which can further help to advance the understanding of physical systems and design postprocessing techniques for a refined solution.

Senwei Liang
Purdue University
liang339@purdue.edu

MS13

Statistical Learning Theory of Deep Neural Networks for Data on a Low-Dimensional Manifold

Many data in real-world applications are in a high-dimensional space but exhibit low-dimensional structures. In mathematics, these data can be modeled as random samples on a low-dimensional manifold. I will talk about regression, classification or distribution estimation using neural networks. It is based on an efficient approximation theory of deep ReLU networks for functions supported on a low-dimensional manifold. We further establish the sample complexity for regression, classification, or distribution estimation with finite samples of data. 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

MS13

New Algorithms for Quantizing Neural Networks

Neural networks are highly non-linear functions often parametrized by a staggering number of parameters, called weights. They have been a subject of intense research due in large part to their remarkable success in a wide range of application areas. Miniaturizing these networks and implementing them in hardware is a direction of research that is fueled by a practical need, and at the same time connects to interesting mathematical problems. For example, by quantizing, or replacing the weights of a neural network with quantized (e.g., binary) counterparts, massive savings in cost, computation time, memory, and power consumption can be attained. Of course, one wishes to attain these savings while preserving the action of the function on domains of interest. We propose new data-driven and computationally efficient methods for quantizing the weights of already trained neural networks, with rigorous and favorable error guarantees. We also discuss sparsity promoting extensions and provide the results of numerical experiments, on large multi-layer networks, to illustrate the performance of our methods. Time permitting, we will also discuss open problems.

Rayan Saab
UCSD Mathematics
rsaab@ucsd.edu

MS14

Leveraging Dataset Structure for Neural Network Prediction

Scientists and engineers are increasingly applying deep neural networks (DNNs) to modelling and design of complex systems. While the flexibility of DNNs makes them an attractive tool, it also makes their solutions difficult to interpret and their predictive capability difficult to quantify. In contrast, scientific models directly expose the

equations governing a process but their applicability is restricted in the presence of unknown effects or when the data are high-dimensional. The emerging paradigm of physics-guided artificial intelligence asks: How can we combine the flexibility of DNNs with the interpretability of scientific models to learn relationships from data consistent with known scientific theories? In this talk, I will discuss my work on incorporating prior knowledge of problem structure (e.g., physics-based constraints) into neural network design. Specifically, I will demonstrate how prior knowledge of task symmetries can be leveraged for improved learning outcomes in convolutional neural network based classification; and how embedding priors from dynamical systems theory can lead to physically plausible neural network based video prediction.

Christine Allen-Blanchette
Princeton University
ca15@princeton.edu

MS14

GFNN: Accurate Long Time Prediction of General Hamiltonian Dynamics via Learning Exactly-Symplectic Maps from Data

We consider the learning and prediction of nonlinear time series generated by a latent symplectic map. A special case is (not necessarily separable) Hamiltonian systems, whose solution flows give such symplectic maps. We do so by representing the symplectic map via a generating function, which we approximate by a neural network (hence the name GFNN). This way, our approximation of the evolution map is always exactly symplectic. This additional geometric structure, together with the fact that we're directly learning a map instead of a vector field, allows the local prediction error at each step to accumulate in a controlled fashion. In fact, by revamping KAM theory in dynamical systems for generating functions, we will prove that the global prediction error grows at most linearly with long prediction time (under reasonable assumptions), which significantly improves an otherwise exponential growth.

Renyi Chen
Georgia Institute of Technology
renyi@gatech.edu

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

MS14

Discovery of Whitney Forms for Construction of Structure Preserving Data-Driven Modeling

As deep learning is increasingly used to build surrogates and data-driven models of physical systems, there remains a number of challenges to achieve models which guarantee: well-posedness, numerical stability, convergence with increasing model size, and structure-preservation. Partial differential equations are the predominant description of physics in these models, as they can easily be incorporated into loss functions with automatic differentiation. Control volume analysis provides an alternative description with more direct ties to structure-preservation, but introduces a challenging geometric component to automate discovery of physically relevant control volumes. We introduce a framework where Whitney forms may be used to discover control volumes from data, with analysis to guarantee preservation of numerical properties lacking in currently popular data-driven models. Examples from semiconductor physics, battery design, and subsurface fracture demonstrate speedups of three orders of magnitude while preserving the robustness of traditional modeling and simulation.

Nathaniel Trask
Sandia National Laboratories
natrask@sandia.gov

MS16

Adaptive Multiscale Sparse Neural Network with Interpretable Basis Expansion for Multiphase Flow Problems

We propose an adaptive sparse learning algorithm that can be applied to discover physical processes and obtain a sparse representation of the solution given an ample snapshot space. Assume that there is a rich class of precomputed basis functions that can be used to approximate the quantity of interest. We design a neural network architecture for learning the coefficients of solutions in the spaces spanned by these basis functions. The information of the basis functions is incorporated in the loss function, which minimizes the differences between the downscaled reduced-order solutions and reference solutions. The network contains multiple submodules and the solutions at different time steps can be learned simultaneously. We propose strategies to identify important degrees of freedom during learning. To find a sparse solution representation, a soft thresholding operator is applied to enforce the sparsity of the output coefficient vectors of the neural network. To avoid over-simplification and enrich the approximation space, some degrees of freedom can be added back to the system through a greedy algorithm. In both scenarios of removing and adding degrees of freedom, the corresponding network connections are pruned or reactivated guided by the magnitude of the solution coefficients obtained from the network outputs. The adaptive learning process is applied to two-phase multiscale flow applications to show the capability and interpretability of the proposed method.

Yating Wang
Hong Kong University
wytgloria@gmail.com

Wing Tat Leung
University of Texas at Austin
sidnet123@gmail.com

Guang Lin
Purdue University
guanglin@purdue.edu

MS16

Multi-Variance Replica Exchange and Its Application to Inverse PDE

In this work, we propose a multi-variance replica exchange stochastic gradient Langevin dynamics method to tackle the challenge of the multiple local optima in the optimization and the challenge of the multiple modal posterior distribution in the inverse problem. Two chains with different temperatures are designed where the low temperature chain aims for the local convergence, and the target of the high temperature chain is to travel globally and explore the whole loss function entropy landscape. However, it may not be efficient to solve mathematical inversion problems by using the vanilla replica method directly since the method doubles the computational cost. To address this issue, we

propose to make different assumptions on the energy function estimation and this facilities one to use solvers of different fidelities in the likelihood function evaluation. Our proposed method significantly lowers the computational cost in the high temperature chain, meanwhile preserving the accuracy and converging very fast. We give an unbiased estimate of the swapping rate and give an estimation of the discretization error of the scheme. We design and solve four inverse problems which have multiple modes. The proposed method is also employed to train the Bayesian PINN to solve the forward and inverse problems; faster and more accurate convergence has been observed when compared to the stochastic gradient Langevin dynamics (SGLD) method and vanilla replica exchange methods.

Zecheng Zhang
Purdue University
zecheng.zhang.math@gmail.com

Yating Wang
The University of Hong Kong
yatingwang@maths.hku.hk

Guang Lin
Purdue University
guanglin@purdue.edu

MS17

Graph Laplacian in $\ell^2 - \ell^q$ regularization for Image Reconstruction

The use of the Laplacian of a properly constructed graph for denoising images has attracted a lot of attention in the last years. Recently, a way to use this instrument for image deblurring has been proposed. Even though the previously proposed method was able to provide extremely accurate reconstructions, it had several limitation, namely it was only applicable when periodic boundary conditions were employed, the regularization parameter had to be hand-tuned, and only convex regularization terms were allowed. In this talk, we propose two automatic methods that do not need the tuning of any parameter and that can be used for different imaging problems. Moreover, thanks to the projection into properly constructed subspaces of fairly small dimension, the proposed algorithms can be used for solving large scale problems.

Alessandro Buccini
University of Cagliari
alessandro.buccini@unica.it

Marco Donatelli
University of Insubria
marco.donatelli@uninsubria.it

MS17

How Neural Networks Can Improve Image Reconstruction?

In this talk I propose approaches using deep learning combined with variational methods for the solution of inverse problems in imaging. Algorithms based on neural networks, can improve in different ways the effectiveness and efficiency of the iterative regularization methods. I will try to answer to the question: how is it possible to exploit information obtained from a large training data set to compensate eventually missing data such as in subsampling reconstruction, super-resolution, inpainting? How well the network approximates the inverse problem solution?

Davide Evangelista
University of Bologna
davide.evangelista5@unibo.it

Elena Morotti
University of Bolognia
elena.morotti4@unibo.it

Elena Loli Piccolomini
University of Bologna
elena.loli@unibo.it

MS17

Asymptotics and Applications of Discretized Graph Cuts

Graph cuts are a popular tool that has found applications in a multitude of fields, in particular in image segmentation and machine learning. While they are most prominently utilized through their relaxation, i.e. spectral clustering, this method yields a cut without any inherent qualitative guarantee, and there are currently no results known regarding asymptotic distributions for spectral clustering. In this talk we introduce an alternative method that relies on discretization in order to superimpose a multinomial framework onto any distribution on a metric space, leading to powerful limit theorems for the original graph cuts. To ensure computability, we propose an algorithm that is able compute restrictions of graph cuts in almost quadratic empirical runtime while also being able to utilize our limit theorems. This makes our results usable in asymptotic testing as well as through bootstrapping. We demonstrate the applicability of our results by showcasing detection of clusters in cell images.

Housen Li
Institute for Mathematical Stochastics
University of Göttingen
housen.li@mathematik.uni-goettingen.de

Axel Munk
University of Goettingen
munk@math.uni-goettingen.de

Leo Suchan
Institute for Mathematical Stochastics
Georg August University of Göttingen
leo.suchan@uni-goettingen.de

MS18

Bayesian Learning of Reduced-Order Systems

In this talk, two Bayesian methods for the learning of reduced-order systems will be discussed. The first method is the Bayesian reduced-order operator inference for time-dependent problems, a non-intrusive, glass-box approach that inherits the basic physics yet without requiring access to the full-order solvers. Inspired by the polynomial structure of projection-based, reduced-state governing equations, a reduced model is learned using linear Bayesian inference with Gaussian priors, and the reduced-order operators are recovered as posterior Gaussian distributions conditioning on projected state data, providing a quantification of modeling uncertainties and a naturally embedded Tikhonov regularization. The second method employs deep kernel learning for the data-driven reduced-

order modeling from high-dimensional observations given by noise-corrupted images. The deep kernel learning technique embeds a deep neural network into a kernel function with which a manifold Gaussian process is formulated, and such a probabilistic deep learning model is utilized in this work for both dimensionality reduction and the representation of reduced-order dynamics. The effectiveness of deep kernel learning in the denoising and uncertainty quantification of reduced models will be demonstrated by numerical results. The first method is joint work with S. A. McQuarrie and K. E. Willcox (UT Austin), while the second with N. Botteghi and C. Brune (UTwente).

Mengwu Guo
University of Twente
m.guo@utwente.nl

MS18

Data-Driven Determination of Problem Well-Posedness

We discuss, implement and illustrate algorithms for the data-driven determination of whether differential equation problems (ordinary and partial differential equations) are well posed (under- or over-parametrized). This is done by repeated variational solution of the problem, followed by data mining of the ensemble of obtained variational solutions. Connections of the approach with randomized linear algebra algorithms as well as weak solutions are explored.

Yannis Kevrekidis
Department of Chemical and Biomolecular Engineering
Johns Hopkins University
ykevrek1@jhu.edu

Tom Bertalan
Johns Hopkins University
USA
tom@tombertalan.com

George Kevrekidis
Johns Hopkins University
gkevrek1@jhu.edu

Elizaveta Rebrova
Princeton University
elre@princeton.edu

MS18

Unraveling Neural Networks with Structure-Preserving Computing

The development of simplified computational models of complex fundamental phenomena in physics, chemistry, astronomy and biology is an ongoing challenge. The purpose of such simplified models is typically to reduce computational cost at a minimal loss of accuracy. At the same time, more importantly, these models can provide fundamental understanding of underlying phenomena. Recently, the following two concepts have gained importance in computational science: (i) machine learning (in particular neural networks) and (ii) structure-preserving (mimetic or invariant-conserving) computing for mathematical models in physics, chemistry, astronomy, biology and more. While neural networks are very strong as high-dimensional universal function approximators, they require enormous datasets for training and tend to perform poorly outside the range of training data. On the other hand, structure-preserving methods are strong in providing accurate solutions to complex mathematical models from science. The goal of the UNRAVEL project is to better understand neural networks to enable the design of highly efficient, tailor-made neural networks built on top of and interwoven with structure-preserving properties of the underlying science problems that can serve as the simplified models mentioned above. This is unexplored terrain, and will lead to novel types of machine learning that are much more effective and have a much lower need for abundant sets of data.

Wil Schilders
Eindhoven University of Technology
w.h.a.schilders@TUE.nl

Barry Koren
Department of Mathematics and Computer Science
Eindhoven University of Technology
b.koren@tue.nl

Jim Portegies
Eindhoven University of Technology
j.w.portegies@tue.nl

Simon Portegies Zwart
Leiden Observatory, Leiden University
spz@strw.leidenuniv.nl

Federico Toschi
Eindhoven University of Technology
f.toschi@tue.nl

Benjamin Sanderse
Centrum Wiskunde & Informatica (CWI)
Amsterdam, the Netherlands
b.sanderse@cwi.nl

MS18

Reinforcement Learning and Reduced Order Modeling for Digital Twins Adaptive Planning

A digital twin is a virtual model evolving in time which is able to mirror a physical asset of interest throughout its operational lifespan. The advantage that a digital asset brings for predictive maintenance, optimization, and planning, among others, is testified by the spread of digital twins in many engineering applications. We present a framework involving reduced order modeling and reinforcement learning for fast and accurate adaptive planning of an unmanned aerial vehicle (UAV). We build a structural digital twin, and we simulate several delivery missions. Each mission is modeled as a Markov decision process with a parametrized state transition probability connected to the structural health state of the UAV. We use reinforcement learning to compute offline the optimal policy for a prescribed set of parameters. We show how to incorporate new information coming from sensors during the operational regime, update the current estimate of the state transition probability, and adapt online the planning strategy to balance the structural self-preservation and time to arrive at client location.

Marco Tezzele
International School of Advanced Studies (SISSA)
Trieste, Italy
marco.tezzele@austin.utexas.edu

Karen E. Willcox
UT Austin

kwillcox@oden.utexas.edu

MS19

Randomized Algorithms for Tikhonov Regularization in Linear Least Squares

We describe three randomized algorithms to efficiently solve regularized linear least squares systems based on sketching. The algorithms compute preconditioners for $\min \|Ax - b\|_2^2 + \lambda\|x\|_2^2$, where $A \in \mathbb{R}^{m\times n}$ and $\lambda > 0$ is a regularization parameter, such that LSQR converges in $\mathcal{O}(\log(1/\epsilon))$ iterations for ϵ accuracy. We focus on the context where the optimal regularization parameter is unknown, and the system must be solved for a number of parameters λ. Our algorithms are applicable in both the underdetermined $m \ll n$ and the overdetermined $m \gg n$ setting. Our algorithms efficiently update preconditioners for new regularization parameters. We introduce an algorithm specifically for an approximately low-rank setting, in which the matrix A has rapidly decreasing singular values and such the problem is of low statistical dimension. The scheme we propose exploits the low statistical dimension while not requiring the computation of the Gram matrix, resulting in a more stable scheme than existing algorithms in this context.

Maike Meier
University of Oxford
maike.meier@maths.ox.ac.uk

MS19

Fast and Accurate Randomized Algorithms for Linear Systems and Eigenvalue Problems

Linear systems and eigenvalue problems are the core problems in numerical linear algebra and ubiquitous in applications. We develop a new class of algorithms for general linear systems and eigenvalue problems. These algorithms apply fast randomized sketching to accelerate subspace projection methods, such as GMRES and Rayleigh-Ritz. This approach offers great flexibility in designing the basis for the approximation subspace, which can improve scalability in many computational environments. The resulting algorithms outperform the classic methods with minimal loss of accuracy. For model problems, numerical experiments show large advantages over MATLAB's optimized routines, including a 100 speedup over gmres and a 10 speedup over eigs.

Yuji Nakatsukasa
University of Oxford
nakatsukasa@maths.ox.ac.uk

Joel A. Tropp
Caltech
jtropp@cms.caltech.edu

MS19

Randomized Riemannian Preconditioning

Recent literature has advocated the use of randomized methods for accelerating the solution of various matrix problems arising throughout computational science. Two leading approaches for leveraging randomization in such methods are sketch-and-solve and randomized preconditioning. In sketch-and-solve, randomization is used to reduce problem size. This approach is intuitive and simple and leads to many interesting algorithms for solving a wide array of problems. However, methods based on this strategy lack sufficient accuracy for some applications. The second approach, randomized preconditioning, uses randomization to precondition the problem, thus allowing for much higher accuracies. The main challenge in using randomized preconditioning is the need for an underlying iterative method, thus randomized preconditioning so far has been applied almost exclusively to solving regression problems and linear systems. In this talk, we will show how to expand the application of randomized preconditioning to another important set of problems: optimization problems with (generalized) orthogonality constraints. Our approach is based on the frameworks of Riemannian optimization and Riemannian preconditioning. We demonstrate it on the problems of computing the dominant canonical correlations and on the Fisher linear discriminant analysis, evaluate the effect of preconditioning on the computational costs and the asymptotic convergence, and present some numerical results.

Boris Shustin
Tel Aviv University
Israel
borisshy@mail.tau.ac.il

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

MS20

CAS4DL: Christoffel Adaptive Sampling for Deep Learning in Data-Scarce Applications

Many problems in computational science and engineering require the approximation of a high-dimensional function from data. In many such applications, data is costly to generate: for example, it each sample may require a costly PDE solve. Therefore, it is imperative to develop highly sample efficient algorithms. Recently, deep neural networks and deep learning have shown great promise to provide breakthrough performance in challenging function approximation tasks. In this work, we propose an adaptive sampling strategy, CAS4DL (Christoffel Adaptive Sampling for Deep Learning) to increase the sample efficiency of DL. Our novel approach is based on interpreting the second to last layer of a DNN as a dictionary of functions defined by the nodes on that layer. With this viewpoint, we then define an adaptive sampling strategy motivated by adaptive sampling schemes recently proposed for linear approximation schemes, wherein samples are drawn randomly with respect to the Christoffel function of the subspace spanned by this dictionary. We present numerical experiments comparing CAS4DL with standard Monte Carlo (MC) sampling. Our results demonstrate that CAS4DL often yields substantial savings in the number of samples required to achieve a given accuracy, particularly in the case of smooth activation functions. These results, therefore, are a promising step toward fully adapting DL to scientific computing applications.

Juan M. Cardenas, Ben Adcock
Simon Fraser University
jcardena@sfu.ca, ben_adcock@sfu.ca

Nick Dexter
Florida State University, U.S.

nick.dexter@fsu.edu

MS20

Leveraging Machine Learning for Subsurface Modeling with Fractional-Order PDEs

Data from both numerical simulations as well as physical experiments of subsurface flows reveal that contaminant tracer is often anomalous, with breakthrough curves and mean-square displacements exhibiting heavy, non-Gaussian tails characteristic of Levy processes, the statistical properties of which are described by fractional-order models. We leverage recent machine learning approaches for inverse problems to demonstrate and compare the efficacy of fractional-order models and local models in a variety of datasets. Our results are both a novel application of machine learning approaches to model discovery for such applications and shed light on what type of models can be expected to accuractely describe subsurface flows in different materials.

Mamikon Gulian, Christian Glusa, Marta D'Elia
Sandia National Laboratories
mgulian@sandia.gov, cglusa@sandia.gov, marti.delia@gmail.com

MS20

Level Set Learning with Pseudo-Reversible Neural Networks for Nonlinear Dimension Reduction in Function Approximation

Inspired by the Nonlinear Level set Learning (NLL) method that uses the reversible residual network (RevNet), we propose a new method of Dimension Reduction via Learning Level Sets (DRiLLS) for function approximation. Our method contains two major components: one is the pseudo-reversible neural network (PRNN) module that effectively transforms high-dimensional input variables to low-dimensional active variables, and the other is the synthesized regression module for approximating function values based on the transformed data in the low-dimensional space. The PRNN not only relaxes the invertibility constraint of the nonlinear transformation present in the NLL method due to the use of RevNet, but also adaptively weights the influence of each sample and controls the sensitivity of the function to the learned active variables. The synthesized regression uses Euclidean distance in the input space to select neighboring samples, whose projections on the space of active variables are used to perform local least-squares polynomial fitting. This helps to resolve numerical oscillation issues present in traditional local and global regressions. Extensive experimental results demonstrate that our DRiLLS method outperforms both the NLL and Active Subspace methods, especially when the target function possesses critical points in the interior of its input domain.

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

Yuankai Teng
University of South Carolina
yteng@email.sc.edu

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

Anthony Gruber
Sandia National Laboratories
adgrube@sandia.gov

Guannan Zhang
Oak Ridge National Laboratory
zhangg@ornl.gov

MS20

Stability and Generalization of Graph Neural Networks

Graph neural networks have seen a steep rise in popularity since their introduction as generalizations of convolutional neural networks to graph structured data, leading to many practical applications with commercial impact. In machine learning settings where the dataset consists of many different graphs, the trained neural network should generalize to graphs outside the training set. In this work, we study the generalization capabilities of graph neural networks. To model the data, we assume that graphs of different classes are sampled from different random graph models. Based on this data distribution, we derive non-asymptotic bounds on the generalization gap between the empirical and statistical loss, that decrease to zero as the graphs become larger. Our generalization bounds depend on the regularity of the network and of the random graph models, but not directly on the number of parameters in the network and not on training. These results can be treated as guarantees that the network will always have some generalization capability, no matter how its weights are specifically chosen. Hence, this innate property of graph neural networks explains one aspect of their success in learning tasks like graph classification and regression.

Ron Levie
Ludwig-Maximilians-Universität München
Mathematisches Institut
levieron@technion.ac.il

Gitta Kutyniok
Ludwig-Maximilians-Universität München
kutyniok@math.lmu.de

Sohir Maskey
Ludwig-Maximilians-Universität München
Mathematisches Insti
maskey@math.lmu.de

MS22

Low Dimensional Embeddings of Phylogenetic β-Diversity Metrics

Modern phylogenetic β-diversity metrics such as UniFrac have given remarkably successful scientific insight into the high dimensional datasets associated with microbial environments. We argue that the effectiveness of UniFrac in high dimensions may be explained by a related metric, Euclidean UniFrac (EUniFrac) that admits a Mahalanobis representation in terms of a strictly ultrametric phylogenetic covariance matrix. Applying a multiscale approximation with a discrete Haar-like basis reveals a hierarchical structure of phylogenetic covariance matrices that may explain their success in generating low-dimensional representations of microbial environment data. This approximation leads to evolutionary insight justifying relationships

between branching events and organism fitness that previous authors have observed. This research has been partially funded by the NSF grant No. 1836914.

Evan Gorman
Department of Applied Mathematics
University of Colorado Boulder
evan.gorman@colorado.edu

MS22

A New Approach to the Numerical Encoding of Bags of Words

A subset of points in a metric space is called resolving when any other point is uniquely determined by its distance to the points in the set. The metric dimension of the metric space is the size of its smallest resolving set. Given a non-empty set X, let 2^X denote its power-set, i.e., the set of all possible subsets of X. The Jaccard distance between $A, B \in 2^X$ is defined as $d(A, B) := 1 - |A \cap B|/|A \cup B|$. d is a metric on 2^X. In this talk, I will present some bounds on the metric dimension of $(2^X, d)$ and show how this concept can be applied to represent tweets as comparatively low-dimensional numerical vectors. This research has been partially funded by the NSF grant No. 1836914.

Alexander J. Paradise
Department of Applied Mathematics
University of Colorado
alexander.paradise@colorado.edu

MS22

Numerical Representation of Words Through Levenshtein Graphs

Motivated by genetic data, we apply the notion of metric dimension to find a low-dimensional numerical representation of words. In this talk, we will consider the metric space over words of length k, with $0 \leq k \leq K$, endowed with the Levenshtein or edit distance, $\ell(u, v)$. This metric space has a graphical representation, which we call a Levenshtein graph $\mathbb{L}_{K;a}$. The nodes of Levenshtein graphs represent words and an edge is placed between two words that are separated by a single substitution, insertion, or deletion. Simple properties of Levenshtein graphs are given alongside their automorphism group. We construct a resolving set of order $O(aK^2)$ composed of two-run strings and an algorithm that computes the edit distance between a string of length k and any single-run or two-run string in $O(k)$ operations. This research has been partially funded by the NSF grant No. 1836914.

Perrin Ruth
Department of Mathematics
University of Maryland, College Park
pruth@umd.edu

MS22

Metric Dimension as a Tool for Symbolic Data Science

A subset of points $R \subseteq M$ of a metric space (M, d) is called resolving when, for each pair of unique points $x, y \in M$ there is at least one $r \in R$ such that $d(r, x) \neq d(r, y)$. The size of the smallest possible resolving sets is called the metric dimension of the space. In this talk, we give an overview of metric dimension and its properties and discuss how resolving sets can be used as a tool for symbolic data science. In general, the computational complexity of determining the exact metric dimension of an arbitrary metric space is prohibitive. However, by focusing on specific types of spaces and by being content with small, but perhaps not minimal, resolving sets, this idea can serve as the foundation of effective embeddings for symbolic data. As a small proof-of-concept, we show how a study of the metric dimension of Hamming graphs can be used to generate representations of biological sequence data. This research has been partially funded by the NSF grant No. 1836914.

Richard C. Tillquist
Computer Science Department
California State University
rctillquist@csuchico.edu

MS23

Incremental Tensor Train Decomposition for Compression and Approximation of High-Dimensional Streaming Datasets

Bayesian inference and other UQ analysis are extremely challenging in high-dimensional spaces. One common approach to mitigate these challenges is to perform dimension reduction in a way that enables inverse problems to be solved in the reduced latent space. In this talk we propose a tensor network-based approach to discover latent structure in large-scale tensor-structured data, e.g., solutions of PDEs. Such an approach would then enable inverse problems to be performed on the discovered latent space. Moreover, we describe a new incremental procedure to sequentially compress large-scale solutions as they arise through solvers or data gathering. Our approaches are developed for the tensor-train decomposition and we discuss and demonstrated advantages compared to the standard TT-SVD algorithm for compression. Time-permitting we compare the results of inference in these latent spaces with latent spaces generated from other models, such as the variational autoencoder. The computational advantages of our approach are showcased on several data-sets of large-scale physical phenomena.

Doruk Aksoy
Department of Aerospace Engineering
University of MIchigan
doruk@umich.edu

Alex Gorodetsky
University of Michigan
goroda@umich.edu

MS23

Advances in Low Rank Tensor Train Taylor Series Approximation for High-Dimensional Bayesian Inversion

Recent advances in randomized methods for constructing tensor trains have made high order Taylor series surrogate models computationally feasible for large-scale Bayesian inverse problems. In this talk we present our latest work in higher order tensor train Taylor series, in which we modify the randomized tensor train construction procedure to exploit the special structure of derivative tensors. This allows us to construct more accurate Taylor series surrogate models at a reduced cost. We use the method to accelerate sampling from the posterior in a distributed parameter PDE constrained inverse problem.

Nicholas Alger

The University of Texas at Austin
Center for Computational Geosciences and Optimization
nalger225@gmail.com

Peng Chen, Omar Ghattas
The University of Texas at Austin
peng@oden.utexas.edu, omar@oden.utexas.edu

MS23

Efficient Randomization Techniques for Solving Bayesian Inverse Problems

Large-scale inverse problems involve fusing incomplete and noisy information from multiple sources, such as model simulations, measurements from sensors, and physical experiments, to obtain a consistent description of the state of the underlying physical system. Solving the Bayesian formulation of these problems enables quantifying the uncertainties associated with the solution. However, solving Bayesian problems presents a major challenge: Solving Bayesian inverse problems is computationally demanding, often requiring hundreds to thousands of expensive simulations to accurately estimate the parameters and their uncertainties. Randomized algorithms provide an attractive means to reduce the computational cost. In this work, we will explore efficient randomization techniques as a means to develop scalable solvers and pre-conditioners to mitigate the computational costs associated with solving Bayesian inverse problems.

Vishwas Rao
Argonne National Laboratory
vhebbur@anl.gov

MS23

Large Data Limit for Bayesian PDE Inversion

Bayesian Inversion for Elliptic and Parabolic PDEs with regard to uncertain coefficients and/or uncertain domains. For log-gaussian prior, we exhibit results on robustness of several approximation algorithms in the large data limit.

Christoph Schwab
ETH-Zurich
schwab@sam.math.ethz.ch

MS24

Computationally Efficient Algorithms for Bayesian Nearest Neighbor Co-Kriging Gaussian Processes with Application in Inter-Satellite Model Calibration

The recently proposed nearest neighbor co-kriging Gaussian process (NNCGP) models make possible inter-satellite calibration by combining measurements of different fidelity and accounting for spatially varying bias correction. The inference of the NNCGP is based on a sequential Markov chain Monte Carlo (MCMC) sampler which involves updating a latent random effect vector. Because of the high-dimensional nature of the latent random effect vector, this sampler may be sensitive to the initial values and has high auto-correlations with the tendency to converge slowly. Here we propose two alternative inferential procedures which target to reduce high-dimensional parametric space, improve convergence, and reduce computing time. The first alternative procedure reduces the posterior sampling space by integrating out the latent processes. The second alternative procedure is a new MCMC free procedure which significantly decreases the computing time without sacrificing prediction accuracy. The good computational and prediction performance of our algorithms are demonstrated on benchmark examples and the analysis of the High-resolution Infrared Radiation Sounder data gathered from two NOAA polar-orbiting satellites and manage to reduce the computational time from two days to just a few minutes.

Alex Konomi
University of Cincinnati
alex.konomi@uc.edu

MS24

Oco-2 Forward Model Emulator Using Kernel Flows

Rigorous uncertainty quantification and methods like MCMC require a substantial amount of forward model evaluations. In NASAs Orbiting Carbon Observatory 2 satellites carbon dioxide retrieval problem, the computationally costly forward model is often the bottleneck for these kinds of experiments. We propose a Gaussian Process emulator as a surrogate model to facilitate this computational challenge and to make large-scale computational experiments more feasible. We employ Kernel Flows method in training the GP, and showcase its forward model emulation capabilities together with an application to CO2 retrievals.

Otto Lamminpää
Jet Propulsion Laboratory
otto.m.lamminpaa@jpl.nasa.gov

MS24

A Two-Stage Adaptive Monte Carlo Algorithm for Atmospheric State Retrievals Using Surrogate Models

We propose a new sampling algorithm combining two quite powerful ideas in the Markov chain Monte Carlo literature - the adaptive Metropolis sampler and the two-stage Metropolis-Hastings sampler. The proposed sampling method is particularly very useful for high-dimensional posterior sampling in Bayesian model calibration which involves a computationally expensive forward model. In the first stage of the algorithm, an adaptive proposal is used based on the previously sampled states, and the corresponding acceptance probability is computed based on an approximated posterior involving an inexpensive surrogate model. The expensive target posterior using the true forward model is evaluated in the second stage only if the proposal is accepted in the inexpensive first stage. While the adaptive nature of the algorithm guarantees faster convergence of the chain and very good mixing properties, the two-stage approach helps in rejecting the bad proposals in the inexpensive first stage, making the algorithm computationally efficient. As the proposals are dependent on the previous states the chain loses its Markov property, but we prove that it retains the desired ergodicity property. The proposed method makes the Bayesian retrieval of atmospheric CO2 states computationally efficient by using a cheap statistical emulator in the first stage and the expensive forward simulator only in the second stage.

Anirban Mondal
Case Western Reserve University
Department of Mathematics Applied Mathematics and Statistics

anirban.mondal@case.edu

MS25

Data-Driven Classification of Stratified Wakes

Bluff body wakes in stratified fluids are known to exhibit a rich variety of dynamic behavior that can be categorized into different dynamic regimes based on Reynolds number (Re) and Froude number (Fr). In this work, we attempt to identify the dynamic regime from limited measurement data in a stratified wake with (nominally) unknown Re and Fr using data-driven techniques. The purely data-driven techniques used here include a library-based sparse regression formulation and convolutional neural networks (CNNs). Both methods require prior high-fidelity data at various points in the regime space. For sparse regression, a large database of candidate basis functions is compiled by pooling the DMD modes obtained in prior Direct Numerical Simulations (DNS). A sparse model is built using the Forward Regression with Orthogonal Least Squares algorithm. The dynamic regime for the measurements is estimated via a projection-weighted average of Re and Fr corresponding to the identified modes. CNNs are trained on a labeled dataset of velocity field snapshots available from DNS. Both these methods are tested for classification accuracy for measurements made in laboratory experiments. These measurements do not always have the same parameter values or field of view as the numerical dataset. We compare these purely data-driven techniques with an expert-defined decision tree-based classification system.

Vamsi Krishna Chinta, Chan-Ye Ohh
University of Southern California
vchinta@usc.edu, ohh@usc.edu

Geoffrey R. Spedding
Aerospace & Mechanical Engineering
University of Southern California
geoff@usc.edu

Mitul Luhar
University of Southern California
luhar@usc.edu

MS25

Finding Scale-Invariant Turbulent Flow Structures for Enhanced Machine Learning

Developing reliable machine-learning (ML) models for turbulence is challenging. There is recent progress in creating complex ML models that exhibit remarkable capabilities for turbulence analyses. However, the reasons for making models successful have often been overlooked. To depart from the trial-and-error developments and black-box uses of ML, this study identifies universal vortical structures over a range of spatial scales that are important for learning turbulence. Our objective is to determine a coordinate that captures the universality of these vortical structures with scale invariance. A sparse Buckingham Pi-based regression is developed for finding the optimal coordinate based on the invariants of velocity gradient tensor. The present method introduces appropriate coordinate transformations, thereby capturing rotational and shear similarities of vortical structures. The identified coordinates also classify seen/unseen structures in training, enabling robust ML construction for turbulence. For demonstration, super-resolution, which reconstructs a high-resolution flow field from its low-resolution counterpart, of isotropic turbulence is considered. We find that ML models achieve a qualitative reconstruction for the identified seen structures, while the performance is decreased towards the region on the coordinate where unseen structures appear. The present findings suggest the importance of embedding scale invariance in ML design for turbulence analyses.

Kai Fukami
University of California, Los Angeles
kfukami1@g.ucla.edu

Kunihiko Taira
University of California, Los Angeles
Mechanical and Aerospace Engineering
ktaira@seas.ucla.edu

MS25

Topological Data Analysis of Pitching and Heaving Wakes

Experimental and computational studies of vortex-dominated flows have shown that changes in the topology of vortices, that is, their structure and arrangement, correlate with the forces exerted on objects immersed in the flow. Quantifying topology using the framework of persistent homology allows for a more-explicit analysis of correlations between the topology and force measurements, both in simulated and experimental flows. Additionally, persistent homology inherently addresses the multiscale nature of determining what is meant by "topology' of any particular fluid flow. We demonstrate multiple approaches to computing the persistent homology of Lagrangian locations of vortices from discrete vortex models and of Eulerian vorticity fields measured in vortex-shedding experiments. In those situations when both types of data are available, we show that either approach results in equivalent topological quantifiers. Finally, we show that time-variation of persistent homology correlates with time variation of measured drag forces. These results suggest that characterizing vortex topology via persistent homology is dynamically meaningful for the study of vortex-dominated fluid flows, therefore opening avenues for the use of persistent homology as an input for optimization, machine learning, and control of fluid flows.

Melissa Green
University of Minnesota
magreen@umn.edu

Marko Budišic
Clarkson University
marko@clarkson.edu

MS27

Robust Modeling of Unknown Systems via Ensemble Averaged Learning

Recent work has focused on data-driven learning of the evolution of unknown systems via deep neural networks (DNNs), with the goal of conducting long time prediction of the evolution of the unknown system. Training a DNN with low generalization error is a particularly important task in this case as error is accumulated over time. Because of the inherent randomness in DNN training, chiefly in stochastic optimization, there is uncertainty in the resulting prediction, and therefore in the generalization error. Hence, the generalization error can be viewed as a random variable with some probability distribution. Well-trained DNNs, particularly those with many hyperparameters, typically result in probability distributions for gen-

eralization error with low bias but high variance. High variance causes variability and unpredictably in the results of a trained DNN. In this talk we present a computational technique which decreases the variance of the generalization error, thereby improving the reliability of the DNN model to generalize consistently. In the proposed ensemble averaging method, multiple models are independently trained, and the average of all the single-step model predictions is used as the initial condition for all models at the next time step. A mathematical foundation for the method is presented, including results regarding the distribution of the local truncation error.

Victor Churchill
The Ohio State University
churchill.77@osu.edu

MS27

Bandit Learning and Budget Allocation for Multi-Fidelity Scientific Computing

Multi-fidelity algorithms operate by recognizing that not all data is created equal: Some data is more trustworthy than other data, and this inequality in data trustworthiness can be exploited if inexpensive data that reveals bulk behavior is appropriately coupled with expensive data that more faithfully adheres to some underlying behavior. We examine this tradeoff in the context of scientific computing, where multiple computational simulations give rise to data of multiple levels of trustworthiness, or "fidelity". As a crude yet clarifying example, simulations with fewer degrees of freedom in discretizing a differential equation are less trustworthy and less costly than simulations employing more degrees of freedom. The question then is how to allocate a fixed computational budget across simulations so that the generated data ensemble reveals as much predictive information as possible. We show that ideas and techniques from budget-limited multi-armed bandit learning can successfully tackle the stochastic decision-making problem of how to allocate computational resources amongst such simulations. We propose and discuss an adaptive multi-armed bandit learning strategy that, with minimal knowledge about given simulations, successfully generates data from simulations in ways that are provably asymptotically optimal and empirically pre-asymptotically near-optimal.

Akil Narayan, Vahid Keshavarzzadeh, Robert Kirby, Yiming Xu
University of Utah
akil@sci.utah.edu, vkeshava@sci.utah.edu, mkirby@cs.utah.edu, yxu@math.utah.edu

MS27

Computational Graph Completion

We present a generalization of Gaussian Process Regression from the approximation of unknown functions to the completion of computational graphs. This generalization is motivated by three observations: (1) Most problems in Computational Sciences and Engineering (CSE) can be formulated as that of completing (from data) a computational graph (or hypergraph) representing dependencies between functions and variables. (2) Replacing unknown functions by Gaussian Processes (GPs) and conditioning on observed data provides a simple but efficient approach to completing such graphs. (3) Since this completion process can be reduced to an algorithm, as one solves $sqrt2$ on a pocket calculator without thinking about it, one could, with the automation of the proposed approach, solve a complex CSE problem by drawing a diagram. We illustrate the proposed framework with applications to system identification, solving/learning PDEs/dynamical systems/SDEs, and ML.

Houman Owhadi
California Institute of Technology
owhadi@caltech.edu

MS27

Reinforced Inverse Problems

Artificial intelligence (AI) has been a powerful tool in science and engineering. Developing AI-aided tools in computational science is a recently popularized research topic. Along this direction, this talk discusses an example of designing intelligent machine that can learn a strategy from data to better solve inverse problems than human-designed algorithms. Especially, we apply reinforcement learning to achieve this goal for solving inverse scattering problems.

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

Yuehaw Khoo, Hanyang Jiang
University of Chicago
ykhoo@uchicago.edu, scottjhy@gmail.com

MS29

Understanding Nesterovs Acceleration via Proximal Point Method

The proximal point method (PPM) is a fundamental method in optimization that is often used as a building block for designing optimization algorithms. In this work, we use the PPM method to provide conceptually simple derivations along with convergence analyses of different versions of Nesterov's accelerated gradient method (AGM). The key observation is that AGM is a simple approximation of PPM, which results in an elementary derivation of the update equations and stepsizes of AGM. This view also leads to a transparent and conceptually simple analysis of AGM's convergence by using the analysis of PPM. The derivations also naturally extend to the strongly convex case. Ultimately, the results presented in this paper are of both didactic and conceptual value; they unify and explain existing variants of AGM while motivating other accelerated methods for practically relevant settings.

Kwangjun Ahn
Massachusetts Institute of Technology, U.S.
kjahn@mit.edu

MS29

On the Complexity of Primal-Dual Coordinate Algorithms

We prove complexity bounds for the primal-dual algorithm with random extrapolation and coordinate descent (PURE-CD), which has been shown to obtain good practical performance for solving convex- concave min-max problems with bilinear coupling. Our complexity bounds either match or improve the best-known results in the literature for both dense and sparse (strongly)-convex-(strongly)-concave problems.

Ahmet Alacaoglu
University of Wisconsin-Madison

alacaoglu@wisc.edu

Volkan Cevher
EPFL
volkan.cevher@epfl.ch

Stephen Wright
University of Wisconsin-Madison
Dept. of Computer Sciences
swright@cs.wisc.edu

MS29

Algorithms with Deviations for Monotone Inclusions and Convex Optimisation

Forward-backward splitting has proven itself as a powerful tool for solving monotone inclusion problems and convex optimisation problems and is also known as ISTA or the proximal-gradient method. This talk will present extensions to forward-backward splitting that introduce deviations, which allow to control where to evaluate the problem data. In contrast to scalar parameters, deviations have the dimension of the problem, allowing for great flexibility. We show new and existing algorithms as special cases of forward-backward splitting with deviations, in particular methods that incorporate information from previous iterations. Moreover we show how deviations allow to use deep learning to develop algorithms that are adapted to a specific problem class by choosing the deviations as the output of a neural network.

Sebastian Banert
Lund University
sebastian.banert@control.lth.se

Jonas Adler
Google DeepMind
jonasadler@google.com

Pontus Giselsson
Lund University, Sweden
pontusg@control.lth.se

Ozan ktem
KTH Royal Institute of Technology
Sweden
ozan@kth.se

Jevgenija Rudzusika
KTH Royal Institute of Technology
jevaks@kth.se

Hamed Sadeghi
Lund University, Sweden
hamed.sadeghi@control.lth.se

MS29

The Restricted Gaussian Oracle as a Proximity Operator for Sampling

Sampling and optimization are fundamental tasks in data science. While the literature on optimization for data science has developed widely in the past decade, with fine convergence rates for some methods, the literature on sampling remained mainly asymptotic until recently. We study the proximal sampler introduced recently by Lee, Shen, and Tian. This sampling algorithm can be seen as a proximal point algorithm for the purpose of sampling. We will discuss the connection with the standard proximal point algorithm in optimization, and how the proximal sampler can be seen as an optimization algorithm over a space of probability measures. Then, we will review an existing convergence guarantee relying on strong convexity, and show new convergence guarantees under weaker assumptions such as convexity and isoperimetry, which allow for nonconvexity. With these results, we obtain new state-of-the-art sampling guarantees for several classes of target distributions.

Yongxin Chen
Georgia tech
yongchen@gatech.edu

Sinho Chewi
Massachusetts Institute of Technology
schewi@mit.edu

Andre Wibisono
Yale University
andre.wibisono@yale.edu

Adil Salim
University of California, Berkeley,U.S.
salim@berkeley.edu

MS30

Simultaneous Neural Network Approximations for Smooth Functions

We establish in this work approximation results of deep neural networks for smooth functions measured in Sobolev norms, motivated by recent development of numerical solvers for partial differential equations using deep neural networks. Our approximation results are nonasymptotic in the sense that the error bounds are explicitly characterized in terms of both the width and depth of the networks simultaneously with all involved constants explicitly determined. Namely, for $f \in C^s([0,1]^d)$, we show that deep ReLU networks of width $\mathcal{O}(N \log N)$ and of depth $\mathcal{O}(L \log L)$ can achieve a nonasymptotic approximation rate of $\mathcal{O}(N^{-2(s-1)/d} L^{-2(s-1)/d})$ with respect to the $\mathcal{W}^{1,p}([0,1]^d)$ norm for $p \in [1,\infty)$. If either the ReLU function or its square is applied as activation functions to construct deep neural networks of width $\mathcal{O}(N \log N)$ and of depth $\mathcal{O}(L \log L)$ to approximate $f \in C^s([0,1]^d)$, the approximation rate is $\mathcal{O}(N^{-2(s-n)/d} L^{-2(s-n)/d})$ with respect to the $\mathcal{W}^{n,p}([0,1]^d)$ norm for $p \in [1,\infty)$.

Sean Hon
Hong Kong Baptist University
seanyshon@hkbu.edu.hk

Haizhao Yang
Purdue University
haizhaoyang@yahoo.com

MS30

Learning Green's Functions Associated with Elliptic and Parabolic PDEs

Can one learn a differential operator from pairs of solutions and right hand sides? If so, how many pairs are required? These two questions have received significant research attention in differential equation learning. Given input-output pairs from an unknown elliptic or parabolic partial differential equation, we will derive a theoretically rigorous scheme for learning the associated Green's func-

tion. By exploiting the hierarchical low-rank structure of Greens functions and randomized linear algebra, we will have a provable learning rate. We will go on to use rational neural networks and deep learning to discover Green's functions from data and reveal mechanistic insights. The talk is based on joint work with Christopher Earls and Nicolas Boullé.

Nicolas Boulle
Mathematical Institute
University of Oxford
boulle@maths.ox.ac.uk

Christopher Earls, Alex J. Townsend
Cornell University
earls@cornell.edu, townsend@cornell.edu

MS30

Reproducing Activation Function for Deep Learning

The speaker will propose reproducing activation functions (RAFs) motivated by applied and computational harmonic analysis to improve deep learning accuracy for various applications ranging from computer vision to scientific computing. The idea is to employ several basic functions and their learnable linear combination to construct neuron-wise data-driven activation functions for each neuron. Armed with RAFs, neural networks (NNs) can reproduce traditional approximation tools and, therefore, approximate target functions with a smaller number of parameters than traditional NNs. In NN training, RAFs can generate neural tangent kernels with a better condition number than traditional activation functions lessening the spectral bias of deep learning. As demonstrated by extensive numerical tests, the proposed RAFs can facilitate the convergence of deep learning optimization for a solution with higher accuracy than existing deep learning solvers for audio/image/video reconstruction, PDEs, and eigenvalue problems. With RAFs, the errors of audio/video reconstruction, PDEs, and eigenvalue problems are decreased by over 14%, 73%, 99%, respectively, compared with baseline, while the performance of image reconstruction increases by 58%.

Chunmei Wang
Department of Mathematics
University of Florida
chunmei.wang@ufl.edu

MS30

Symmetry-Preserving Machine Learning for Computer Vision, Scientific Computing, and Distribution Learning

Symmetry is ubiquitous in machine learning and scientific computing. Robust incorporation of symmetry prior into the learning process has shown to achieve significant model improvement for various learning tasks, especially in the small data regime. In the first part of the talk, I will explain a principled framework of deformation-robust symmetry-preserving machine learning. The key idea is the spectral regularization of the (group) convolutional filters, which ensures that symmetry is robustly preserved in the model even if the symmetry transformation is contaminated by nuisance data deformation. In the second part of the talk, I will demonstrate how to incorporate additional structural information (such as group symmetry) into generative adversarial networks (GANs) for data-efficient distribution learning. This is accomplished by developing new variational representations for divergences between probability measures with embedded structures. We study, both theoretically and empirically, the effect of structural priors in the two GAN players. The resulting structure-preserving GAN is able to achieve significantly improved sample fidelity and diversityalmost an order of magnitude measured in Frchet Inception Distanceespecially in the limited data regime.

Wei Zhu
University of Massachusetts Amherst
zhu@math.umass.edu

MS31

Learning Partially Observed Stochastic Dynamical Systems

In this talk we discuss probabilistic formulations of system identification, with particular focus on handling sparse, noisy, and indirect data. We introduce the problem from a Bayesian perspective and discuss how it provides a principled mechanism for fusing information and data. First, we describe a filtering-based approximate marginal Markov Chain Monte Carlo scheme to obtain the Bayesian posterior for both linear and nonlinear problems. We then discuss several other variational inference approaches. Finally, we discuss how to incorporate energy and symmetry preservation properties of a dynamical system by embedding both Hamiltonian approximations and symplectic integrators into a structured learning approach. Our methods work with both shallow and deep learning, and our numerical experiments demonstrate that obtaining the marginal posterior of the parameter dynamics and making predictions by extracting optimal estimators (e.g., mean, median, mode) can yield orders of magnitude improvement over existing approaches. We attribute this performance to the fact that the Bayesian approach captures parameter, model, and measurement uncertainties, whereas the other methods typically neglect at least one type of uncertainty.

Nicholas Galioto, Alex Gorodetsky
University of Michigan
ngalioto@umich.edu, goroda@umich.edu

Harsh Sharma, Boris Kramer
University of California San Diego
hasharma@ucsd.edu, bmkramer@ucsd.edu

MS31

Variational Integration of Learned Dynamical Systems

The principle of least action is one of the most fundamental physical principle. It says that among all possible motions connecting two points in a phase space, the system will exhibit those motions which extremize an action functional. Many qualitative features of dynamical systems, such as the presence of conservation laws and energy balance equations, are related to the existence of an action functional, i.e. the presence of variational structure. In this talk I will show how to incorporate variational structure into learning algorithms for dynamical systems. More precisely, a quantity related to a discrete Lagrangian is learned from discrete position data of observed trajectories. Velocities or conjugate momenta do not need to be observed or approximated. The technique compensates discretisation errors when trajectories are computed from the learned system.

This is important when moderate to large step-sizes are used and high accuracy is required. The approach is validated using variational backward error analysis. Moreover, variational BEA is used to identify the Lagrangian of the system. It is remarkable that this can be done from position data only without further prior knowledge and without discretization errors.

Christian Offen, Sina Ober-Blöbaum
University of Paderborn
christian.offen@uni-paderborn.de, sinaob@math.uni-paderborn.de

MS31

Preserving Lagrangian Structure in Data-Driven Reduced-Order Modeling of Large-Scale Mechanical Systems

We present a nonintrusive physics-preserving method to learn reduced-order models (ROMs) of Lagrangian mechanical systems. Existing intrusive projection-based model reduction approaches construct structure-preserving Lagrangian ROMs by projecting the Euler-Lagrange equations of the full-order model (FOM) onto a linear subspace. This Galerkin projection step requires complete knowledge about the Lagrangian operators in the FOM and full access to manipulate the computer code. In contrast, the proposed Lagrangian operator inference approach embeds the mechanics into the operator inference framework to develop a data-driven model reduction method that preserves the underlying Lagrangian structure. The method does not require access to FOM operators or computer code. The numerical results demonstrate Lagrangian operator inference on an Euler-Bernoulli beam model and a large-scale discretization of a soft robot fishtail with 779,232 degrees of freedom. Accurate long-time predictions of the learned Lagrangian ROMs far outside the training time interval illustrate their generalizability.

Harsh Sharma, Boris Kramer
University of California San Diego
hasharma@ucsd.edu, bmkramer@ucsd.edu

MS33

A Multi-Fidelity Monte Carlo Method for Kinetic BGK Model with Uncertainties via the Moment Method

Kinetic equations, usually derived from the N-body Newtons second law, typically have integral operators modeling interactions between particles, thus the model may contain uncertainties. In this talk, we study the BGK equation with random parameters and develop a multi-level Monte Carlo method in the multi-fidelity framework, where each fidelity is chosen by different moment methods. We will give theoretical results on the proposed method, as well as how to allocate the computation resource on each level of the Monte Carlo method. Extensive numerical experiments will be shown to demonstrate the efficiency and accuracy of our method.

Liu Liu
The Chinese University of Hong Kong
lliu@math.cuhk.edu.hk

MS33

Deep Learning Theories for Problems with Low-Dimensional Structures

Deep neural networks have demonstrated a great success on many applications, especially on problems with high-dimensional data sets. In spite of that, most existing theories are cursed by data dimension and cannot explain such a success. To bridge the gap between theories and practice, we exploit the low-dimensional structures of data set and establish theoretical guarantees with a fast rate that is only cursed by the intrinsic dimension of the data set. This presentation addresses our recent work on function approximation and operator learning by deep neural networks. The first part function approximation on low-dimensional manifolds. For Sobolev functions defined on a low-dimensional manifold, we show that neural networks can approximate both the function value and its gradient well. The network size critically depends on the intrinsic dimension of the manifold and only weakly depends on the ambient dimension. In the second part, we consider a general encoder-decoder framework to learn Lipschitz operators between infinite dimensional spaces by feedforward neural networks. Such a framework covers most scenarios in real applications. We develop non-asymptotic upper bounds for the generalization error of the empirical risk minimizer. When the problem have low-dimensional structures, our error bounds have a fast rate depending on the intrinsic dimension. Our results show that neural networks are adaptive to the low-dimensional structures of the problem.

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

Haizhao Yang
Purdue University
haizhaoyang@yahoo.com

Minshuo Chen
Georgia Institute of Technology
minshuochen@gmail.com

Tuo Zhao
Georgia Tech
tourzhao@gatech.edu

Wenjing Liao
Georgia Institute of Technology
wliao60@gatech.edu

MS33

MultiAuto-Deeponet: A Multi-Resolution Autoencoder Deeponet for Nonlinear Dimension Reduction, Uncertainty Quantification and Operator Learning of Forward and Inverse Stochastic Problems

A new data-driven method for operator learning of stochastic differential equations(SDE) is proposed in this talk. The central goal is to solve forward and inverse stochastic problems more effectively using limited data. Deep operator network(DeepONet) has been proposed recently for operator learning. Compared to other neural networks to learn functions, it aims at the problem of learning nonlinear operators. However, it can be challenging by using the original model to learn nonlinear operators for high-dimensional stochastic problems. We propose a new multi-resolution autoencoder DeepONet model referred to as MultiAuto-DeepONet to deal with this difficulty with

the aid of convolutional autoencoder. We conduct several numerical experiments to illustrate the effectiveness of our proposed MultiAuto-DeepONet model with uncertainty quantification.

Jiahao Zhang, Guang Lin, Shiqi Zhang
Purdue University
jiahzhang2@outlook.com, guanglin@purdue.edu, zhan2585@purdue.edu

MS34

From Unlabeled Sampling to Manifold Learning and Back

In unlabeled sensing we want to reconstruct an unknown vector $x \in \mathbb{R}^m$ from shuffled observations of its image under a known linear map $A = [a_1^T, \dots, a_n^T]^T$. Equivalently, we want to reconstruct x given the multiset $\{\!\{a_i^T x : i = 1, \dots, n\}\!\}$. The problem is related to simultaneous localization and mapping (SLAM) and to modern sampling theorems with unknown sample locations. Set-structured data are common in machine learning, with point clouds sampled from low-dimensional embedded manifolds being a prototypical example. In this talk I will explore connections between unlabeled sampling and manifold learning and show that both can be interpreted as instances of a certain abstract unlabeled sampling problem. In particular, I will show how manifold learning can be seen as a variant of unlabeled sampling with additional structure. I will then present an instance of this analogy in the context of tomography with unknown view angles and show how ideas of graph denoising and graph isomorphism arise naturally and how the manifold structure affords us better algorithms than in standard unlabeled sensing. I will conclude with a list of interesting problems induced by this perspective.

Ivan Dokmanic
University of Basel
University of Illinois at Urbana-Champaign
ivan.dokmanic@unibas.ch

MS34

Matrix Reordering for Noisy Disordered Matrices: Optimality and Computationally Efficient Algorithms

Motivated by applications in single-cell biology and metagenomics, we consider matrix reordering based on the noisy disordered matrix model. We first establish the fundamental statistical limit for the matrix reordering problem in a decision-theoretic framework and show that a constrained least square estimator is rate-optimal. Given the computational hardness of the optimal procedure, we analyze a popular polynomial-time algorithm, spectral seriation, and show that it is suboptimal. We then propose a novel polynomial-time adaptive sorting algorithm with guaranteed improvement on the performance. The superiority of the adaptive sorting algorithm over the existing methods is demonstrated in simulation studies and in the analysis of two real single-cell RNA sequencing datasets.

Tony Cai
University of Pennsylvania
tcai@wharton.upenn.edu

Rong Ma
Stanford University
rongm@stanford.edu

MS34

Permuted and Unlinked Monotone Regression in $\mathbb{R}^d$: An Approach Based on Mixture Modeling and Optimal Transport

Suppose we want to learn a map between d-dimensional inputs and d-dimensional noisy outputs, without observing (input, output)-pairs, but only separate unordered lists of inputs and outputs. We show that the notion of cyclical monotonicity of the underlying map is sufficient for identification and estimation in the unordered setting. We study restoration of the correct correspondence of (input,output)-pairs ("permutation recovery") and develop a computationally efficient and easy-to-use algorithm for denoising based on the Kiefer-Wolfowitz nonparametric maximum likelihood estimator and techniques from the theory of optimal transport. We provide explicit upper bounds on the associated mean squared denoising error under Gaussian noise. Numerical studies corroborate our theoretical analysis.

Martin P. Slawski
George Mason University
mslawsk3@gmu.edu

Bodhisattva Sen
Department of Statistics, Columbia University
1255 Amsterdam Ave, Room # 1032 SSW, New York, NY 10027
bodhi@stat.columbia.edu

MS34

Homomorphic Sensing: An Algebraic-Geometric Theory

In abstract terms, unlabeled sensing is the problem of uniquely recovering points in a linear subspace, after they have undergone an unknown permutation composed by an unknown coordinate projection. Homomorphic sensing is the natural generalization, that replaces permutations and projections with arbitrary linear transformations. This talk will present the main aspects and findings of the homomorphic sensing theory, built upon notions of algebraic geometry that will be introduced. It will be shown how known results for unlabeled sensing can be retrieved, as a special case, from the homomorphic sensing framework. Extensions of homomorphic sensing from a single subspace to a subspace arrangement will be made, as well as to the sparse and noisy cases. As a byproduct, this will yield in a unified way results about problems such as real phase retrieval, that have been obtained by diverse methods in the literature.

Manolis Tsakiris
Chinese Academy of Sciences, China
manolis@amss.ac.cn

Liangzu Peng
Johns Hopkins University
lpeng25@jhu.edu

MS35

An Inverse Problem in Mean Field Games from Partial Boundary Measurement

In this work, we consider a novel inverse problem in mean-

field games (MFG). We aim to recover the MFG model parameters that govern the underlying interactions among the population based on a limited set of noisy partial observations of the population dynamics under the limited aperture. Due to its severe ill-posedness, obtaining a good quality reconstruction is very difficult. Nonetheless, it is vital to recover the model parameters stably and efficiently in order to uncover the underlying causes for population dynamics for practical needs. Our work focuses on the simultaneous recovery of running cost and interaction energy in the MFG equations from a finite number of boundary measurements of population profile and boundary movement. To achieve this goal, we formalize the inverse problem as a constrained optimization problem of a least squares residual functional under suitable norms. We then develop a fast and robust operator splitting algorithm to solve the optimization using techniques including harmonic extensions, three-operator splitting scheme, and primal-dual hybrid gradient method. Numerical experiments illustrate the effectiveness and robustness of the algorithm. A future direction will be to develop a faster algorithm for inverse problems in higher dimensions with the help of machine learning techniques and neural network architecture.

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

MS35

Towards Closed-loop Deep Brain Stimulation via Optimal Control

Recent advances in machine learning and optimal control have opened new avenues to tackle learning problems in computational neuroscience. In this work, we consider deep brain stimulation (DBS), a highly efficacious treatment for various neurological disorders including medically refractory Parkinsons disease and epilepsy. While conventional open-loop DBS remains an effective treatment strategy, its success is limited by trial and error parameter selection by clinicians and its inability to adapt and personalize to a patients evolving clinical state. Closed-loop neurostimulation has the potential to mitigate these issues and maximize the therapeutic benefit of stimulation while reducing stimulation side effects. Focusing on the Hodgkin-Huxley neuronal model, we formulate the problem of finding an optimal neurostimulation strategy as a control problem. We state and prove the optimality conditions of the Hamilton-Jacobi-Bellman equations and derive the value function from which an optimal control can be recovered. We lay down the mathematical foundation for designing closed-loop treatment strategies and empirically demonstrate the utility of this approach.

Malvern Madondo
Emory University
malvern.madondo@emory.edu

MS35

Overcoming the Curse of Dimensionality for Solving High-dimensional Hamilton-Jacobi Partial Differential Equations Using Neural Networks

Hamilton-Jacobi (HJ) PDEs and optimal control problems are widely used in many practical problems in control engineering, physics, financial mathematics, and machine learning. For instance, controlling an autonomous system is important in everyday modern life, and it requires a scalable, robust, efficient, and data-driven algorithm for solving optimal control problems. Traditional grid-based numerical methods cannot solve these high dimensional problems, because they are not scalable and may suffer from the curse of dimensionality. To overcome the curse of dimensionality, we developed several neural network architectures for solving different classes of high dimensional HJ PDEs and optimal control problems. In this talk, I will show these architectures which have solid theoretical guarantees. These theoretical guarantees are given by the theory of HJ PDEs and optimal controls, instead of the universal approximation theorem. Moreover, solving certain optimal control problems using the proposed architectures does not require a training process. By leveraging dedicated efficient hardware designed for neural networks, these methods have potential for real-time applications in the future.

Tingwei Meng
University of California, Los Angeles, U.S.
tingwei_meng@brown.edu

MS35

Optimal Neural Network Approximation of Wasserstein Gradient Direction via Convex Optimization

The computation of Wasserstein gradient direction is essential for posterior sampling problems and scientific computing. The approximation of the Wasserstein gradient with finite samples requires solving a variational problem. We study the variational problem in the family of two-layer networks with squared-ReLU activations, towards which we derive a semi-definite programming (SDP) relaxation. This SDP can be viewed as an approximation of the Wasserstein gradient in a broader function family including two-layer networks. By solving the convex SDP, we obtain the optimal approximation of the Wasserstein gradient direction in this class of functions. Numerical experiments including PDE-constrained Bayesian inference and parameter estimation in COVID-19 modeling demonstrate the effectiveness of the proposed method.

Yifei Wang
Stanford University, U.S
wangyf18@stanford.edu

MS36

Constrained Multi-Modal Data Mining Using Coupled Matrix and Tensor Factorizations

The complexity of systems such as the brain and the human metabolome requires collection of data from multiple sources recording the behavior of the system from complementary viewpoints. An effective way of jointly analyzing such multi-modal data is through coupled matrix and tensor factorizations (CMTF). However, different characteristics of datasets require to incorporate various loss functions, constraints and coupling structures when using CMTF models. Existing algorithmic approaches for CMTF can incorporate constraints, linear couplings and different loss functions but none of them has been shown to achieve the flexibility to incorporate all. We introduce a flexible algorithmic framework for coupled matrix and tensor factorizations, which utilizes Alternating Optimization (AO) and the Alternating Direction Method of Multipliers (ADMM). The framework facilitates the use of a variety of constraints, loss functions and couplings with linear transformations. Numerical experiments on simulated and real data demonstrate that the proposed approach is

accurate, flexible and computationally efficient with comparable or better performance than available CMTF algorithms. While we focus on CMTF models using the CANDECOMP/PARAFAC (CP) tensor method to analyze higher-order tensors, we will also discuss the extension of the framework to incorporate the PARAFAC2 model, which has shown promising performance in terms of revealing evolving patterns in time-evolving data analysis.

Carla Schenker, Marie Roald
Simula Metropolitan Center for Digital Engineering
Oslo Metropolitan University
carla@simula.no, mariero@simula.no

Xiulin Wang
Affiliated Zhongshan Hospital of Dalian University
xiulin.wang@foxmail.com

Jeremy Cohen
Universite de Rennes, INRIA, CNRS, IRISA
jeremy.cohen@cnrs.fr

Evrim Acar
Simula Metropolitan Center for Digital Engineering
evrim@simula.no

MS36

Accelerating the Computation of the Unscented Transform with Tensor Decompositions for Probabilistic Machine Learning

One way of lifting the Curse of Dimensionality in probabilistic machine learning is by using tensor decompositions. Making predictions in that setting requires the evaluation of nonlinear functions of the tensor decomposition components. This can be done with the Unscented Transform. A commonly used tensor decomposition is the tensor train. However, computing the Unscented Transform in tensor train format requires computationally expensive rounding. In fact, these rounding operations are the computational bottleneck of the algorithm. We propose a novel method that exploits the structure of the tensor train decomposition and reduces the amount of rounding operations without changing the results compared to the conventional Unscented Transform. The performance of the algorithm is compared to standard Monte Carlo simulations.

Eva Memmel
TU Delft
e.m.memmel@tudelft.nl

Clara Menzen
Delft University of Technology
c.m.menzen@tudelft.nl

Kim Batselier
TU Delft
k.bateslier@tudeflt.nl

MS36

The Subspace Power Method for Symmetric Tensor Decomposition

I will discuss a recent method we proposed for low-rank CP symmetric tensor decomposition. We reduce the problem of tensor decomposition to finding rank-1 tensors in a subspace of symmetric tensors, which we find using the power method. In simulations, this method is roughly one order of magnitude faster than existing decomposition algorithms, with similar accuracy. We also study the relevant non-convex optimization problem, and obtain several guarantees, including robustness results when the tensor is only approximately low-rank. Time-permitting, I will also discuss a modification in the algorithm enables it to obtain a symmetric block-term decomposition, which may be used for generalized PCA.

Joao M. Pereira
IMPA - Instituto de Matemática Pura e Aplicada
Rio de Janeiro, Brazil
jpereira@impa.br

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

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

Timo Klock
Deeptech Consulting
tmklock@googlemail.com

MS36

Alternating Mahalanobis Distance Minimization for Well-Conditioned CP Decomposition

CP decomposition (CPD) is prevalent in chemometrics, signal processing, data mining and many more fields. Alternating least squares (ALS) remains one of the most widely used algorithm for computing the decomposition. Recent works have introduced the notion of eigenvalues and singular values of a tensor and explored applications in various fields. We introduce a new formulation for deriving singular values and vectors of a tensor by considering the critical points of a function different from the one used in the previous work. Computing these critical points in an alternating manner minimizes an objective function which is different from the commonly used least squares loss for CPD. Alternating optimization of this new objective leads to simple updates to the factor matrices with the same asymptotic computational cost as ALS. We show that a subsweep of this algorithm can achieve a superlinear convergence rate for exact CPD with rank smaller than the mode lengths. We then view the algorithm as optimizing a Mahalanobis distance with respect to each factor. This perspective allows us to generalize our approach to interpolate between updates corresponding to the ALS and the new algorithm to manage the tradeoff between stability and fitness of the decomposition. Our experimental results show that for approximating synthetic and real-world tensors, this algorithm and its variants converge to a better conditioned decomposition with comparable fitness as compared to the ALS algorithm.

Navjot Singh
University of Illinois at Urbana Champaign
navjot2@illinois.edu

Edgar Solomonik
University of Illinois at Urbana-Champaign
solomon2@illinois.edu

MS37

Few-Shot Transfer via Low-Rank Quadratic Regu-

larization

Few-shot learning aims to learn a model from a very small number of training data. As such, it critically depends on having a good inductive bias in order to offset data scarcity. In this talk, I consider regularization with a general quadratic form for shaping inductive bias. More specifically, we consider the scenario where the coefficient matrix associated with the quadratic form has low rank. We present a theoretical analysis of the effect of such a regularizer in the few-shot setting linear regression tasks. The analysis leads to a principled approach for learning such quadratic regularizers from (readily abundant) source task data and then transferring to the few-shot target task.

Hossein Mobahi
Google Research
hmobahi@google.com

MS37

Overparameterized Learning Beyond the Lazy Regime

In this talk I will focus on demystifying the generalization and feature learning capability of modern overparameterized learning where the parameters of the learning model (e.g. neural network) exceed the size of the training data. Our result is based on an intriguing spectral bias phenomena for gradient descent, that puts the iterations on a particular trajectory towards solutions that are not only globally optimal but also generalize well. Notably this analysis overcomes a major theoretical bottleneck in the existing literature and goes beyond the "lazy" training regime which requires unrealistic hyperparameter choices (e.g. very small step sizes, large initialization or wide models).

Mahdi Soltanolkotabi
University of Southern California
soltanol@usc.edu

MS38

Adjoint-Based Direct Data Assimilation for Optimizing Frictional Parameters and Predicting Postseismic Deformation

Reproduction and prediction of fault slip behavior are one of the important topics for earthquake forecasting. We have developed an adjoint-based data assimilation method for reproducing and predicting fault slips along the subducting plate by combining physics-based simulation and observation of crustal deformation. The method was applied to the postseismic crustal deformation following the 2003 Tokachi-oki earthquake (e.g., Kano et al., 2015 GJI, 2020 EPS). In the presentation, I will introduce our purpose and summarize our recent works.

Masayuki Kano
Tohoku University
Solid Earth Physics Laboratory, Department of Geophysics, To
masayuki.kano.a3@tohoku.ac.jp

Shin'ichi Miyazaki
Kyoto University
miyazaki.shinichi.2m@kyoto-u.ac.jp

Yoichi Ishikawa
JAMSTEC
ishikaway@jamstec.go.jp

Kazuro Hirahara
RIKEN, Kagawa University
khirahara@krb.biglobe.ne.jp

MS38

Model Geometry Uncertainty Quantification for Improved Intracranial Aneurysm Prediction Using Data Assimilation

The anatomy of the main artierial bloody supply system of the brain, known as the Circle of Willis (CoW), strongly differs between individuals resulting in highly variable intracranial vascularization patterns. Modelling and predicting patient-specific intracranial vessel flow fields can benefit greatly from merging various sources of medical imaging data and computational fluid dynamics models. Our initial work has showcased the potential of data assimilation, specifically a local ensemble transform Kalman filter, in providing detailed information on vascular flow and estimates of the uncertainty in a computationally feasible manner. However effective data assimilation requires accurate characterization of both model and observation uncertainty. In such modelling problems as in hemodynamic blood flow, arguably the most significant source of uncertainty is in specification of the model geometry. The model geometry is often determined from the either manual, semi-automatic or fully automatic segmentation of MRI images. Our focus is on investigating methodologies for machine learning based biomedical image segmentation (primarily U-Net). The CoW has a fairly complex structure with many junctions and ambiguous vessels that make segmentation a difficult task. We investigate both mathematically and experimentally recently proposed approaches to quantifying uncertainty in the outputs of such segmentations (e.g. the Probabilistic U-Net) and suggest improvements.

Sahani Pathiraja
UNSW
sahanidpa@gmail.com

MS38

Ensemble-Based Data Assimilation Techniques for State and Parameter Estimation in Earthquake Forecasting

Our ability to forecast earthquakes is hampered by the very limited information on the fault state and parameters. Ensemble data-assimilation methods provide a means to estimate these by combining physics-based models and observations while considering their uncertainties. Using an advanced simulator for earthquake recurrence with an ensemble Kalman Filter (EnKF), we explore state estimation of a fault governed by rate-and-state friction in 1D and 2D simulations of a laboratory setup. The data assimilation estimates the earthquake evolution, even when having highly uncertain observations. In these experiments, we assume the parameters to be perfectly known, which is typically not the case. To address how well state estimation and state-parameter estimation can account for parameter bias, we use a 0D model for earthquake recurrence with a particle filter with sequential importance resampling and a large number of particles. This highlights the advantage of a particle filter over an EnKF for this nonlinear system. Based on the comparison of the various experiments, we discuss the choice of data-assimilation method and -approach for data assimilation in earthquake simulation

and suggest directions for future research.

Ylona van Dinther
Utrecht University
y.vandinther@uu.nl

Hamed A. Diab-Montero
Delft University of Technology (TU Delft)
Department of Geoscience and Engineering
h.a.diabmontero@tudelft.nl

Arundhuti Banerjee
Delft University of Technology
arundhuti.banerjee@emi.fraunhofer.de

Femke C. Vossepoel
Delft University of Technology, NL
f.c.vossepoel@tudelft.nl

MS39

On the Consistency of Some Hill-Climbing Clustering Methods

Fukunaga and Hostetler proposed, in the 1970s, to cluster points in space according to the gradient flow of the underlying sampling density. Although their work has generated a good amount of interest, including in recent years, the theoretical underpinnings of this proposal have not been fully developed. In this talk, we will present novel consistency results for the main methods for this approach to clustering. Joint work with Wanli Qiao (George Mason University).

Ery Arias-Castro
UCSD
eariasca@ucsd.edu

MS39

Active Learning with Graphs for Remote Sensing Applications

We present a novel method for classification of remote sensing data by combining ideas from graph-based learning and neural network methods within an active learning framework. Graph-based methods in machine learning are based on a similarity graph constructed from the data. When the data consists of raw images composed of scenes, extraneous information can make the classification task more difficult. In recent years, neural network methods have been shown to provide a promising framework for extracting patterns from images. These methods, however, require ample training data to avoid overfitting. At the same time, such training data are often unavailable for applications of interest. We use a neural network to embed the data into a feature space, and then construct a similarity graph from the embedded data and apply graph-based semi-supervised learning techniques. The feature embedding and graph construction requires no labeled data, which reduces overfitting and improves the generalization performance of graph learning at low label rates. Furthermore, the method easily incorporates a human-in-the-loop for active learning in the data-labeling process.

Andrea L. Bertozzi
UCLA Department of Mathematics
bertozzi@math.ucla.edu

MS39

On Multiclass Adversarial Training and Multimarginal Optimal Transport Problems

Adversarial training is a framework widely used by machine learning practitioners to enforce robustness of learning models. Despite the development of several computational strategies for adversarial training and some theoretical development in the broader distributionally robust optimization literature, there are still several theoretical questions about adversarial training that remain relatively unexplored. In this talk, I will discuss an equivalence between adversarial training in the context of non-parametric multiclass classification problems and multimarginal optimal transport problems. This is another analytical interpretation of adversarial training that expands recently studied connections to perimeter minimization problems. One of the implications of the connection discussed during the talk is computational: to solve a certain adversarial problem, we may as well solve a multimarginal optimal transport problem. We will discuss many of the nuances of this interpretation and of its computational consequences.

Nicolas Garcia Trillos
Department of Statistics
University of Wisconsin-Madison
garciatrillo@wisc.edu

MS39

A Similarity Graph-Based Max-Flow Approach To Data Classification and Image Segmentation

The max-flow problem entails the computation of a maximum feasible flow from a source to a sink through a network under constraints. Its connection to total variation presents an opportunity to apply the problem to machine learning tasks by incorporating a similarity graph-based setting. In this paper, we integrate max-flow and duality techniques, similarity graph-based frameworks and semi-supervised procedures to derive algorithms for machine learning tasks, such as classification, and image segmentation. An important advantage of the methods is that they require only a small set of labeled samples for good accuracy, in part due to the integration of graph-based and semi-supervised techniques; this is an important advantage due to the scarcity of labeled data.

Ekaterina Rapinchuk
Michigan State University
Department of Mathematics
merkurje@msu.edu

MS40

Inverse Design via Deep Learning

The main bottleneck of classical methods for inverse design, e.g., the SIMP (Solid Isotropic Material with Penalisation) algorithm, is the necessity to solve multiple, sometimes hundreds of PDEs (partial differential equations) for every single design. This slows down practitioners who use these methods or constrains them to the use of coarse models. Thanks to the advent of DL (deep learning) based amortized optimization and PDE solvers, this seems to change. We want to discuss general developments and our work in

the field.

Sören Dittmer
University of Cambridge
sd870@cam.ac.uk

MS40

Operator Correction

Iterative model-based reconstruction approaches for high-dimensional problems with non-trivial forward operators can be highly time consuming. Thus, it is desirable to employ model reduction techniques to speed-up reconstructions in variational approaches as well as to enable training of learned model-based techniques. Nevertheless, reduced or approximate models can lead to a degradation of reconstruction quality and need to be accounted for. For this purpose, we discuss in this talk the possibility of learning a nonlinear data-driven explicit model correction for inverse problems and whether such a model correction can be used within a variational framework to obtain regularized reconstructions.

Andreas Hauptmann
University of Oulu
University College London
andreas.hauptmann@oulu.fi

MS40

Data-Driven Regularization for Imaging Inverse Problems: Algorithms and Guarantees

Inverse problems arise in numerous applications where imaging is used as a tool to advance our understanding of an object of interest or discover new scientific phenomena. Variational regularization with analytically designed regularizers has by far been the most widely adopted approach for imaging inverse problems. While such hand-crafted analytical regularizers are well-understood mathematically, they are generic and cannot adapt well to a specific application. The recent success of deep learning has motivated the quest for data-adaptive regularizers with tunable parameters that can be learned from available training images. Such data-driven regularizers have been shown to outperform their analytical variants in different applications. This talk will focus on the recent advances in data-driven variational regularization, both from algorithmic and theoretical perspectives. In particular, the adversarial regularization framework will be introduced and different learning protocols and network constructions will be discussed. The talk will highlight the trade-off between mathematical guarantees and empirical performance using the examples of convex adversarial regularizers and a recent optimal transport-based approach that combines regularizer learning with algorithm unrolling.

Subhadip Mukherjee
University of Cambridge
subhadipju@gmail.com

Carola Bibiane Schoenlieb
University of Cambridge, United Kingdom
cbs31@cam.ac.uk

Marcello Carioni
Faculty of Mathematics
University of Cambridge
mc2250@cam.ac.uk

Sören Dittmer
University of Cambridge
sd870@cam.ac.uk

Ozan ktem
KTH Royal Institute of Technology
Sweden
ozan@kth.se

Martin Burger
Friedrich-Alexander University
Erlangen-Nürnberg
martin.burger@fau.de

Zakhar Shumaylov
University of Cambridge
zs334@cam.ac.uk

MS40

Regularisation of Inverse Problems: Deep Equilibrium Models vs Bi-Level Optimisation

Variational regularisation methods are commonly used to approximate solutions of inverse problems. In recent years, model-based variational regularisation methods have often been replaced with data-driven ones such as the fields-of-expert model (Roth and Black 2009). Training the parameters of such data-driven methods can be formulated as a bi-level optimisation problem. In this talk, we compare the framework of bi-level optimisation for the training of data-driven variational regularisation models with the novel framework of deep equilibrium models (Bai, Kolter, Colton 2019) that has recently been introduced in the context of inverse problems (Gilton, Ongie, Willett 2021). We show that computing the lower-level optimisation problem within the bi-level formulation with a fixed-point iteration is a special case of the deep equilibrium framework. We compare both approaches computationally, with a variety of numerical examples for the inverse problems of denoising, inpainting and deconvolution.

Danilo Riccio
Queen Mary University of London
s.d.riccio@qmul.ac.uk

Matthias Ehrhardt
University of Bath
me549@bath.ac.uk

Martin Benning
Queen Mary University of London
m.benning@qmul.ac.uk

MS41

A Novel A.I Enhanced Automatic History Matching with a Black-Oil Physics Constrained Neural Network Forward Model

A key bottleneck in history matching is turnaround time due to the need to run many complex simulations. And while the recent advances in parallel algorithms have allowed reducing turnaround time by adding more hardware, fundamentally, reservoir simulation remains a computationally expensive exercise. In this work, we propose a different approach that promises to reduce computational expense in addition to turnaround time. We use tools from AI to achieve up to 30X faster results compared to traditional numerical methods while maintaining

accuracy. We make two contributions. First, we propose an AI-based (forward) black oil simulator which combines ideas on Physics Informed Neural Networks (PINNs) with a mixture-of-experts variant we call Cluster Classify Regress (CCR). And second, we propose an inversion workflow we call Regularized Ensemble Kalman inversion (REKi) derived using the Discrepancy Principle. The developed workflow was run on a synthetic reservoir model, and it successfully recovered the bimodal unknown channelized permeability and porosity fields using variational convolutional autoencoders as exotic priors.

Clement Etienam, Pavel Dimitrov
NVIDIA
cetienam@nvidia.com, pdimitrov@nvidia.com

Issam Said
Nvidia
isaid@nvidia.com

Ken Hester
NVIDIA
khester@nvidia.com

MS41

Multiscale Dropblock DNN for LSRTM Hessian Computation

Conventional seismic depth migration suffers from low resolution, incorrect amplitude compensation, limited bandwidth, and migration artifacts. Compared to conventional depth migration, LSRTM can generate balanced amplitudes, less migration artifacts, higher spatial resolution, and improved images for complicated structures by fitting the reflection wavefield with a linearized inverse scattering formula iteratively. LSRTM can be implemented as an iteration inversion in data domain or approximate inverse Hessian computation using a single operator like point spread functions (PSFs) or nonstationary matching filter in image domain. Recently deep learning neural network (DNN) has been used to approximate the inverse Hessian computation and produce seismic images with correct amplitudes at a high resolution in highly flexible and efficient way. Here we developed a UNet based deep learning neural network using multiresolution dilated convolution operators to enable the multiscale feature learning for different illumination patterns from different seismic events and also dropblock layer which are new regularization to alleviate the overfitting and generate more stable and generalizable model by adding ensemble learning idea for neural network training. We test the feasibility of our results using both synthetic and field datasets successfully.

Chaoshun Hu
Chevron
chaoshun.hu@chevron.com

Yue Wang
chevron
yue.wang@chevron.com

Xinhui Min
Chevron Corporation, U.S
xinhuimin@chevron.com

Chunyan Xie
Chevron Corporation, U.S.
chunyan.xie@chevron.com

MS41

Physics-Guided Cnn Approach to Solve Classical Geophysical Workflow

Correlating subsurface geology to observed seismic data is essential for most subsurface evaluation workflows. A reliable seismic-to-well-tie algorithm that involves matching depth-domain well data with time-domain seismic data is commonly done for correlation. Although this is a well-known workflow, the traditional algorithm is typically run manually one well at a time. However, tying multiple wells at once is pretty tedious, hardly reproducible, and requires an interpreters experience. We proposed a fully automated workflow to accurately estimate subsurface properties using a physics-guided convolutional neural network. The method takes advantage of the physics nature of the process, to establish an unsupervised neural network. It allows the fully automated workflow to be adjusted and solve for either:

- time depth relationship on seismic to well tie analysis,
- the acoustic impedance on an inversion problem.

We further add Bayesian layers to the network design to encapsulate the uncertainties of the model. This probabilistic approach to deep learning allows one to estimate the uncertainty, which enhance the interpretability of the model.

David Thanoon
Shell
david.thanoon@shell.com

MS42

Adaptive Residual Sub-Sampling Methods for Kernel Interpolation Based on Maximum Likelihood Estimations

In this talk we propose an enhanced version of the residual sub-sampling method (RSM) in [T.A. Driscoll, A.R.H. Heryudono, Adaptive residual subsampling methods for radial basis function interpolation and collocation problems, Comput. Math. Appl. 53 (2007) 927-939] for adaptive interpolation by radial basis functions (RBFs). More precisely, we introduce in the context of sub-sampling methods a maximum profile likelihood estimation (MPLE) criterion for the optimal selection of the RBF shape parameter. This choice is completely automatic, provides highly reliable and accurate results for any RBFs, and, unlike the original RSM, guarantees that the RBF interpolant exists uniquely. The efficacy of this new method, called MPLE-RSM, is tested by numerical experiments on some 1D and 2D benchmark target functions.

Roberto Cavoretto
Department of Mathematics
University of Torino, Italy
roberto.cavoretto@unito.it

Alessandra De Rossi
Department of Mathematics
University of Torino
alessandra.derossi@unito.it

MS42

Physics-Informed Neural Networks Approach for

Solving Gray-Scott Systems

A physics-informed neural network (PINN) is employed to approximate the solution of nonlinear partial differential equation systems [S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, F. Piccialli, Scientific Machine Learning through Physics-Informed Neural Networks: Where we are and What's next, (2022); arXiv preprint arXiv:2201.05624]. In this talk, we present an approach for solving different configurations for the Gray-Scott, a reaction-diffusion system that involves an irreversible chemical reaction between two reactants. Computational results show that the PINN can successfully provide an approximated solution in a variety of conditions, also reproducing the characteristic Turing patterns in the unstable region of the model's parameter space, through a supervised approach that relies on a finite difference method (FDM) [M. Raei, S. Cuomo, G. Colecchia, G. Severino, Solving 3-D GrayScott Systems with Variable Diffusion Coefficients on Surfaces by Closest Point Method with RBF-FD, Mathematics 9 (2021), 924]. Joint work with Francesco Piccialli and M.O.D.A.L Laboratory.

Salvatore Cuomo
University of Napoli Federico II, Italy
salvatore.cuomo@unina.it

MS42

Approximation of Complex Functions via Multinode Shepard Operators

The idea beyond the multinode Shepard operators is the fact that, working with scattered data, it is always possible to find, in a set of p nearby scattered nodes, a subset σ_m of $m \leq p$ nodes which is good for polynomial interpolation of reasonable degree d s.t. $m = \binom{d+s}{s}$, that is the set σ_m is unisolvent for interpolation in a polynomial space $\Pi_d(\mathbb{R}^s)$ of total degree d with a good Lebesgue constant at least in the smallest disc centered at the barycenter of σ_m and containing all of them. These configurations of nodes can be obtained by minimizing the 1-norm condition number of Vandermonde matrices [F. DellAccio, F. Di Tommaso, N. Siar, On the numerical computation of bivariate Lagrange polynomials, Applied Mathematics Letters 112 (2021) 106845], or by using the so called Leja sequences [F. DellAccio, F. Di Tommaso, N. Siar, M. Vianello, Numerical differentiation on scattered data through multivariate polynomial interpolation, BIT Numerical Mathematics (2021) 129]. In the complex case, any set of $d + 1$ pairwise distinct nodes is unisolvent for interpolation by polynomials of complex variable of degree at least d. After the inspiring work [O. Duman, B. Della Vecchia, Complex Shepard Operators and Their Summability, Results in Mathematics 76 (4) (2021) 119], in this talk we introduce and discuss the complex version of the multinode Shepard operator for the interpolation of functions of complex variable.

Francesco Dell'Accio
University of Calabria
Department of Mathematics and Computer Science
francesco.dellaccio@unical.it

Maria Rosaria Capobianco
Istituto per le Applicazioni Del calcolo Mauro Picone, CNR,
mariarosaria.capobianco@cnr.it

Filomena Di Tommaso
University of Calabria
Department of Mathematics and Computer Science
filomena.ditommaso@unical.it

Federico Nudo
Department of Mathematics and Computer Science,
University o
federico.nudo@unical.it

MS42

DISC: An Adaptive Numerical Differentiator by Local Polynomial Interpolation on Multivariate Scattered Data

In this talk, we present a pointwise numerical differentiation formula on multivariate scattered data [F. DellAccio, F. Di Tommaso, N. Siar, M. Vianello, Numerical differentiation on scattered data through multivariate polynomial interpolation, Bit. Numer. Math., https://doi.org/10.1007/s10543-021-00897-6, 2021], based on the coefficients of local polynomial interpolation at Discrete Leja Points [L. Bos, S. De Marchi, A. Sommariva, M. Vianello, Computing multivariate Fekete and Leja points by numerical linear algebra, SIAM J. Numer. Anal., vol. 48, pp. 19841999, 2010], written in Taylors formula monomial basis [F. DellAccio, F. Di Tommaso, N. Siar, On the numerical computation of bivariate Lagrange polynomials, Appl. Math. Lett., p. 106845, 2020]. Error bounds for the approximation of partial derivatives of any order compatible with the function regularity are provided, as well as sensitivity estimates to functional perturbations, in terms of the inverse Vandermonde coefficients that are active in the differentiation process. Moreover, we discuss a MatLab implementation of an adaptive pointwise numerical differentiator on multivariate scattered data. Several numerical tests are presented showing the accuracy of the approximation.

Filomena Di Tommaso, Francesco Dell'Accio
University of Calabria
Department of Mathematics and Computer Science
filomena.ditommaso@unical.it,
francesco.dellaccio@unical.it

Najoua Siar
University Ibn Tofail
University of Calabria
najoua.siar@uit.ac.ma

Marco Vianello
University of Padova (Italy)
marcov@math.unipd.it

MS43

Noise, Fake News, and Tenacious Bayesians

A modelling framework, based on the theory of signal processing, for characterizing the dynamics of systems driven by the unravelling of information is outlined, and is applied to describe the process of decision making. The model input of this approach is the specification of the flow of information. This enables the representation of (i) reliable information, (ii) noise, and (iii) disinformation, in a unified framework, and is particularly suited for synthetic data generation for large-scale simulations and scenario analysis to quantify the impact of information control, including those resulting from the dissemination of disinformation. It will be shown that if a decision maker assigns an exceptionally high weight on one of the alternative realities, then under the Bayesian logic their perception hardly changes in

time even if evidences presented indicate that this alternative corresponds to a false reality. It follows that confirmation bias need not be entirely incompatible with Bayesian updating.

Dorje C. Brody
Department of Mathematics,
University of Surrey
d.brody@surrey.ac.uk

MS43

Filtered Arcade Martingales

Randomized arcade processes constitute a class of stochastic processes which by construction strongly (ω-by-ω) interpolate between a set of arbitrary random variables indexed by fixed, prespecified dates. The structure of these processes introduces a stochastic filtering interpretation whereby the randomized interpolator is viewed as a signal component that is obfuscated by the arcade noise process. Some randomized arcade processes are neither infinite-memory nor memoryless processes, and so we introduce the nearly-Markov property with respect to their natural filtration. Nearly-Markov randomized arcade processes allow for the construction of so-called filtered arcade martingales thus providing filtering solutions to the martingale interpolation problem. These filtered martingales depend on the martingale coupling of the target random variables in between which they interpolate. This feature can be used to bridge filtered arcade martingales to martingale optimal transport problems by considering optimally coupled target random variables. Filtered martingales play a fundamental role in information-based asset pricing because they serve as stochastic filtering models for no-arbitrage price processes. A potential link to stochastic filtering optimal transport is suggested with possible applications in data science, e.g., financial informatics. Work carried out in collaboration with Andrea Macrina, University College London.

Georges Kassis
University College London
georges.kassis.19@ucl.ac.uk

MS43

Strong Stochastic Interpolation with Arcade Processes

A new class of interpolating stochastic processes is introduced called arcade processes. These are continuous processes that interpolate strongly (ω-by-ω) between an arbitrary number of zeros which are fixed at prespecified times. Arcade processes have two components: the deterministic interpolating coefficients and a continuous stochastic driver. They may be seen as multi-period anticipative stochastic bridges and so readily generalize continuous bridge processes, e.g., standard Brownian bridges. We investigate the statistical properties of arcade processes and show that in a Gauss-Markov setting one can always construct a Markovian arcade process by relying on the covariance structure of the driver. The additive structure of arcade processes lends itself well to the randomization of the target values, the zeros in the standard setting, at the end of each arc. An arbitrary number of random variables indexed by pre-specified times is thus introduced, which give rise to the so-called randomized arcade processes. By construction, these stochastic interpolators match the target random variables in the strong sense. Randomized arcade processes provide an important extension to the continuous information processes employed in the theory of information-based asset pricing, since they can be seen as a sum of a signal component and a noise process. Work carried out in collaboration with Georges Kassis, University College London.

Andrea Macrina
University College London
a.macrina@ucl.ac.uk

MS43

Darwin Among the Cryptocurrencies

The paper highlights some commonalities between the development of cryptocurrencies and the evolution of ecosystems. Concepts from evolutionary finance embedded in toy models consistent with stylized facts are employed to understand what survival of the fittest means in cryptofinance. Stylized facts for ownership, trading volume and market capitalization of cryptocurrencies are selectively presented in terms of scaling laws. Work carried out in collaboration with Henry C W Price, Department of Physics and Centre for Complexity Science, Imperial College London.

Bernhard Meister
MK2 Finance, Vienna, Austria
bernhard.k.meister@gmail.com

MS44

Towards Certifiably Safe Nonlinear Control in Uncertain Systems

Modern nonlinear control theory seeks to endow systems with properties such as stability and safety. Despite its successful deployment in various domains, uncertainty remains a significant challenge, while data offers a potential solution. In this talk, I will discuss data-driven nonlinear control, where uncertainty arises from the dynamics model or the sensing model. I will introduce robust control synthesis procedures based on Control Lyapunov and Control Barrier Functions. Using this framework, we will show data-dependent guarantees. These results allow us to quantify tolerable error in learned models, and they can serve as a guide for designing sample-efficient data collection approaches. Based on joint work with Aaron Ames, Ryan Cosner, Victor Dorobantu, Ben Recht, Andrew Taylor, and Yisong Yue.

Sarah Dean
Cornell
sdean@cornell.edu

MS44

Learning Near-Best Polynomial and Neural Network Approximations to High-Dimensional, Banach-valued Functions from Scarce Data

A key problem in many applications in computational science and engineering is the approximation of high-dimensional functions from limited amounts of sample data. In such applications, the data may be scarce or incomplete due to cost of acquiring samples and the samples themselves may even be corrupted by measurement errors. Furthermore, many such problems involve the approximation of a parameter to solution map associated with an underlying PDE, taking values in an infinite-dimensional Hilbert or Banach space. Sparse polynomial approximation has become indispensable for approximating smooth

functions of many variables, and techniques based on compressed sensing and ℓ_1-minimization have proven highly effective at such tasks. Neural networks are also increasingly being applied to such problems, although many questions about their stability and robustness properties and the key issue of sample complexity of deep learning in scientific computing applications remain. This talk presents recent results connecting the theory of best s-term polynomial approximations, (weighted) ℓ_1-minimization, and sparse neural network approximation to answer these questions. We first establish the existence of a training procedure based on sparse polynomial approximation which allows for near-optimal approximation of smooth high-dimensional parameter to solution maps relevant to such applications in computational science and uncertainty quantification. We next connect these results to the problem of approximating such maps with deep neural networks. Our results show the existence of an architecture and training procedure for neural networks that achieves the same rates as sparse polynomial approximation. Our theory is supplemented by numerical experiments demonstrating the practical efficacy of these algorithms.

Ben Adcock
Simon Fraser University
ben_adcock@sfu.ca

Simone Brugiapaglia
Concordia University
simone.brugiapaglia@concordia.ca

Juan M. Cardenas
Simon Fraser University
jcardena@sfu.ca

Nick Dexter
Florida State University, U.S.
nick.dexter@fsu.edu

Sabastian Moraga
Simon Fraser University
sebastian_moraga_scheuermann@sfu.ca

MS44

Double Descent and Conditioning of Random Feature Matrices

We will discuss a sparse random feature method with applications to learning equations from data. We will provide an overview of our theoretical results on the concentration of these random feature matrices, the connections to generalization and complexity bounds, and the design and applications of the method. Examples and applications to high-dimensional modeling and dynamics will be included.

Hayden Schaeffer
Carnegie Mellon University
schaeffer@cmu.edu

MS45

The Approximation Properties of Relu^k Neural Networks

A natural space of functions which can be efficiently approximated by shallow neural networks is the variation space corresponding to the dictionary of single neuron outputs. We will precisely define this space and study its approximation properties. Specifically, we develop techniques for bounding the metric entropy and n-widths of the unit ball in this variation space. These are fundamental quantities in approximation theory that control the limits of linear and non-linear approximation. Consequences of these results include: the optimal approximation rates which can be attained for shallow neural networks, that shallow neural networks dramatically outperform linear methods of approximation, and indeed that shallow neural networks outperform all stable methods of approximation on the associated variation space. Finally, we introduce a class of greedy algorithms and show that they construct asymptotically optimal shallow neural network approximations.

Jonathan Siegel, Jinchao Xu
Penn State University
jus1949@psu.edu, jxx1@psu.edu

MS45

Iterative Regularization for Low Complexity Regularizers

Iterative regularization exploits the implicit bias of an optimization algorithm to regularize ill-posed problems. Constructing algorithms with such built-in regularization mechanisms is a classic challenge in inverse problems but also in modern machine learning, where it provides both a new perspective on algorithms analysis, and significant speed-ups compared to explicit regularization. In this work, we propose and study the first iterative regularization procedure able to handle biases described by non smooth and non strongly convex functionals, prominent in low-complexity regularization. Our approach is based on a primal-dual algorithm of which we analyze convergence and stability properties, even in the case where the original problem is unfeasible. Our theoretical results are complemented by experiments showing the computational benefits of our approach.

Silvia Villa
University of Genova, Italy
villa@dima.unige.it

Lorenzo Rosasco
DIBRIS, MaLGA, Università degli studi di Genova
Italy
lorenzo.rosasco@unige.it

Cesare Molinari
DIMA, MaLGa, Università degli Studi di Genova, Italy
cecio.molinari@gmail.com

Mathurin Massias
MaLGa, University of Genova, Italy
mathurin.massias@gmail.com

MS45

Momentum-Based Variance-Reduced Proximal Stochastic Gradient Method

Stochastic gradient methods (SGMs) have been extensively used for solving stochastic problems or large-scale machine learning problems. Recent works employ various techniques to improve the convergence rate of SGMs for both convex and nonconvex cases. Most of them require a large number of samples in some or all iterations of the improved SGMs. In this talk, we will present a new SGM, named PStorm, for solving nonconvex nonsmooth stochastic problems. With a momentum-based variance reduction tech-

nique, PStorm can achieve the optimal complexity result $O(\varepsilon^{-3})$ to produce a stochastic ε-stationary solution, if a mean-squared smoothness condition holds. Different from existing optimal methods, PStorm can achieve the $O(\varepsilon^{-3})$ result by using only one or $O(1)$ samples in every update. With this property, PStorm can be applied to online learning problems that favor real-time decisions based on one or $O(1)$ new observations. In addition, for large-scale machine learning problems, PStorm can generalize better by small-batch training than other optimal methods that require large-batch training and the vanilla SGM, as we demonstrate on training a sparse fully-connected neural network and a sparse convolutional neural network.

Yangyang Xu
Rensselaer Polytechnic Institute
xuy21@rpi.edu

Yibo Xu
Clemson University
yibox@g.clemson.edu

MS45

Stability and Generalization of Stochastic Gradient Methods

Stochastic gradient methods (SGMs) have become the workhorse behind machine learning. A substantial number of studies are devoted to studying their convergence behavior. In contrast, there is relatively little work on understanding their statistical generalization, i.e., how the learning models built from training data would behave on the future test data. In this talk, I present our systematic work on studying the generalization of both SGD for minimization problems and SGDA for minimax problems using the algorithmic stability approach in the framework of statistical learning theory (SLT). Our theoretical studies remove restrictive assumptions in the literature and significantly improve the existing generalization bounds for SGMs. We show that optimal statistical bounds can be achieved by trading off the computational optimization error and stability results. I will also discuss the application of our new stability study to analyze the generalization of differentially private SGMs.

Yiming Ying
SUNY Albany
yying@albany.edu

MS46

Infeasibility Detection with Primal-Dual Hybrid Gradient for Large-Scale Linear Programming

In this talk, we consider the problem of detecting infeasibility of large-scale linear programming problems using the primal-dual hybrid gradient method (PDHG) of Chambolle and Pock (2011). The literature on PDHG has mostly focused on settings where the problem is assumed to be feasible. When the problem has no solution, the iterates of the algorithm do not converge. In this scenario, we show that the iterates diverge at a controlled rate towards a well-defined ray. The direction of this ray is known as the infimal displacement vector. The first contribution of our work is to prove that this vector recovers certificates of primal and dual infeasibility whenever they exist. Based on this fact, we propose a simple way to extract approximate infeasibility certificates from sequences generated by the iterates of PDHG. Our second contribution is to establish tight convergence rates for these sequences. We prove a convergence rate of $O(k^{-1})$, improving over the known rate of $O(k^{-1/2})$. This rate is general and applies to any fixed-point iteration of a nonexpansive operator. Thus, it is a result of independent interest since it covers a broad family of algorithms, including, for example, ADMM. Further, we show that, under non-degeneracy assumptions, one of these sequences exhibits eventual linear convergence. Numerical experiments support our theoretical findings.

Mateo Diaz
California Institute of Technology,U.S.
mateodd@caltech.edu

David Applegate
Google
dapplegate@google.com

Haihao Lu
University of Chicago
haihao.lu@chicagobooth.edu

Miles Lubin
Google Research
mlubin@google.com

MS46

High Probability Bounds for a Class of Nonconvex Algorithms with AdaGrad Stepsize

In this paper, we propose a new, simplified high probability analysis of AdaGrad for smooth, non-convex problems. More specifically, we focus on a particular accelerated gradient (AGD) template (Lan, 2020), through which we recover the original AdaGrad and its variant with averaging, and prove a convergence rate of $\mathcal{O}(1/\sqrt{T})$ with high probability without the knowledge of smoothness and variance. We use a particular version of Freedman's concentration bound for martingale difference sequences (Kakade Tewari, 2008) which enables us to achieve the best-known dependence of $\log(1/\delta)$ on the probability margin δ. We present our analysis in a modular way and obtain a complementary $\mathcal{O}(1/T)$ convergence rate in the deterministic setting. To the best of our knowledge, this is the first high probability result for AdaGrad with a truly adaptive scheme, i.e., completely oblivious to the knowledge of smoothness and uniform variance bound, which simultaneously has best-known dependence of $\log(1/\delta)$. We further prove noise adaptation property of AdaGrad under additional noise assumptions.

Ali Kavis
EPFL
ali.kavis@epfl.ch

Kfir Levy
Technion
kfirylevy@technion.ac.il

Volkan Cevher
EPFL
volkan.cevher@epfl.ch

MS46

Convergence and Complexity of Stochastic Block Majorization-Minimization

Stochastic majorization-minimization (SMM) is an online extension of the classical principle of majorization-

minimization (MM), which consists of sampling i.i.d. data points from a fixed data distribution and minimizing a recursively defined majorizing surrogate of an objective function. In this talk, we introduce stochastic block MM, where the surrogates can now be only block multi-convex and a single block is optimized at a time within a diminishing radius or with a proximal regularization. Relaxing standard strong convexity requirements for surrogates in SMM, our framework gives wider applicability including online CANDECOMP/PARAFAC (CP) dictionary learning and yields greater computational efficiency especially when the problem dimension is large. We provide an extensive convergence analysis on the proposed algorithm, which we derive under dependent data streams, relaxing the standard i.i.d. assumption on data samples. We show that the proposed algorithm converges almost surely to the set of stationary points of a nonconvex objective under constraints at a rate $O((\log n)^{1+\epsilon}/n^{1/2})$ for the empirical loss and $O((\log n)^{1+\epsilon}/n^{1/4})$ for the expected loss, where n denotes the number of data samples. Under an additional assumption, the latter rate can be improved to $O((\log n)^{1+\epsilon}/n^{1/2})$. Our results provide first convergence rate bounds for various online matrix and tensor decomposition algorithms under a general Markovian data setting.

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

MS46

Three Operator Splitting with Subgradients, Stochastic Gradients, and Adaptive Learning Rates

Three Operator Splitting (TOS) [Davis Yin, 2017] can minimize the sum of multiple convex functions effectively when an efficient gradient oracle or proximal operator is available for each term. This requirement often fails in machine learning applications: (i) instead of full gradients only stochastic gradients may be available; and (ii) instead of proximal operators, using subgradients to handle complex penalty functions may be more efficient and realistic. Motivated by these concerns, we analyze three potentially valuable extensions of TOS. The first two permit using subgradients and stochastic gradients, and are shown to ensure a $\mathcal{O}(1/\sqrt{t})$ convergence rate. The third extension AdapTOS endows TOS with adaptive step-sizes. For the important setting of optimizing a convex loss over the intersection of convex sets AdapTOS attains universal convergence rates, i.e., the rate adapts to the unknown smoothness degree of the objective function. We compare our proposed methods with competing methods on various applications.

Alp Yurtsever
Umeå University
alp.yurtsever@umu.se

Alex Gu, Suvrit Sra
Massachusetts Institute of Technology
gua@mit.edu, suvrit@mit.edu

MS47

Multifidelity Reinforcement Learning

In many computational science and engineering applications, the output of a system of interest corresponding to a given input can be queried at different levels of fidelity with different computational footprints. Typically, low-fidelity data is cheap and abundant, while high-fidelity data is expensive and scarce. Within online learning and control, multi-fidelity data can be exploited in many stages of the decision-making pipeline, including estimation, exploration, and exploitation. In this talk, we present our research on improving the performance of reinforcement learning (RL) agents with multi-fidelity estimation. Specifically, a multifidelity estimator that exploits the cross-correlations between the low- and high-fidelity returns is proposed to reduce the variance in the estimation of the agent's state-action value function. The proposed estimator, which is based on the method of control variates, is used to design a multifidelity Monte Carlo RL (MFMC-RL) algorithm that improves the learning of the agent in the high-fidelity environment. The impacts of variance reduction on policy evaluation and policy improvement are theoretically analyzed by using probability bounds.

Sami Khairy, Prasanna Balaprakash
Argonne National Laboratory
skhairy@hawk.iit.edu, pbalapra@anl.gov

MS47

A Graph Policy Network Approach for Volt-Var Control in Power Distribution Systems

Volt-var control (VVC) is the problem of operating power distribution systems within healthy regimes by controlling actuators in power systems. Existing works uses the conventional routine of representing power systems (a graph with tree topology) as vectors to train deep reinforcement learning (RL) policies. We propose a framework that combines RL with graph neural networks and study the benefits and limitations of graph-based policy in the VVC setting. Our results show that graph policies converge to the same rewards asymptotically, however at a slower rate than the vector representation counterpart. We conduct further analysis on the impact of both observations and actions: On the observation end, we examine the robustness of graph policies on two typical data acquisition errors in power systems, namely sensor communication failure and measurement misalignment. Furthermore, we study the robustness of graph policies to erroneous topological information in the graph representation. Our results reveal that graph policies are significantly more robust than policies with conventional dense network representations. On the action end, we show that actuators have various impacts on the system, thus using a graph representation induced by the physical power systems topology may not be the optimal choice. In the end, we conduct a case study to demonstrate that the choice of readout function architecture and graph augmentation can further improve training performance and robustness.

Xian Yeow Lee, Soumik Sarkar
Iowa State University
xylee@iastate.edu, soumiks@iastate.edu

Yubo Wang
Siemens
yubo.wang@siemens.com

MS47

CityLearn Standardizing Reinforcement Learning for Grid-Interactive Smart Communities

Advanced controllers such as model predictive control and reinforcement learning control have both advantages and disadvantages that prevent them from being implemented in real world problems. Inspired by recently introduced

challenges for real life reinforcement learning control, here we discuss nine real world challenges for reinforcement learning control in grid-interactive buildings. We argue that research in this area should be expressed in this framework. By focusing on the challenges, we can investigate the performance of the controllers under a variety of situations and generate a fair comparison between algorithms. As a demonstration, we implement the offline learning challenge in CityLearn and study the impact of different levels of domain knowledge and complexity of RL algorithms. We show that the sequence of operations utilized in a rule-based controller (RBC) used for offline training affects the performance of the RL agents when evaluated on a set of four energy flexibility metrics. We call for a more interdisciplinary effort of the research community to address the real world challenges and unlock the potential of grid-interactive smart communities.

Zoltan Nagy
University of Texas, U.S.
nagy@utexas.edu

MS47

Learning-Based Control

Control systems with learning abilities could cost-effectively address societal issues like energy reliability, decarbonization, climate security and enable autonomous scientific discovery. Innovation in deep learning methods, tools, and technology offers an unprecedented opportunity to transform the control engineering practice and bring much excitement to control systems theory research. In this talk, I will introduce recent results in modeling dynamic systems with deep learning representations that embed domain knowledge. I will also discuss differentiable predictive control, a data-driven approach that uses physics-informed deep learning representations to synthesize predictive control policies. Ill close by considering the implications of differentiable programming in the broader control systems context.

Draguna L. Vrabie, Jan Drgona, Aaron Tuor
Pacific Northwest National Laboratory
draguna.vrabie@pnnl.gov, jan.drgona@pnnl.gov, aaron.tuor@pnnl.gov

MS48

Multilevel Preconditioning for Stein Variational Gradient Descent with Applications to Bayesian Inverse Problems

We present a multilevel variant of Stein variational gradient descent to more efficiently sample from target distributions. The key ingredient is a sequence of distributions with growing fidelity and costs that converges to the target distribution of interest. For example, such a sequence of distributions is given by a hierarchy of ever finer discretization levels of the forward model in Bayesian inverse problems. Multilevel Stein variational gradient descent moves most of the iterations to lower, cheaper levels with the aim of requiring only a few iterations on the higher, more expensive levels when compared to the traditional, single-level Stein variational gradient descent variant that uses the highest-level distribution only. Under certain assumptions, in the mean-field limit, the error of the proposed multilevel Stein method decays by a log factor faster than the error of the single-level counterpart with respect to computational costs. Numerical experiments with Bayesian inverse problems show speedups of more than one order of magnitude of the proposed multilevel Stein method compared to the single-level variant that uses the highest level only.

Terrence Alsup
New York University
Courant Institute of Mathematical Sciences
alsup@cims.nyu.edu

Luca Venturi
Courant Institute of Mathematical Sciences
New York, United States
lv800@nyu.edu

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

MS48

Neural Operators: Learn to Solve Partial Differential Equations

Traditional deep neural networks are maps between finite dimension spaces, and hence, are not suitable for modeling phenomena such as those arising from the solution of partial differential equations (PDE). We introduce neural operators that can learn operators, which are maps between infinite dimension spaces. By framing neural operators as non-linear compositions of kernel integrations, we establish that they are universal approximators of operators. They are independents of the resolution or grid of training data and allow for zero-shot generalization to higher resolution evaluations. We find that neural operators can solve turbulent fluid flow, seismic wave equation, co2 storage, and many more hard problems with 100000x speedup compared to numerical solvers. I will outline several applications where neural operators have shown order of magnitude speedup.

Kamyar Azizzadenesheli
Purdue University
kaazizzad@gmail.com

MS49

Combining Sequential Monte Carlo Methods with Normalizing Flows: New Perspectives

Annealed Importance Sampling (AIS) and its Sequential Monte Carlo (SMC) extensions are state-of-the-art methods for estimating normalizing constants of probability distributions. We propose here a novel Monte Carlo algorithm, Annealed Flow Transport (AFT), that builds upon AIS and SMC and combines them with normalizing flows (NFs) for improved performance. This method transports a set of particles using not only importance sampling (IS), Markov chain Monte Carlo (MCMC) and resampling steps-as in SMC, but also relies on NFs which are learned sequentially to push particles towards the successive annealed targets. We provide limit theorems for the resulting Monte Carlo estimates of the normalizing constant and expectations with respect to the target distribution. Additionally, we show that a continuous-time scaling limit of the population version of AFT is given by a FeynmanKac measure which simplifies to the law of a controlled diffusion for expressive NFs. We demonstrate experimentally the benefits and limitations of our methodology on a variety of appli-

cations.

Michael Arbel
INRIA Grenoble Rhône-Alpes
michael.n.arbel@gmail.com

MS49

Stochastic Normalizing Flows for Inverse Problems

The aim of this talk will be to present a unified framework for combining many of the commonly used posterior sampling: MCMC, Variational Autoencoder, Langevin sampling and Normalizing Flows. We will outline the basic ideas of those algorithms and present a unified framework with a loss, which allows for the combination of arbitrary layers of those models. We will discuss some theoretical insights, and validate our finding with numerical results. Advantages and disadvantages will be discussed and the relation to the annealed flow transport will be drawn.

Paul Hagemann
TU Berlin
hagemann-p@web.de

MS49

Bayesian Inference with Data-Driven Image Priors Encoded by Conditional Normalising Flows

This talk presents a new empirical Bayesian methodology for performing inference in imaging inverse problems where prior knowledge is available in the form of training data. Adopting a generative modelling approach, and by considering the training data available as a sample from the marginal distribution of the unknown image of interest, we construct a data-driven prior that is encoded by a conditional normalising flow. Following an empirical Bayesian strategy, we use a super-resolution flow that is trained to accurately represent the distribution of high-resolution images conditional to a low-resolution image, which we regard as an unknown parameter of the model that we automatically adjust from the measurement data (not the training data) by marginal maximum likelihood estimation. This empirical Bayesian strategy is efficiently implemented by using a state-of-the-art stochastic approximation proximal gradient algorithm that couples a stochastic optimisation scheme with a proximal MCMC method in order to jointly calibrate the flow model and draw Monte Carlo samples from the posterior distribution of interest. We establish the existence and well-posedness of the posterior distribution and posterior moments under easily verifiable conditions, as well as the convergence of the Bayesian computation scheme, providing a rigorous underpinning for Bayesian estimators and uncertainty quantification analyses. The proposed approach is illustrated with a range of imaging problems.

Matthew Holden
School of Mathematical and Computer Sciences
Heriot Watt University and Maxwell Institute
mdh4@hw.ac.uk

Marcelo A. Pereyra
Maxwell Insitute for Mathematical Science
Heriot-Watt University
m.pereyra@hw.ac.uk

Kostas Zygalakis
University of Edinburgh
K.Zygalakis@ed.ac.uk

MS50

Representation in Fourier Space Reveals a Physical Mechanism for Instability in Data-Driven Models of Multi-Scale Dynamical Systems

Data-driven weather prediction (DDWP) with deep learning (DL) models has seen a surge of interest in recent times. While these models are still not as accurate as operational weather forecasting models, recent successes have seen DDWP models being comparable to them in terms of short-term accuracy at orders-of-magnitude less computational cost. These successes of DDWP models promise us improved performance in terms of sub-seasonal-to-seasonal and extreme weather predictions. However, one major drawback of DDWP models that currently hinders this, is the absence of long-term stability (when integrated autoregressively) and a physical interpretation of the mechanism and cause of this instability. In this paper, we show that a spectral analysis in Fourier space reveals a physical mechanism by which convolution-based DDWP models become unstable. We systematically explore the physics by which instability is introduced in these models by exploring the connection between the representation skills of these networks in Fourier space and a consistent growth of meridional heat flux that leads to an increase in transfer of eddy momentum flux to the mean flow. We further speculate how convolution-based models may have inductive bias that would always lead to such instability and further explorations in novel loss functions or architectures should be explored for autoregressive predictions of weather, climate, and generally fully turbulent flow.

Ashesh K. Chattopadhyay, Ebrahim Nabizadeh
Rice University
akc6@rice.edu, en10@rice.edu

Pedram Hassanzadeh
Mechanical Engineering and Earth Science
Rice University
pedram@rice.edu

MS50

Learning Continuous Models for Continuous Physics

Dynamical systems that evolve continuously over time are ubiquitous throughout science and engineering. Machine learning (ML) provides data-driven approaches to model and predict the dynamics of such systems. A core issue with this approach is that ML models are typically trained on discrete data, using ML methodologies that are not aware of underlying continuity properties, which results in models that often do not capture the underlying continuous dynamics of a system of interest. As a result, these ML models are of limited use for for many scientific and engineering applications. To address this challenge, we develop a convergence test based on numerical analysis theory. Our test verifies whether a model has learned a function that accurately approximates a system's underlying continuous dynamics. Models that fail this test fail to capture relevant dynamics, rendering them of limited utility for many scientific prediction tasks; while models that pass this test enable both better interpolation and better extrapolation in multiple ways. Our results illustrate how principled numerical analysis methods can be coupled with existing ML training/testing methodologies to validate models for sci-

ence and engineering applications.

Michael Mahoney
UC Berkeley
mahoney@stat.berkeley.edu

Aditi Krishnapriyan
UC Berkeley/Lawrence Berkeley National Laboratory
akrishnapriyan@lbl.gov

MS50

Towards Proving the Capacity of Deep Operator Nets for Solving Differential Equations

In recent times machine learning methods have made significant advances in becoming a useful tool for analyzing physical systems. A particularly active area in this theme has been "physics informed machine learning" which largely revolves around using neural nets for numerically solving differential equations. Some of the recent methodological advances in this field include Physics Inspired Neural Nets, Fourier Neural Operators, Deep Ritz Method, Deep Galerkin Method - and maybe most interestingly DeepONets (DONs), whose architecture involves taking an inner-product of two nets, the "branch net" and the "trunk net". There is increasing evidence of the power of DONs in solving differential equations and in this work we start to advance the theory of generalization error bounds for DeepONets. Our key technique is to uncover an upper-bound for the Rademacher complexity of DeepONets with arbitrary dimensional branch and trunk outputs in terms of a sequence of Rademacher complexities for DeepONet classes of lower depths and having one dimensional branch and trunk outputs. Along the way we prove a new kind of a Talagrand contraction lemma which might be of independent mathematical interest.

Anirbit Mukherjee
University of Manchester
anirbit.mukherjee@manchester.ac.uk

Sayar Karmakar
University of Florida
sayarkarmakar@ufl.edu

Pulkit Gopalani
IIT Kanpur
gpulkit@iitk.ac.in

MS50

Developing and Learning from Trust in Neural Networks: Reflections and Progress

Neural networks are becoming established tools for oceanographic study and beyond, showing great potential for the future. The adoption of machine learning as a tool is also timely due to the significant increase in data available from models and observations, where classical tools of analysis may not be sufficient. However, gaining precision in predictions based on, for example, neural networks need not be rooted in the network 'learning' about the system in a meaningful way. Methods exist for assessing the source of skill for neural networks for example through Additive Feature Attribution, but these are themselves uncertain. In this talk, I will present work that addresses the problem of quantifying uncertainty in out-of-sample predictions and the extent to which methods of understanding the source of predictive skill are fit for purpose within oceanography and beyond. This is made possible using an oceanographic framework rooted in theoretical understanding, enabling a critical assessment of methods.

Maike Sonnewald
Princeton University
maikes@princeton.edu

MS51

Deep Ensembles for Graphs with Higher Order Dependencies

Graph neural networks (GNNs) continue to achieve state-of-the-art performance on many graph learning tasks. However, in the presence of higher-order sequential dependencies, traditional graph representations can tend to underfit each node's neighborhood to diminish the performance of a GNN. To address this, we developed a novel Deep Graph Ensemble (DGE), which captures neighborhood variance by training an ensemble of GNNs on different neighborhood subspaces of the same node within a higher-order network structure. We show that DGE consistently outperforms existing a single GNN on semisupervised and supervised tasks on four real-world data sets with known higher-order dependencies, even under a similar parameter budget.

Nitesh Chawla
University of Notre Dame
nchawla@nd.edu

MS51

Superfast Generation of Large Hypergraph Data

We present a straightforward paradigm for designing simple and efficient generators for a wide variety of hypergraph models. This approach yields asymptotically optimal algorithms and concise implementations that outperform existing implementations by more than an order of magnitude. Our generalization of ball dropping, called functional ball dropping, allows us to generate models that have not been previously treated by ball dropping. We identify natural building blocks of a given graph model (e.g an edge) that can be concatenated to form a sensible representation of a graph (e.g. a list of edges). Next, we design a sampler subject to two constraints: the samplers probability distribution must exactly match the expectation values of the model we generate, and the runtime of the sampler must be equal to the size of the component it samples. The resulting algorithm incrementally generates the graph by repeated or parallelized execution of the sampling function.

Nicole Eikmeier
Grinnell College
eikmeier@grinnell.edu

MS51

Can't Keep Up: Extracting Dense Subgraphs under High Data Rates

Identifying dense structures within large, sparse graphs has been shown to be useful for numerous problems. Many graphs today are dynamic: they are derived from continuously changing data and hence undergo continuous vertex and edge changes. Can we identify dense structures quickly within such graphs, especially when the data change rates are high? We focus on nuclei, a generalization of cores and trusses. Prior work has developed scalable, batch parallel maintenance algorithms for nucleus decompositions with trade-offs between minimum latency and variability. The

main approach is to first maintain a special hypergraph derived from small cliques in the original graph and second maintain cores on that hypergraph. We provide an overview of the existing maintenance algorithms. We then consider when the data rate continues to rise beyond the throughput limits of such algorithms. In this case we cannot keep up; we extend prior algorithms to provide heuristic results during periods of high data rates and then recover to exact results during less demanding periods. We empirically demonstrate the ability to handle high data rates and show that maintenance of dense subgraph extraction is feasible on today's real-world rapidly changing graphs.

Kasimir Gabert
Georgia Institute of Technology
kasimir@gatech.edu

MS52

Data Augmentation in High Dimensional Low Sample Size Setting Using a Geometry-Based Variational Autoencoder

In this presentation, we propose a new method to perform data augmentation in a reliable way in the High Dimensional Low Sample Size (HDLSS) setting using a geometry-based variational autoencoder. Our approach combines a proper latent space modeling of the VAE seen as a Riemannian manifold with a new generation scheme which produces more meaningful samples especially in the context of small data sets. The proposed method is tested through a wide experimental study where its robustness to data sets, classifiers and training samples size is stressed. It is also validated on a medical imaging classification task on the challenging ADNI database where a small number of 3D brain MRIs are considered and augmented using the proposed VAE framework. In each case, the proposed method allows for a significant and reliable gain in the classification metrics. For instance, balanced accuracy jumps from 66.3% to 74.3% for a state-of-the-art CNN classifier trained with 50 MRIs of cognitively normal (CN) and 50 Alzheimer disease (AD) patients and from 77.7% to 86.3% when trained with 243 CN and 210 AD while improving greatly sensitivity and specificity metrics.

Clement Chadebec
Université de Paris, INRIA / INSERM
clement.chadebec@inria.fr

Stephanie M. Allassonniere
CMAP Ecole Polytechnique
Stephanie.Allassonniere@polytechnique.edu

MS52

Predicting Solar Wind Streams from the Inner-Heliosphere to Earth via Shifted Operator Inference

Solar wind conditions are predominantly predicted via three-dimensional numerical magnetohydrodynamic (MHD) models. Despite their ability to produce highly accurate predictions, MHD models require computationally intensive high-dimensional simulations. This renders them inadequate for making time-sensitive predictions and for large-ensemble analysis required in uncertainty quantification. This paper presents a new data-driven reduced-order model (ROM) capability for forecasting heliospheric solar wind speeds. Traditional model reduction methods based on Galerkin projection have difficulties with advection-dominated systems—such as solar winds—since they require a large number of basis functions and can become unstable. A core contribution of this work addresses this challenge by extending the non-intrusive operator inference ROM framework to exploit the translational symmetries present in the solar wind caused by the Sun's rotation. The numerical results show that our method can adequately emulate the MHD simulations and outperforms a reduced-physics surrogate model, the Heliospheric Upwind Extrapolation model.

Opal Issan
University of California, San Diego
oissan@ucsd.edu

Boris Kramer
University of California San Diego
bmkramer@ucsd.edu

MS52

Exploiting Derivative Information in Parametric Surrogate Modeling

Outer-loop problems arising in scientific applications (such as optimization, uncertainty quantification and inverse problems) require repeated evaluation of computationally intensive numerical models, such as those arising from discretization and solution of ordinary and partial differential equations. The cost of these evaluations makes solution using the model prohibitive, and efficient accurate surrogates are a key to solving these problems in practice. In this talk we will discuss how compressed derivative information can be exploited to aid the design of and training of parsimonious neural network architectures to be deployed in high dimensional inference. These reduced-basis architectures outperform conventional data-driven approaches when limited training data are available due to computational costs of evaluating high dimensional nonlinear PDEs. Additionally these trained surrogates can provide derivative information that can be used to accelerate derivative based inference in high dimensions, where other strategies cannot provide such information.

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

MS52

Stabilizing Dynamical Systems in the Scarce Data Regime

Stabilizing dynamical systems in science and engineering is challenging, especially in edge cases and limit states where typically little data are available. In this work, we propose a data-driven approach that guarantees finding stabilizing controllers from as few data samples as the dimension of the unstable dynamics, which typically is orders of magnitude lower than the state dimension of the system. The key is learning stabilizing controllers directly from data without learning models of the systems, which would require larger numbers of data points. Numerical experiments with chemical reactors and fluid dynamics behind obstacles demonstrate that the proposed approach stabilizes systems after observing fewer than five data samples even though the dimension of states is orders of magnitude higher.

Steffen W. R. Werner, Benjamin Peherstorfer

Courant Institute of Mathematical Sciences
New York University
steffen.werner@nyu.edu, pehersto@cims.nyu.edu

MS53

Dictionary-Based Low-Rank Approximations

Constrained tensor and matrix factorization models allow to extract interpretable patterns from multiway data. Therefore crafting efficient algorithms for constrained low-rank approximations is nowadays an important research topic. This work deals with columns of factor matrices of a low-rank approximation being sparse in a known and possibly overcomplete basis, a model coined as Dictionary-based Low-Rank Approximation (DLRA). While earlier contributions focused on finding factor columns inside a dictionary of candidate columns, *i.e.* one-sparse approximations, this work is the first to tackle DLRA with sparsity larger than one. I propose to focus on the sparse-coding subproblem coined Mixed Sparse-Coding (MSC) that emerges when solving DLRA with an alternating optimization strategy. Several algorithms based on sparse-coding heuristics (greedy methods, convex relaxations) are provided to solve MSC. The performance of these heuristics is evaluated on simulated data. Then, I show how to adapt an efficient MSC solver based on the LASSO to compute Dictionary-based Matrix Factorization and Canonical Polyadic Decomposition in the context of hyperspectral image processing and chemometrics. These experiments suggest that DLRA extends the modeling capabilities of low-rank approximations, helps reducing estimation variance and enhances the identifiability and interpretability of estimated factors.

Jeremy Cohen
Universite de Rennes, INRIA, CNRS, IRISA
jeremy.cohen@cnrs.fr

MS53

Extension of Correspondence Analysis to Multi-Way Data-Sets Through HOSVD: a Geometric Framework

We present an extension of Correspondence Analysis (CA) to tensors through High Order Singular Value Decomposition (HOSVD) from a geometric viewpoint. Correspondence analysis is a well-known tool in data science, developed from principal component analysis and applied to contingency tables. Indeed, CA associates a two-ways contingency table with two point clouds, linked by a characteristic barycentric relation. Different algebraic extensions of CA to multi-way tables have been proposed over the years, nevertheless neglecting its geometric meaning, as far as we are aware. Relying on the Tucker model and the HOSVD, we propose a direct way to associate with each tensor mode a point cloud. We prove that the point clouds are related to each other. Specifically using the CA metrics we show that the barycentric relation is still true in the tensor framework. Finally, numerical examples are used to underline the advantages and the drawbacks of our strategy with respect to the classical matrix approaches.

Olivier Coulaud
INRIA
olivier.coulaud@inria.fr

Alain Franc
INRAE, Université de Bordeaux
33610 Cestas, France alain.fra
alain.franc@inrae.fr

Martina Iannacito
INRIA Bordeaux Sud-Ouest
martina.iannacito@inria.fr

MS53

Towards Practical Lipschitzness of Neural Networks Using Tensor Decompositions

Functional properties of convolutional neural networks (CNN) such as their Lipschitz constant remain an under-explored area of machine learning research, impacting its practical applications. The CNN Lipschitz constant affects mission-critical model properties, such as generalization and adversarial robustness. Modern approaches to enforcing the convolutional layer Lipschitz constant rely on expensive parameterizations, resort to coarse approximations, or present an intractable computational challenge. In this talk I will present a new approach based on tensor decompositions that alleviates constraints of the prior art at the expense of an insignificant reduction in layer expressivity.

Alexandra Senderovich, Ekaterina Bulatova
HSE University
alexsend57@gmail.com, evbulatova@edu.hse.ru

Anton Obukhov
ETH Zurich
anton.obukhov@vision.ee.ethz.ch

Maxim Rakhuba
National Research University Higher School of Economics
mrakhuba@hse.ru

MS53

Tensor Decompositions for Multi-Aspect Graph Analytics and Beyond

Tensors decompositions have been very popular and effective tools for analyzing multi-aspect data in a wide variety of fields, ranging from Psychology to Chemometrics, and from Signal Processing to Data Mining and Machine Learning. In this talk, we will demonstrate the effectiveness of tensor decompositions in modeling and mining multi-aspect graphs. We will showcase tensor-based node embeddings and their performance compared to state-of-the-art. Subsequently, we will explore recent results that demonstrate the effectiveness of tensor methods in alleviating state-of-the-art adversarial attacks in Graph Convolutional Networks. Finally, we conclude with the interplay of tensor methods and deep generative models, going over some recent preliminary results and exploring the future outlook of the intersection of tensor methods and deep learning.

Vagelis Papalexakis
UC Riverside
epapalex@cs.ucr.edu

William Shiao
University of California, Riverside, U.S.
wshia002@ucr.edu

MS54

Settling the Sample Complexity of Model-Based

Offline Reinforcement Learning

This work is concerned with offline reinforcement learning (RL), which learns using pre-collected data without further exploration. Effective offline RL would be able to accommodate distribution shift and insufficient data coverage. However, prior algorithms either suffer from suboptimal sample complexities or incur high burn-in cost, thus posing an impediment to efficient RL in sample-starved applications. In this work, we demonstrate that the model-based (or "plug-in") approach achieves minimax-optimal sample optimality with minimal burn-in cost. Our algorithms are pessimistic variants of value iteration with Bernstein-style penalties, which do not rely on sophisticated schemes like variance reduction.

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

MS54

Discovering Physically Meaningful Low-Dimensional Structures in Spatiotemporal Data

Current deep learning models for spatiotemporal forecasting struggle with generalization. They can only forecast in a specific domain and fail when applied to systems with different parameters, external forces, or boundary conditions. In this talk, I will discuss how to leverage physically meaningful low-dimensional structures to improve generalization. We propose a model-based meta-learning method called DyAd that can generalize across heterogeneous forecasting tasks. DyAd has two parts: an encoder that infers the time-invariant hidden features using weak supervision, and a forecaster which learns the shared dynamics. The encoder adapts and controls the forecaster during inference. Theoretically, we guarantee the generalization error of such a procedure. Experimentally, we demonstrate that our model outperforms state-of-the-art approaches on both turbulent flow and real-world ocean data forecasting tasks.

Rui Wang, Rose Yu
University of California, San Diego, U.S.
ruw020@ucsd.edu, roseyu@ucsd.edu

MS54

Robust Learning by Over-Parameterization

Recently, over-parameterized models (e.g., deep neural networks) with more parameters than the amount of data have dominated the performances of modern machine learning. However, when the training data is corrupted, it is well-known that over-parameterized models overfit and do not generalize. While there exist many explicit regularizers such as weight decay for alleviating overfitting, they cannot fully separate data corruption and recover clean data. In this talk I will show that overfitting to training data corruption can be simply addressed by introducing an extra over-parameterization term to the already over-parameterized model. Our method, called Double Over-Parameterization (DOP), contradicts classical wisdom that more parameters aggregate rather than alleviate overfitting. Crucially, we avoid overfitting by leveraging the implicit regularization of gradient descent with a particular choice of learning rates on different sets of parameters. When applied to the tasks of image recovery from corrupted measurement and image classification under label noise, we show that DOP exhibits superior empirical performance. Finally, when applied to the task of robust principal component analysis, we prove that DOP can exactly recover a low-rank matrix from sparsely corrupted measurements, without prior knowledge on neither rank of the matrix nor sparsity of the corruption.

Chong You
Google
cyou@google.com

MS55

Depth-Resolved Diffraction with Coded-Apertures

We will introduce a computational imaging method to image the internal microstructure of crystalline materials using Laue diffraction. Our method relies on scanning a coded aperture across the diffracted x-ray beams from a broadband illumination, and a reconstruction algorithm to resolve Laue patterns as a function of depth along the incident illumination path. This method provides a rapid access to full diffraction information at sub-micrometer volume elements in bulk materials. We will describe the underlying theory and the experimental validation, as well as the roadmap for this new imaging approach.

Doga Gursoy
Advanced Photon Source
Argonne National Laboratory
dgursoy@anl.gov

Dina Sheyfer, Michael Wojcik, Wenjun Liu, Jonathan Tischler
Argonne National Laboratory
dsheyfer@anl.gov, mwojcik@anl.gov, wjliu@anl.gov, tischler@anl.gov

MS55

Optimal Control Framework for Deep Autoencoders

In this talk, we introduce an optimal control and low rank tensor framework for autoencoder type deep neural networks (DNNs). The learning problem is an optimization problem subject to two differential equations, representing encoder and decoder, as constraints. This approach is mathematically rigorous and offers multiple advantages, like a compressed network due to rank reduction, significant memory savings, and the ability to train on small data. We show successful application of this network in various computer vision tasks like image denoising and image deblurring. DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

Ratna Khatri
U.S. Naval Research Laboratory
ratna.khatri@nrl.navy.mil

MS56

Boundary Estimation from Point Clouds

We investigate identifying the boundary of a domain from sample points in the domain. We introduce new estimators for the normal vector to the boundary, distance of a point to the boundary, and a test for whether a point lies within a boundary strip. The estimators can be efficiently computed and are more accurate than the ones present in the literature. We use the detected boundary points to solve boundary-value problems for PDE on point clouds and con-

sider applications to data depth in machine learning. This is joint work with Sangmin Park and Dejan Slepcev.

Jeff Calder
University of Minnesota
jcalder@umn.edu

MS56

From Fermat Graph Laplacians to Density Accelerated Diffusions

Fermat distances are an optimal path metric which balance density-based and geometric information present in data. Graph Laplacians constructed with Fermat distance provide a useful tool for obtaining sparse graph cuts, for dealing with elongated data structures, and for automatically detecting the number of clusters. As the sample size converges to infinity, the spectrum of the discrete Fermat Graph Laplacian converges to the spectrum of a continuum Kolmogorov operator which generates a diffusion on the associated manifold. Unlike the Euclidean case where the diffusion occurs at a constant speed, the resulting diffusion is accelerated in regions of high density, allowing for the rapid exploration of elongated data structures. This talk will discuss some of the desirable properties of Fermat Graph Laplacians and highlight how the continuum limit provides a theoretical framework for understanding these properties.

Anna Little
University of Utah
little@math.utah.edu

MS56

Efficient Distribution Classification Tasks via Optimal Transport Embeddings

Detecting differences and building classifiers between distributions, given only finite samples, are important tasks in a number of scientific fields. This motivated the introduction of Linear Optimal Transportation (LOT), which embeds the space of distributions into an L2-space using the optimal transport framework. The transform is defined by computing the optimal transport of each distribution to a fixed reference distribution, and has a number of benefits when it comes to speed of computation and to determining classification boundaries. In this talk, we characterize settings in which LOT embeds families of distributions into a space in which they are linearly separable. This is true in arbitrary dimension, and for families of distributions generated by simple group actions. The proposed framework significantly reduces both the computational effort and the required training data in supervised settings. We demonstrate the benefits in pattern recognition tasks in imaging and provide some medical applications.

Caroline Moosmueller
University of California, San Diego
cmoosmueller@ucsd.edu

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

Harish Kannan, Varun Khurana
University of California, San Diego
hkannan@ucsd.edu, vkhurana@ucsd.edu

Jinjie Zhang
University of Chicago
University of California San Diego
jiz003@ucsd.edu

MS56

How Spectral Methods Adapt to Data Sets of Mixed Dimensions

High dimensional data often consist of parts with different intrinsic dimension. We study how spectral methods on graphs adapt to data containing intersecting pieces of different dimensions. We show that unnormalized Laplacian only selects one dimension, while appropriately normalized Laplacian converges to multiple of Laplace-Beltrami operator in all dimensions simultaneously. For intersecting manifolds we identify the trace condition between the manifold Laplacians.

Leon Bungert
University of Bonn
leon.bungert@hcm.uni-bonn.de

Simone Di Marino
Università di Genova
simone.dimarino@unige.it

Dejan Slepcev
Carnegie Mellon University
slepcev@math.cmu.edu

MS57

Intrinsic Subspaces of High-Dimensional Inverse Problems and Where to Find Them

The high-dimensionality is a central challenge faced by many numerical methods for solving large-scale Bayesian inverse problems. In this talk, we will present some recent developments in the identification of low-dimensional subspaces that offer a viable path to alleviating this dimensionality barrier. Utilising concentration inequalities, we are able to identify the intrinsic subspaces from the solutions of certain eigenvalue problems and derive corresponding dimension-truncation error bounds. The resulting low-dimensional subspace enables the design of inference algorithms that can scale sub-linearly with the apparent dimensionality of the problem.

Tiangang Cui
Monash University
tiangang.cui@monash.edu

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

Xin T. Tong
National University of Singapore
Department of Mathematics
tongthomson@gmail.com

Olivier Zahm
Univ. Grenoble Alpes, Inria
olivier.zahm@inria.fr

MS57

Bayesian Inversion of a Coupled Acoustic-Gravity

Model for Predictive Tsunami Simulation

To improve tsunami preparedness, early-alert systems and real-time monitoring are essential. We propose a novel approach for predictive tsunami modeling within the Bayesian inversion framework. This effort focuses on informing the immediate response to an occurring tsunami event using near-field data observation. Our forward model is based on a coupled acoustic-gravity model (e.g., Lotto and Dunham, Comput Geosci (2015) 19:327–340). Similar to other tsunami models, our forward model relies on transient boundary data describing the location and magnitude of the seafloor deformation. In a real-time scenario, these parameter fields must be inferred from a variety of measurements, including observations from pressure gauges mounted on the seafloor. One particular difficulty of this inference problem lies in the accurate inversion from sparse pressure data recorded in the near-field where strong hydroacoustic waves propagate in the compressible ocean; these acoustic waves complicate the task of estimating the hydrostatic pressure changes related to the forming surface gravity wave. Furthermore, the forward model incurs a high computational complexity, since the pressure waves must be resolved in the 3D compressible ocean over a sufficiently long time span. Due to the infeasibility of rapidly solving the corresponding inverse problem for the fully discretized space-time operator, we explore options for using surrogate operators of the parameter-to-observable map.

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

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

MS57

Joint Parameter and State Dimension Reduction for Bayesian Ice Sheet Inverse Problems

Solving large-scale Bayesian inverse problems governed by complex models suffers from the twin difficulties of the high dimensionality of the uncertain parameters and computationally expensive forward models. In this talk, we focus on 1. reducing the computational cost when exploring the posterior via joint parameter and state dimension reduction and 2. accounting for the error due to using a reduced order forward model via Bayesian Approximation Error (BAE). To reduce the parameter dimension, we exploit the underlying problem structure (e.g., local sensitivity of the data to parameters, the smoothing properties of the forward model, the fact that the data contain limited information about the (infinite-dimensional) parameter field, and the covariance structure of the prior) and identify a likelihood-informed parameter subspace that shows where the change from prior to posterior is most significant. For the state dimension reduction, we employ a proper orthogonal decomposition (POD) combined with the discrete empirical interpolation method (DEIM) to approximate the nonlinear term in the forward model. We illustrate our approach with a model ice sheet inverse problem governed by the nonlinear Stokes equation for which the basal sliding coefficient field is inferred from the surface ice flow velocity. The results show the potential to make the exploration of the full posterior distribution of the parameter or subsequent predictions more tractable.

Noemi Petra
University of California, Merced
npetra@ucmerced.edu

Ki-Tae Kim
University of California Merced
kkim107@ucmerced.edu

Tiangang Cui
Monash University
tiangang.cui@monash.edu

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

MS58

A Comparative Study of Time Series Machine Learning Algorithms for Field Production Forecasting

Production forecasting is at the heart of maximizing recovery from a hydrocarbon fields. Production forecasting is critical to many field management tasks such as production optimization and wells and equipment maintenance. Such data is particularly important to fields with high well count due to the need to prioritize actions. Challenges for accurate forecasting of production data include lack, or inadequacy, of physics-based models, data availability and quality, and system disruption. Manual identification of candidate wells and appropriate optimization or maintenance actions is time consuming, prone to inconsistency, and suffers from data quality issues. In this presentation we present a performance compassion between few machine learning models for time series forecasting. The list of algorithm considered include Tree-based models (LightGBM, XGBoost, Random Forest), Neural Network (MLP and LSTM) and Generalized Linear Models. We finally discuss key factors for successful implementation such as design, validation, deployment, and human-centric use of forecasting uncertainty.

Akriti Sharma
Exxonmobil
akriti.sharma@exxonmobil.com

Alex Ahn
ExxonMobil
chong.h.ahn@exxonmobil.com

Jihoon Park
Exxonmobil
jihoon.park@exxonmobil.com

Sharad Shashank
Georgia Institute of Technology
shashank.sharad@gmail.com

Abhishek Mishra
exxonmobil.com
abhishek.mishra1@exxonmobil.com

Amr El-Bakry
ExxonMobil Upstream Research
amr.s.el-bakry@exxonmobil.com

MS58

Building Predictive Models for Rare eEvents in

Unconventional Production Wells

Unplanned downtime events are significant contributors to production loss in oil and gas operations, particularly in unconventional settings with high well counts. Predicting such events before they occur allows operations to become more proactive, remediating conditions to avoid downtime or reducing the amount of time a well is shut-in for maintenance. Predictive maintenance approaches have been applied with great success in several similar manufacturing and industrial applications. However, there are unique challenges that need to be considered for application to producing wells. Notably, the reservoirs in which wells operate are continuously evolving and the state at any time is only partially known from data. The available time series data used for prediction is dominated by pressure related data, which are sensitive not only on the operational health of the well but also to non-stationary reservoir conditions. Furthermore, data sets are highly imbalanced; both with respect to the amount of time spent in normal operation versus near downtime and in the context of identifying and differentiating downtime related to a specific system or condition from all other possible downtime events in the well. This talk will review how we have addressed general challenges of predictive maintenance and specific challenges related to oil and gas producing wells in field application, as well as how these challenges are prompting new research into physics informed approaches.

Vanessa Nenna, Divakar Gollu
ExxonMobil Upstream Integrated Solutions
Vanessa.nenna@exxonmobil.com,
divakar.gollu@exxonmobil.com

Amit Kumar
ExxonMobil Upstream Research Company
amit.x.kumar@exxonmobil.com

Xiao Zhang, Frank Ceballos
ExxonMobil Upstream Integrated Solutions
xiao.zhang@exxonmobil.com,
frank.ceballos@exxonmobil.com

MS59

Accelerating Hydrodynamics Simulations with Reduced Order Models

As a mathematical model of high-speed flow and shock wave propagation in a complex multimaterial setting, Lagrangian hydrodynamics is characterized by moving meshes, advection-dominated solutions, and moving shock fronts with sharp gradients. These challenges hinder the existing projection-based model reduction schemes from being practical. We present several variations of projection-based reduced order model techniques for Lagrangian hydrodynamics by introducing three different reduced bases for position, velocity, and energy fields, with a manifold decomposition approach to address the challenge imposed by the advection-dominated solutions. Lagrangian hydrodynamics is formulated as a nonlinear problem, which requires a proper hyper-reduction technique. Therefore, we present the over-sampling DEIM and SNS approaches to reduce the complexity due to the nonlinear terms. Finally, we also present both a posteriori and a priori error bounds associated with our reduced order model, and the performance comparison of the reduced order modeling approaches in terms of accuracy and speed-up with respect to the corresponding full order model for several numerical examples, namely Sedov blast, Gresho vortices, Taylor-Green vortices, triple-point problem and Rayleigh-Taylor instability problem.

Siu Wun Cheung, Youngsoo Choi
Lawrence Livermore National Laboratory
cheung26@llnl.gov, choi15@llnl.gov

Dylan M. Copeland
Lawrence Livermore National Laboratory, US
copeland11@llnl.gov

Kevin Huynh
Lawrence Livermore National Laboratory
huynh24@llnl.gov

MS59

Filtering in Non-Intrusive Data-Driven Reduced Modeling of Large-Scale Systems

We present a method for enhancing data-driven reduced-order modeling with a preprocessing step in which the training data are filtered prior to training the reduced model. Filtering the data prior to training has a number of benefits for data-driven modeling: it attenuates (or even eliminates) frequency content that would otherwise be difficult or impossible to capture with the reduced model, it smoothens discontinuities in the data that would be difficult to capture in a low-dimensional representation, and it reduces noise in the data. This makes the reduced modeling learning task numerically better conditioned, less sensitive to numerical errors in the training data, and, especially important in real-world applications, less prone to overfitting when the amount of training data is limited. We illustrate the effects of filtering in large-scale rotating detonation rocket engine simulations with millions of spatial degrees of freedom for which only a few hundred downsampled training snapshots are available. A reduced-order model is derived from these snapshots using operator inference. Our results indicate the potential benefits of filtering to reduce overfitting, which is particularly important for complex physical systems where the amount of training data is limited.

Ionut-Gabriel Farcas
The University of Texas at Austin
Oden Institute for Computational Engineering and Sciences
ionut.farcas@austin.utexas.edu

Ramakanth Munipalli
Edwards Air Force Base
Air Force Research Laboratory
ramakanth.munipalli@us.af.mil

Karen E. Willcox
UT Austin
kwillcox@oden.utexas.edu

MS59

Data-Driven Enhanced Model Reduction for Bifurcating Models in Computational Fluid Dynamics

We present various data-driven methods to enhance projection-based model reduction techniques with the aim of capturing bifurcating solutions. To show the effectiveness of the data-driven enhancements, we focus on the incompressible Navier-Stokes equations and different types of bifurcations. To recover solutions past a Hopf bifurcation, we propose an approach that combines proper orthogonal

decomposition with Hankel dynamic mode decomposition. To approximate solutions close to a pitchfork bifurcation, we combine localized reduced models with artificial neural networks. Several numerical examples are shown to demonstrate the feasibility of the proposed approaches.

Martin Hess
SISSA, International School for Advanced Studies
Trieste, Italy
mhess@sissa.it

Annalisa Quaini
Department of Mathematics, University of Houston
quaini@math.uh.edu

Gianluigi Rozza
SISSA, International School for Advanced Studies,
Trieste, Italy
grozza@sissa.it

MS59

Advances in Parameter Space Reduction with Active Subspaces

Nowadays, parameter space reduction represents a critical aspect for many outer-loop applications such as optimization. This is due to the increased available computational power and to the need of more sophisticated industrial artifacts. One of the most widespread linear methods to reduce the dimensionality of the input design space is the active subspaces (AS) technique. In this talk we present a recent localization extension of AS which exploits a supervised distance metric for regression and classification tasks. This represents a further step to go beyond global linear models. We also present a multi-fidelity extension involving AS to increase the accuracy of Gaussian process response surfaces. We show that incorporating a low-dimensionality bias could greatly improve the regression capabilities for scalar quantities of interest without the need of running low fidelity simulations. We are able to construct better models with the same amount of data. Industrial applications will be presented.

Gianluigi Rozza
SISSA, International School for Advanced Studies,
Trieste, Italy
grozza@sissa.it

Marco Tezzele
International School of Advanced Studies (SISSA)
Trieste, Italy
marco.tezzele@austin.utexas.edu

Francesco Romor
SISSA, International School for Advanced Studies
Trieste, Italy
fromor@sissa.it

MS60

Uncovering the Spatial Landscape of Molecular Interactions Within the Tumor Microenvironment Using Latent Spaces

Spatial transcriptomics (ST) technologies enable us to measure gene expression in the tissue samples while retaining their spatial context. Such spatially-resolved data enables *in situ* resolution of the regulatory pathways in the heterogeneous tumor and its microenvironment (TME). Direct characterization of cellular co-localization using spatial technologies enables quantification of molecular changes caused by direct cell-cell interaction, such as that seen in tumor-immune interactions. Spot-based ST technologies, on the other hand, do not measure gene expression in individual cells but rather in groups of 1-10 cells, obscuring the constituent cell types as well as their states. Matrix factorization methods can be used to deconvolve the ST data to infer constituent cell populations and cell activities in each spot. I will discuss how unsupervised nonnegative matrix factorization (NMF) of the ST data can be used to identify spatially-resolved latent features that mirror pathologists annotations on the tissue samples. I will then describe our work to identify molecular changes and pathways associated with cell-cell interactions under the assumption that spatially overlapping latent features associated with different cell types interact in regions where they overlap. We apply these strategies to infer molecular changes from tumor-immune interactions in ST data from metastasis, invasive and precursor lesions, and immunotherapy treatments.

Atul Deshpande
Division of Biostatistics and Bioinformatics, Department of
Johns Hopkins University
adeshpande@jhu.edu

MS60

Pattern Discovery in Time-Course Omics Data Using Non-Negative CP tensor decomposition (NCPD)

Datasets associated with bioinformatics studies that include time-course analysis of omics data naturally lend themselves to a 3-mode tensor structure of features-by-subject-by-time. A non-negative CANDECOMP/PARAFAC (CP) decomposition (NCPD) of such datasets can reveal temporal patterns of feature (such as gene) expression that are associated with different subject groups, and may be assessed for association with clinical features. We show the application of an NCPD pipeline to reveal novel structure and biological observations from immune response profiling studies against the Influenza (flu) and Bordetella Pertussis pathogens. We further consider an extension of NCPD which can capture the underlying geometry of the data using a Wasserstein distance in lieu of the Frobenius distance, which is used in the ALS and OPT objective functions of the original NCPD decomposition. We compare the performance of NCPD-F (NCPD with the Frobenius distance) to NCPD-W (NCPD with the Wasserstein distance) using both real and synthetic datasets and discuss ongoing NCPD-W research to improve the quality and robustness of NCPD application to bioinformatics datasets.

Shoaib Bin Masud
Dept. Electrical and Computer Engineering
Tufts University
Shoaib_Bin.Masud@tufts.edu

Anna Konstorum
Department of Pathology
Yale School of Medicine
anna.konstorum@yale.edu

Misha E. Kilmer
Mathematics Department
Tufts University
misha.kilmer@tufts.edu

Shuchin Aeron
Tufts University
shuchin@ece.tufts.edu

MS60

Mechanistic and Data-Driven Dissection of Cell Communication Through Tensor Decompositions

Studies of even simple cell responses to their environment are hindered by how responses are multi-dimensional. For example, a simple receptor-ligand pathway can display differing responses based on timescale, cell type, stimulation, type of response measured, and context. Interrogating and manipulating these systems is thus almost always constrained by an incomplete view of the overall pathway. Like how principal component analysis uses a low-rank approximation for dimensionality reduction of matrix-structured data, tensor generalizations provide solutions for pattern recognition in data with a higher-dimensional structure. Using several recent and unpublished applications, including engineering cell-type selective IL-2 therapies and serology analysis, I will describe some of the unique benefits of tensor-based analysis and the biological discoveries it has revealed. Specifically, tensor approximations enable more effective dimensionality reduction, separation of dimension-specific effects, and a natural, flexible solution to data integration. Finally, I will discuss some of the reasons tensor-based methods remain limited in their application to molecular biology. Resolving these limitations, and applying tensor methods in a more widespread manner, will help provide a complete view of cellular communication.

Aaron S. Meyer
Dept. of Bioengineering
UCLA
ameyer@ucla.edu

MS60

Tensor and Higher-Order Generalizations of the GSVD with Applications to Personalized Cancer Medicine

The number of high-dimensional datasets recording multiple aspects of interrelated phenomena is increasing in many areas, from medicine to finance. This drives the need for mathematical frameworks that can simultaneously identify the similar and dissimilar among multiple matrices and tensors, and thus create a single coherent model from the multiple datasets. The generalized singular value decomposition (GSVD) was formulated as such a comparative spectral decomposition of two column-matched but row-independent matrices. I will, first, define a higher-order GSVD (HO GSVD) and a tensor GSVD and prove that they extend almost all of the mathematical properties of the GSVD to multiple matrices and two tensors. Second, I will describe the development of a tensor HO GSVD for multiple tensors. Third, I will describe the use of these decompositions in the comparisons of cancer and normal genomes, where they uncover patterns that predict survival and response to treatment. The data had been publicly available for years, but the patterns remained unknown until the data were modeled by using the decompositions, illustrating their ability to find what other methods miss.

Sri Priya Ponnapalli
Amazon Machine Learning (ML) Solutions Lab
sripriyaponnapalli@gmail.com

MS61

Neural Galerkin and Active Learning for High-Dimensional Evolution Equations

While machine learning methods have been shown to provide accurate predictions when trained on sufficient data, many of the scientifically most interesting phenomena happen in regimes where there is no data available a priori and where it is even unclear how to collect informative data at all. In this work, we propose the Neural Galerkin methodology that integrates data acquisition into the process of solving partial differential equations with deep learning so that new data samples are collected in a self-informed manner that is guided by the dynamics of the solution itself. Numerical experiments demonstrate that the adaptive data collection of Neural Galerkin is key to providing accurate approximations of solutions in high dimensions, especially if features of the solutions are local such as in interacting particle systems described by kinetic equations and when advecting coherent structures and waves in high dimensions.

Joan Bruna
New York University
Courant Institute
bruna@cims.nyu.edu

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

Eric Vanden Eijnden
Courant Institute
New York University
eve2@cims.nyu.edu

MS61

Cost-Accuracy Trade-Offs in Learning PDE Operators with Neural Networks

The term 'surrogate modeling' in computational science and engineering refers to the development of computationally efficient approximations for expensive simulations, such as those arising from numerical solution of partial differential equations (PDEs). Surrogate modeling is an enabling methodology for many-query computations in science and engineering, which include iterative methods in optimization and sampling methods in uncertainty quantification. Over the last few years, several approaches to surrogate modeling for PDEs using neural networks have emerged, motivated by successes in using neural networks to approximate nonlinear maps in other areas. In principle, the relative merits of these different approaches can be evaluated by understanding, for each one, the cost required to achieve a given level of accuracy. However, the absence of a complete theory of approximation error for these approaches makes it difficult to assess this cost-accuracy trade-off. In this talk we provide a careful numerical study of this issue, comparing a variety of different neural network architectures for operator approximation across a range of problems arising from PDE models in continuum mechanics.

Maarten V. de Hoop
Rice University
mdehoop@rice.edu

Daniel Huang, Elizabeth Qian
California Institute of Technology
dzhuang@caltech.edu, eqian@caltech.edu

Andrew Stuart
Computing + Mathematical Sciences
California Institute of Technology
astuart@caltech.edu

MS61

Learning with Minimal Supervision in Medical Imaging - from Classifying Chest X-rays to Fast MRI

Machine learning has entered many areas of science, engineering and every day life. Medical imaging is no exception and indeed, many success stories in the enhancement and analysis of medical images have happened in particular in more recent years with the resurrection of neural networks. Most state-of-the-art approaches in this context rely on large sets of supervised training data. These, however, do not exist or are very costly to obtain in most real life medical imaging applications. In this talk we will investigate machine learning approaches for medical imaging with minimal (or no) supervision and which are as such applicable in the small data regime. We will discuss deep graphical models for semi-supervised classification and un- and semi-supervised tomographic image reconstruction for low-dose CT and fast MRI.

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

MS61

Markov Chain Monte Carlo Sampling with Deep Learning

Sampling high-dimensional probability distributions is a common task in computational chemistry, Bayesian inference, etc. Markov Chain Monte Carlo (MCMC) is the method of choice to perform these calculations, but it is often plagued by slow convergence properties. I will discuss how methods from deep learning (DL) can help enhance the performance of MCMC via a feedback loop in which we simultaneously use DL to learn better samplers based e.g. on generative models, and MCMC to obtain the data for the training of these models. I will illustrates these techniques via several examples, including the sampling of reaction paths in metastable systems and the calculation of free energies and Bayes factors.

Eric Vanden Eijnden
Courant Institute
New York University
eve2@cims.nyu.edu

MS62

Benign Overfitting of Constant-Stepsize SGD for Linear Regression

There is an increasing realization that algorithmic inductive biases are central in preventing overfitting; empirically, we often see a benign overfitting phenomenon in overparameterized settings for natural learning algorithms, such as stochastic gradient descent (SGD), where little to no explicit regularization has been employed. In this talk, I will discuss benign overfitting of constant-stepsize SGD in arguably the most basic setting: linear regression in the overparameterized regime. Our main results provide a sharp excess risk bound, stated in terms of the full eigenspectrum of the data covariance matrix, that reveals a bias-variance decomposition characterizing when generalization is possible. This is joint work with Difan Zou, Jingfeng Wu, Vladimir Braverman, Dean P. Foster, and Sham M. Kakade.

Quanquan Gu
UCLA
qgu@cs.ucla.edu

MS62

Parameter-Free Stochastic Optimization of Variationally Coherent Functions

Consider the problem of finding the minimizer of a differentiable function $F : \mathbb{R}^d \to \mathbb{R}$ using access only to noisy gradients. A plethora of algorithms have been proposed to solve this problem. However, the choice of the algorithm and its parameters crucially depends on the (unknown) characteristics of the function F. On the other hand, for convex Lipschitz functions it is possible to design so-called parameter-free optimization algorithms that guarantee almost optimal performance without any hyperparameter to tune. Unfortunately, they do not seem to work on non-convex functions. In an effort to go beyond convex functions, we focus on variationally coherent functions: for these functions, at any point $\boldsymbol{x}$ the vector pointing towards the optimal solution $\boldsymbol{x}^*$ and the negative gradient form an acute angle. We propose a new algorithm based on the Dual Averaging framework with the added twists of using rescaled gradients and time-varying linearithmic regularizers. We can prove almost sure convergence to the global minimizer of variationally coherent functions. Additionally, the very same algorithm with the same hyperparameters, after T iterations guarantees on convex functions that the expected suboptimality gap is bounded by $O(\|\boldsymbol{x}^* - \boldsymbol{x}_0\| T^{-1/2+\epsilon})$ for any $\epsilon > 0$, up to polylog terms. This is the first algorithm to achieve both these properties at the same time.

Francesco Orabona
Boston University
fo@bu.edu

MS62

Stochastic Linear Optimization Never Overfits with Quadratically-Bounded Losses on General Data

In this talk I will show that a diverse collection of linear optimization methods, when run on general data, fail to overfit, despite lacking any explicit constraints or regularization: with high probability, their trajectories stay near the curve of optimal constrained solutions over the population distribution. This analysis is powered by an elementary but flexible proof scheme which can handle many settings, summarized as follows. Firstly, the data can be general: unlike other implicit bias works, it need not satisfy large margin or other structural conditions, and moreover can arrive sequentially IID, sequentially following a Markov chain, as a batch, and lastly it can have heavy tails. Secondly, while the main analysis is for mirror descent, rates are also provided for the Temporal-Difference fixed-point method from reinforcement learning; all prior high probability analyses in these settings required bounded iterates, bounded updates, bounded noise, or some equivalent. Thirdly, the losses are general, and for instance the logistic

and squared losses can be handled simultaneously, unlike other implicit bias works. In all of these settings, not only is low population error guaranteed with high probability, but moreover low sample complexity is guaranteed so long as there exists any low-complexity near-optimal solution, even if the global problem structure and in particular global optima have high complexity.

Matus Telgarsky
University of Illinois
mjt@illinois.edu

MS62

Grand++: Graph Neural Diffusion with a Source Term

In this talk, I will present GRAph Neural Diffusion with a source term (GRAND++) for graph deep learning with a limited number of labeled nodes, i.e., low-labeling rate. GRAND++ is a class of continuous-depth graph deep learning architectures whose theoretical underpinning is the diffusion process on graphs with a source term. The source term guarantees two interesting theoretical properties of GRAND++: (i) the representation of graph nodes, under the dynamics of GRAND++, will not converge to a constant vector over all nodes even as the time goes to infinity, which mitigates the over-smoothing issue of graph neural networks and enables graph learning in very deep architectures. (ii) GRAND++ can provide accurate classification even when the model is trained with a very limited number of labeled training data.

Hedi Xia
University of California, Los Angeles, U.S.
hedixia@math.ucla.edu

Bao Wang
Data Science Research Lab in the DEPARTMENT OF MATHEMATICS a
bwang@math.utah.edu

MS63

Geometric Considerations when Surrogate Modeling for Multiobjective Optimization Problems

Surrogate modeling for multiobjective optimization problems presents unique challenges and opportunities from a data science perspective. Similarly as in a regression problem, the goal of surrogate modeling is to approximate the underlying objective function(s) using a database of real-world observations. However, in this context, the optimization algorithm gets to choose the locations for its data points, instead of being presented with a predetermined dataset. A multiobjective optimization algorithm must balance its goal of locating numerous Pareto optimal solutions, while still placing exploratory data points that improve the predictive capabilities of its surrogate models. Proper placement of these data points requires an understanding of the geometry of the input space. In this talk, we will discuss geometric issues in the context of the response surface methodology (RSM) framework for solving multiobjective optimization problems.

Tyler H. Chang
Argonne National Laboratory
tchang@anl.gov

MS63

Data-Driven Geometric Scale Detection via Delaunay Interpolation

In data analysis and machine learning, knowing how much data is sufficient for the task at hand is an important consideration. Too little data can lead to underfitting of models, while too much data suggests inefficient use of resources. There are few systematic approaches to determining the proper amount of data to use, with most approaches relying on heuristics or post-hoc justification based on performance. In this talk, we describe a new systematic method for determining how much data is necessary for the specific task of predicting a scalar function over an input space of medium dimension (typically between 2 and 10). The method relies on using the data to construct successively refined Delaunay interpolants, allowing us to precisely characterize the scales of geometric features present in the scalar function and determine when the data density is sufficiently high to properly resolve such features. We will discuss both the theoretical foundations of the method as well as numerical experiments demonstrating its use and validity.

Eugene Kur
Lawrence Livermore National Laboratory
kur1@llnl.gov

Andrew Gillette
Lawrence Livermore National Laboratory
Center for Applied Scientific Computing
gillette7@llnl.gov

MS63

Scalable Topological Data Analysis for Evaluating Data-Driven Models

With the rapid adoption of machine learning techniques for large-scale applications in science and engineering comes the convergence of two grand challenges in visualization. First, the utilization of black box models (e.g., deep neural networks) calls for advanced techniques in exploring and interpreting model behaviors. Second, the rapid growth in computing has produced enormous datasets that require techniques that can handle millions or more samples. Although some solutions to these interpretability challenges have been proposed, they typically do not scale beyond thousands of samples, nor do they provide the high-level intuition scientists are looking for. Here, we present the first scalable solution to explore and analyze high-dimensional functions often encountered in the scientific data analysis pipeline. By combining a new streaming neighborhood graph construction, the corresponding topology computation, and a novel data aggregation scheme, namely topology aware datacubes, we enable interactive exploration of both the topological and the geometric aspect of high-dimensional data. Following a use cases from high-energy-density (HED) physics we demonstrate how these capabilities have led to crucial new insights in real world data-driven application.

Shusen Liu
Lawrence Livermore National Laboratory

liu42@llnl.gov

MS63

Geometric Priors for Scientific Generative Models in Inertial Confinement Fusion

Deep generative models have advanced our ability to synthesize and create highly realistic samples in domains ranging from images to speech and text. These advances are especially useful in scientific applications to characterize simulation data, which can then lead to effective coupling with experiments, to improved prediction capabilities, and eventually to an improved understanding of the underlying physics. Unfortunately, in the scientific setting generative models can fail in unexpected ways, often leading to samples that are implausible as defined by known physical laws or constraints. This is due to a mismatch between the true (unknown) generative process, and the one estimated using the generative model. One mitigation strategy is to engineer more complex, structured latent spaces that not only possess improved expressivity but also impact any downstream task (transfer learning, search and retrieval, design optimization etc.). In this work, we study how enforcing geometric priors in the latent space impacts generative models trained using ICF simulation data. Specifically, we explore a new class of geometry-driven generative models that go beyond the conventional Euclidean assumption. We utilize analytical distributions defined on non-Euclidean spaces, such as Poincare and hypersphere, in an attempt to improve the generation of samples thar are more interpretable while lying on the physics manifold.

Ankita Shukla
Arizona State University
ashukl20@asu.edu

Rushil Anirudh
Lawrence Livermore National Laboratory
anirudh1@llnl.gov

Jayaraman Thiagarajan
Lawrence Livermore National Labs
jayaramanthi1@llnl.gov

Eugene Kur
Lawrence Livermore National Laboratory
kur1@llnl.gov

Peer-Timo Bremer
lawrence livermore national laboratory
bremer5@llnl.gov

Brian Spears, Tammy Ma
Lawrence Livermore National Laboratory
spears9@llnl.gov, ma8@llnl.gov

Pavan Turaga
Arizona State University
pturaga@asu.edu

MS64

Finite-Time Analysis of Natural Actor-Critic with Neural Network Approximation

Natural actor-critic (NAC), equipped with the representation power of neural networks, have demonstrated impressive empirical success in solving Markov decision problems with large state spaces. In this talk, we present a finite-time analysis of NAC with neural network approximation, and identify the roles of regularization and optimization techniques (e.g., gradient clipping and averaging) to achieve sharp bounds on the sample complexity, iteration complexity and network width. In particular, we prove that (i) the uniform approximation power of the actor neural network is important for global optimality in policy optimization due to distributional shift, (ii) entropy regularization, averaging and gradient clipping ensure sufficient exploration to guarantee optimality under minimal assumptions on concentrability coefficients, and (iii) regularization leads to sharp sample complexity and network width bounds in the regularized MDPs, yielding a favorable bias-variance tradeoff in policy optimization.

Semih Cayci
The Ohio State University
Department of Electrical and Computer Engineering
scayci@illinois.edu

Niao He
ETH Zurich
niao.he@inf.ethz.ch

Rayadurgam Srikant
UIUC
rsrikant@illinois.edu

MS64

Target Network and Truncation Overcome The Deadly Triad in Q-Learning

Q-learning with function approximation is one of the most empirically successful while theoretically mysterious reinforcement learning (RL) algorithms, and was identified in Sutton (1999) as one of the most important theoretical open problems in the RL community. Even in the basic linear function approximation setting, there are well-known divergent examples. In this work, we propose a stable design for Q-learning with linear function approximation using target network and truncation, and establish its finite-sample guarantees. Our result implies an $O(\epsilon^{-2})$ sample complexity up to a function approximation error. This is the first variant of Q-learning with linear function approximation that is provably stable without requiring strong assumptions or modifying the problem parameters, and achieves the optimal sample complexity.

Zaiwei Chen
Georgia Institute of Technology
zchen458@gatech.edu

John-Paul Clarke
University of Texas at Austin
johnpaul@utexas.edu

Siva Theja Maguluri
Georgia Institute of Technology
siva.theja@gatech.edu

MS64

Policy Gradient Methods in Multi-Agent Reinforcement Learning: Do They Even Converge?

In this talk I will present work on understanding the behavior and convergence of policy-gradient algorithms in multi-agent reinforcement learning. In particular, I will

discuss their performance in simple Linear Quadratic Dynamic Games (the extension of the LQR problem to multiple agents) where I will discuss the optimization landscape of these problems and some convergence (and non-convergence) guarantees for this class of algorithms.

Eric Mazumdar
California Institute of Technology
mazumdar@caltech.edu

MS64

Overcoming the Long Horizon Barrier for Sample-Efficient Reinforcement Learning with Latent Low-Rank Structure

The practicality of reinforcement learning algorithms has been limited due to poor scaling with respect to the problem size, as the sample complexity of learning an epsilon-optimal policy scales with $|S|times|A|$ over worst case instances of an MDP with state space S, action space A, and horizon H A key question is when can we design probably efficient RL algorithms that exploit nice structure? We consider a class of MDPs that exhibit low rank structure, where the latent features are unknown. We argue that a natural combination of value iteration and low-rank matrix estimation results in an estimation error that grows doubly exponentially in the horizon length. We then provide a new algorithm along with statistical guarantees that efficiently exploits low rank structure given access to a generative model, achieving a sample complexity that scales linearly in $|S| + |A|$ and polynomially in the horizon and the rank. In contrast to literature on linear and low-rank MDPs, we do not require a known feature mapping, our algorithm is computationally simple, and our results hold for long time horizons. Our results provide insights on the minimal low-rank structural assumptions required on the MDP with respect to the transition kernel versus the optimal action-value function.

Christina L. Yu
Cornell University
Cornell University
cleeyu@cornell.edu

Tyler Sam
Cornell University
tjs355@cornell.edu

Yudong Chen
UW Madison
yudong.chen@wisc.edu

MS65

Learning Continuous Networks for Continuous Physics

Deep learning is playing a growing role in many areas of science and engineering for modeling dynamical systems that evolve continuously over time, including fluid flows, and climate data. In this talk, I am going to discuss continuous neural networks that can approximate a system's underlying continuous dynamics and enable both better interpolation and better extrapolation in multiple ways. However, depending on the training procedure, these models can also easily fail to capture relevant dynamics, rendering them of limited utility for many scientific prediction tasks. Hence, I will discuss a convergence test for continuous neural networks, based on numerical analysis theory, to validate models for science and engineering applications.

N. Benjamin Erichson
University of Pittsburgh
erichson@pitt.edu

MS65

Model Discrepancy Sensitivities to Enable Decision-Making

Scientific machine learning has opened new avenues of research to enable efficient outer loop analysis through learned models of physical systems. However, these learned models are imperfect representations of complex physical processes. The discrepancy between such models and the underlying truth may be amplified by outer loop analysis such as optimization. We present a novel approach to compute the sensitivity of optimization problems with respect to model discrepancy and use this information to improve the solution obtained using learned models. By posing a Bayesian inverse problem to calibrate the discrepancy, we compute a posterior discrepancy distribution and then propagate it through post-optimality sensitivities to compute a posterior distribution on the optimal solution.

Joseph Hart
Sandia National Laboratories
joshart@sandia.gov

MS65

Advances and Challenges in Solving HJB Equations Arising in Optimal Control

We present a neural network approach for approximately solving high-dimensional stochastic as well as deterministic control problems. Our network design and the training problem leverage insights from optimal control theory. We approximate the value function of the control problem using a neural network and use the Pontryagin maximum principle and Dynamic Programming principle to express the optimal control (and therefore the sampling) in terms of the value function. Our training loss consists of a weighted sum of the objective functional of the control problem and penalty terms that enforce the Hamilton Jacobi Bellman equations along the sampled trajectories. As a result, we can obtain the value function in the regions of the state space traveled by optimal trajectories to avoid the Curse of Dimensionality. Importantly, training is unsupervised in that it does not require solutions of the control problem. Our approach for stochastic control problem reduces to the method of characteristics as the system dynamics become deterministic. In our numerical experiments, we compare our method to existing solvers for a more general class of semi-linear PDEs. Using a two-dimensional toy problem, we demonstrate the importance of the PMP to inform the sampling. For a 100-dimensional benchmark problem, we demonstrate that approach improves accuracy and time-to-solution. Finally we consider a PDE based dynamical system to demonstrate the scalability of our approach.

Deepanshu Verma, Xingjian Li
Emory University
dverma4@emory.edu, xingjian.li@emory.edu

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

Bart van Bloeme Waanders
Sandia National Laboratories
bartv@sandia.gov

Nickolas Winovich
Sandia National Laboratory, U.S.
nwinovi@sandia.gov

MS65

Rare Event Sampling for the Planetary Sciences

Rare events can be highly impactful. Yet, estimating the probability p of a rare event by direct numerical simulation requires a very large sample size ($> 100/p$). To reduce these sample size requirements, a 'splitting and killing' scheme can be used instead to estimate rare event probabilities. In this approach, certain trajectories are 'split' to promote progress toward a rare event while other trajectories are randomly 'killed' to control the computational cost. We have proved splitting and killing can estimate rare event probabilities at exponentially lower cost compared to direct numerical simulation. Additionally, we recently applied splitting and killing to estimate the rare probability of Mercury's orbit destabilizing after a close encounter with Venus.

Robert Webber
California Institute of Technology
rwebber@caltech.edu

MS66

Wasserstein Patch Prior for Inverse Problems

We introduce the Wasserstein Patch Prior (WPP) for image superresolution, which yields a prior for unsupervised training of neural networks based on the comparison of empirical patch distributions with respect to the Wasserstein-2 distance. Instead of requiring access to a large database of registered high- and low-resolution images, we only assume to know a large database of low resolution images, the forward operator and one high-resolution reference image, which has a similar patch distribution as the unknown high-resolution images, fulfilled for e.g. textures or material data. In a first step, we explain the idea of WPP and show that using the WPP for training non-invertible CNNs performs better than other methods with similar assumptions. In particular, it provides a stable behavior under an inaccurate forward operator as given e.g. in real world applications. In a second step, we present a possibility of unsupervised training of Normalizing Flows using the so-called backward Kullback-Leibler divergence in order to approximate the superresolution space of a given low-resolution image. Here we use the WPP as a prior distribution.

Fabian Altekrüger
HU Berlin
fabian.altekrueger@hu-berlin.de

MS66

Normalizing Flows Based Importance Sampling for Estimating Rare Event Probabilities

Estimating the probabilities of rare events is a challenging but critical task in many fields of engineering and science. In this work, we propose the use of normalizing flows (NFs) to construct importance sampling distributions that can be used to estimate the probabilities of rare events. NFs are generative modeling techniques amenable to exact, but efficient, density evaluation. We evaluate our proposed approach on a number of challenging benchmark reliability estimation problems and compare it against two techniques widely adopted for similar tasks: subset simulation and cross-entropy method. The results show that the proposed approach can be used to estimate rare-event probability in cases that have extremely low failure probability ($< 10^{-7}$), high-dimensionality, and multiple failure modes. Moreover, we show empirically that the estimator has desirable statistical properties. The proposed approach is also applied to the reliability estimation of two physics-based problems.

Agnimitra Dasgupta, Erik Johnson
University of Southern California
adasgupt@usc.edu, johnsone@usc.edu

MS66

Generative Modelling for Digital Pathology

Driven by requirements to increase efficiency and the desire for mobile working, digitization is becoming increasingly important in pathology. Optical scanners gradually replace the traditional microscope to examine histologically stained tissue sections. However, only switching to optical scanners does not justify the high costs of the digitization process. The missing link is deep learning models that significantly speed up the diagnosis and reporting process and provide another layer of security by assisting the pathologist. A couple of problems must be overcome to apply deep learning approaches in digital pathology. First, the number of annotated samples is limited for many diseases because they are rare, or the annotation process is time-consuming. Second, the models must be robust to variations between individual laboratories, e.g., different scanners or staining solutions. In addition, the scanned whole slide images (WSI) are taken at different magnification levels to mimic the behavior of the microscope. The results are gigapixel images at the highest magnification level. Our talk investigates how generative models can be applied to digital pathology and tackle the mentioned challenges. The model selection includes auto-encoders, generative adversarial networks, and normalizing flows. We test these on true pathological WSI in applications such as out-of-distribution detection or robustification of segmentation models.

Maximilian Schmidt
Center for Industrial Mathematics, University of Bremen, Ge
maximilian.schmidt@uni-bremen.de

MS67

Piecewise-Affine Near-Optimal Feedback Control via Deep Reinforcement Learning

We study deep reinforcement learning and show that it is not so mysterious but rather closely related to classical ideas from optimal control theory and spline approximation. Specifically, we connect two recent lines of research that have demonstrated that (i) reinforcement learning is equivalent to approximate dynamic programming and (ii) a large class of deep (neural) networks are equivalent to continuous piecewise-affine spline mappings. We show that training a reinforcement learning agent with an internal policy parametrized by a deep network naturally leads to the construction of a near-optimal continuous piecewise-

affine controller.

Randall Balestriero, Cannon Lewis, Saad Saleh
Rice University
randallbalestriero@gmail.com, wcl2@rice.edu,
ss212@rice.edu

Richard G. Baraniuk
Rice University
Electrical and Computer Engineering Department
richb@rice.edu

MS67

This Looks Like That There: Interpretable Neural Networks for Image Tasks When Location Matters

We develop and demonstrate a new interpretable deep learning model specifically designed for image analysis in earth system science applications. The neural network is designed to be inherently interpretable, rather than explained via *post hoc* methods. This is achieved by training the network to identify parts of training images that act as prototypes for correctly classifying unseen images. The new network architecture extends the interpretable prototype architecture of a previous study in computer science to incorporate absolute location. This is useful for earth system science where images are typically the result of physics-based processes, and the information is often geo-located. Although the network is constrained to only learn via similarities to a small number of learned prototypes, it can be trained to exhibit only a minimal reduction in accuracy compared to non-interpretable architectures. We apply the new model to two earth science use cases: a synthetic data set that loosely represents atmospheric high- and low-pressure systems, and atmospheric reanalysis fields to identify the state of tropical convective activity associated with the Madden-Julian oscillation. In both cases, we demonstrate that considering absolute location greatly improves testing accuracies. Furthermore, the network architecture identifies specific historical dates that capture multivariate, prototypical behaviour of tropical climate variability.

Elizabeth Barnes
Colorado Stete University
eabarnes@colostate.edu

MS67

Explaining Transfer Learning in Closure Modeling of Turbulent Thermal Convection

In this work, we develop a data-driven subgrid-scale (SGS) model for large eddy simulation (LES) of 2D turbulent thermal convection using a fully convolutional neural network (CNN). In the big-data regime or with data augmentation, the LES-CNN reproduces correct statistics in terms of kinetic energy and entropy (scalar energy) spectra and Nusselt number. Although LES-CNN works when the training and online testing systems have the same control parameters, e.g., Rayleigh number and Prandtl number, the network suffers from inaccurate inferences when it is applied to a different flow with different control parameters. We propose to use transfer learning to address the generalization/extrapolation issue. Transfer learning requires a small fraction of the big data obtained from the target system and a trained network with big data obtained from the base system. To understand the physics of transfer learning, we explain what is learned in transfer learning to a different turbulent flow, which is based around changes in the convolution kernels of the base network after re-training to the target network and the kernels physical interpretation. We show a general a priori framework to guide transfer learning of similar systems.

Yifei Guan, Ashesh K. Chattopadhyay
Rice University
yg62@rice.edu, akc6@rice.edu

Pedram Hassanzadeh
Mechanical Engineering and Earth Science
Rice University
pedram@rice.edu

MS68

Biclustering with Feature Selection

Convex clustering has recently gained popularity due to its attractive theoretical and computational properties. Though it confers many advantages over traditional clustering methods, its merits become limited in the face of high-dimensional data due to numerical instability. To address these issues, biconvex clustering has been proposed recently to simultaneously optimize the feature weights as well as the centroids. In the biclustering problem, we seek to simultaneously group observations and features. For high-throughput bioinformatics data, we often also wish to perform feature selection. In this talk, we propose a novel biconvex algorithm to tackle the biclustering problem while simultaneously performing feature selection. Called 'biconvex biclustering', the proposed method performs by selecting proper subsets in both the groups as well as the features throughout the clustering task. We demonstrate our methods utility for exploring single cell RNA sequencing data.

Eric Chi
Rice University
echi@rice.edu

MS68

Limit Theorems for Sum-of-Norms Clustering

I will present two results concerning sum-of-norms clustering in the limit as N, the number of datapoints, becomes large. The first concerns unweighted sum-of-norms clustering, and gives a local-global characterization of the clusters that, at least for simple datapoint distributions, can be computed explicitly in the large-N limit. The second concerns weighted sum-of-norms clustering when the range of the weight function goes to zero as N goes to infinity. It yields a PDE problem in the limit that again can be computed explicitly in simple situations. I will also discuss applications of these results to the stochastic ball model.

Alexander Dunlap
New York University
alexander.dunlap@cims.nyu.edu

Jean-Christophe Mourrat
ENS Lyon
CNRS
jean-christophe.mourrat@ens-lyon.fr

MS68

A New Family of Features That Strengthens Re-

covery Guarantees of Sum-of-Norms Clustering

Sum-of-norms (SON) clustering is a convex formulation of the clustering problem that comes with several desirable guarantees. One guarantee is that it produces a hierarchical family of clusterings (Chiquet et al., 2017). A second is that the method is guaranteed to recover the central points sampled from a mixture of Gaussians provided that the standard deviations are bounded in terms of the distances between the means (Jiang et al., 2020). However, SON clustering also has some limitations including its inability to recover clusters if the convex hulls of the supports of the clusters in the underlying distribution are not disjoint (Nguyen and Mamitsuka, 2021). Even when points are sampled from a distribution supported on two disjoint disks in the plane, the clusters are not recovered if the two disks are closer than a certain fraction of the radius (Dunlap and Mourrat, 2021). We propose a new family of features, called "leapfrog distances," that improve the recovery guarantees of SON clustering. When these new features are appended to the original data points, clusters are recovered from points sampled from arbitrary distributions provided clusters have disjoint supports. Points sampled from mixtures of Gaussians are recovered for a less restrictive choice of standard deviations.

Tao Jiang, Samuel Tan
Cornell University
tj293@cornell.edu, sst76@cornell.edu

Stephen A. Vavasis
University of Waterloo
Dept of Combinatorics and Optimization
vavasis@uwaterloo.ca

MS68

Efficient Algorithms and Acceleration Techniques for Solving Convex Clustering Problems

We develop a semismooth Newton based augmented Lagrangian (SSNAL) algorithm for solving large-scale convex clustering problems. Extensive numerical experiments on both simulated and real data demonstrate that our algorithm is highly efficient and robust for solving large-scale problems. We also introduce an adaptive sieving technique to reduce the dimension of the problems we have to solve. As a result, we can accelerate our SSNAL algorithm by more than 7 times and the ADMM algorithm by more than 14 times

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

Yancheng Yuan
Hong Kong PolyU
yancheng.yuan@polyu.edu.hk

Defeng Sun
The Hong Kong Polytechnic University, Hong Kong
defeng.sun@polyu.edu.hk

MS69

On Adversarial Training and Perimeter Minimization Problems

Adversarial training is a framework widely used by machine learning practitioners to enforce robustness of learning models. However, despite the development of several computational strategies for adversarial training and some theoretical development in the broader distributionally robust optimization literature, there are several theoretical questions about adversarial training that remain relatively unexplored. In this talk, I will present a series of connections between adversarial training and several problems in the calculus of variations, and geometric measure theory. These connections reveal a rich geometric structure of adversarial problems and conceptually all aim at answering the question: what is the regularization effect induced by adversarial training? I will highlight how these connections provide novel theoretical results on robust training of learning models as well as provide a directly interpretable statistical motivation for a family of regularized risk minimization problems involving perimeter/total variation. This talk is based on joint works with Ryan Murray, Camilo A. Garcia Trillos, Leon Bungert, and Jakwang Kim.

Nicolas Garcia Trillos
Department of Statistics
University of Wisconsin-Madison
garciatrillo@wisc.edu

MS69

Deep Networks and Energy Minimization

Energy is one of the fundamental concepts in computational representation. The introduction of weak form has been a fundamental tool for the development of variational techniques. In this talk we derive a class of networks that are a kin to energy minimization techniques. We show how such networks has predictable properties and how they can be used to solve problems in fields such as molecular dynamics and mesh design.

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

MS69

Approximation Theory of Deep Learning for Sequence Modelling

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

Qianxiao Li
National University of Singapore
Singapore
qianxiao@nus.edu.sg

MS69

Continuous Limit of Bayesian Bandits

For multi-armed bandits (MAB), many algorithms are designed for solving frequentist regrets, but fewer algorithms

are available for solving Bayesian regrets. One of the main challenges of the MAB problem in the bayesian setting is that the computation of the optimal policy appears intractable even for simple settings. The computational cost increases exponentially as the horizon increases. We derive a continuous control problem for the bayesian bandits as the horizon goes to infinity, and a Hamilton-Jacobi-Bellman (HJB) equation is used to solve the limiting control problem. Based on the HJB equation, we propose a new algorithm for solving Bayesian bandits with large horizons. Numerical results show that the proposed algorithm can beat the UCB and Thomsom Sampling algorithms with the same computational cost when the horizon is large.

Yuhua Zhu
Stanford University
yuhuazhu@stanford.edu

MS70

Addressing Computational Bottlenecks in Higher-Order Graph Matching with Tensor Kronecker Product Structure

We present and apply a theory of tensor Kronecker products to tensor based graph alignment algorithms to reduce their runtime complexity from quadratic to linear with no appreciable loss of quality. The key to our methods is a decoupling theorem on the dominant tensor Z-eigenvector of a tensor Kronecker product. This helps these tensor-based graph alignment methods scale in terms of memory and computational demands that are quadratic (or worse) in terms of problem size.

David F. Gleich
Purdue University
dgleich@purdue.edu

MS70

Simultaneously-Updated Support Tensor Train Machine

An increasing amount of data collected are high-dimensional and efficient learning algorithms must utilize the tensorial structure as much as possible. The ever-present curse of dimensionality for high dimensional data and the loss of structure when vectorizing the data motivates the use of tailored low-rank tensor methods. We propose an algorithm of computing a Canonical Polyadic (CP) decomposition by avoiding the NP-hard issue of finding the best CP rank by computing first a Tensor Train (TT) decomposition and calling it TT-CP factorization. Along with it, we define a nonlinear classification machine learning model. We build a full Gradient Descent Primal (GDP) optimization problem which takes initialized variables from the partial GDP model optimized via Support Tensor Machines (STM). In turn, the full GDP enhances a potential suboptimal CP decomposition computed in the first step. This leads to better classification accuracy and a reliable deterministic algorithm for computing the nonlinear boundary, each step of which admits a reasonable explanation. Hence, the full GDP can thus be seen as both a tensor decomposition method tailored to the classification problem, and a classification method that exploits the low-rank model of the data. With numerical examples, we show that this approach benchmarked than other state-of-the-art techniques.

Peter Benner
Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
benner@mpi-magdeburg.mpg.de

Sergey Dolgov
University of Bath
Department of Mathematical Sciences
s.dolgov@bath.ac.uk

Kirandeep Kour
Max Planck Institute for Dynamics of Complex Technical Systems, Magdbeurg
kour@mpi-magdeburg.mpg.de

Martin Stoll
Department of Mathematics,
Technische Universitaet Chemnitz
martin.stoll@math.tu-chemnitz.de

MS70

Group-invariant Tensor Train Networks for Supervised Learning

Symmetry equivariance has recently proven to be a powerful inductive bias in various machine learning models. Another novel method in the machine learning community is the use of tensor networks as either predictive or generative models. This paper provides a way to construct tensors of any order that are equivariant under local symmetry operations for arbitrary discrete symmetry groups and how these can be combined to form a symmetry-equivariant tensor network. Our proposed class of networks is demonstrated on a number of benchmark problems where it outperform models that do not take symmetry into account and shows to be competitive with state-of-the-art models.

Nick Vannieuwenhoven
Departement Computerwetenschappen
KU Leuven
nick.vannieuwenhoven@cs.kuleuven.be

Brent Sprangers
KU Leuven
brent.sprangers@kuleuven.be

MS71

On the Regularity of PDEs in Barron Spaces

Barron functions arise naturally as infinite-width limit of two-layer neural networks. These functions can be well approximated by two-layer networks without curse of dimensionality. Motivated by the recent advancements of neural network methods for solving high dimensional PDEs, we will discuss in this talk some recent theoretical results on the regularity of PDEs in Barron spaces. More precisely, we show that for a class of elliptic equations and eigenvalue problems, the solutions are Barron provided that the coefficients of the PDEs are Barron.

Yulong Lu
Department of Mathematics and Statistics
University of Massachusetts Amherst, U.S
lu@math.umass.edu

MS71

Discretization-Invariant Operator Learning

Discretization-Invariant Learning aims at learning in the infinite-dimensional function spaces with the capacity to

process heterogeneous discrete representations of functions as inputs and/or outputs of a learning model. In this talk, I will introduce a novel deep learning framework based on integral autoencoders (IAE-Net) for discretization-invariant learning. This will be done by introducing the basic building block of IAE-Net, as well as how these building blocks are composed to form the IAE-Net structure. Together with randomized data augmentation that generates training data with heterogeneous structures to facilitate discretization invariant learning, we will show how IAE-Net can be applied to a wide range of applications in predictive data science, solving forward and inverse problems in scientific computing, and signal/image processing, achieving state-of-the-art performance and creating a wide range of new applications.

Yong Zheng Ong
National University of Singapore
e0011814@u.nus.edu

Haizhao Yang
Purdue University
haizhaoyang@yahoo.com

MS71

Partial Differential Equations in Spaces of Neural Network Functions

We consider the solution of PDEs from the regularity perspective (if the data are 'nice', are the solutions 'nice'?), but with data in spaces of neural network functions. We provide both positive and negative results in special cases. The presentation is based on joint work with Weinan E.

Stephan J. Wojtowytsch
Texas A&M University
swoj@tamu.edu

MS71

Towards Neural Network Approaches for High Dimensional Problems

Neural networks are effective tools for solving high dimensional problems. In this talk, I will summarize the popular methods to solve high dimensional problems with neural networks. Then I will briefly introduce two of my works based on the DeepBSDE method. In the first work, we solve the eigenvalue problem by transforming it into a fixed-point formulation, which is a diffusion Monte Carlo like approach. In the second work, we leverage the actor-critic framework from reinforcement learning to solve the static HamiltonJacobiBellman equations. We propose a variance reduced temporal difference method for the critic and apply an adaptive step size algorithm for the actor to improve accuracy.

Mo Zhou
Duke University
mo.zhou366@duke.edu

MS72

A Theoretically Justified Neural-Network Approach for Sparsity Promotion in Linear Inverse Problems

Sparsity promotion is a popular regularization technique for inverse problems, reflecting the prior knowledge that the exact solution is expected to have few non-vanishing components, for example, in a suitable wavelet decomposition. In this talk, I will present a novel convolutional neural network, called ΨDONet, designed for sparsity-promoting regularization for linear inverse problems. Such a network is able to replicate the application of the Iterative Soft Thresholding Algorithm (ISTA), a classical reconstruction algorithm for sparsity promoting regularization, and to actually outperform it, by learning a suitable pseudodifferential correction. By a combination of techniques and tools from regularization theory of inverse problems, multiresolution wavelet analysis, and microlocal analysis, we are able to theoretically deduce the architecture of the network and to prove some convergence results. Our case study is limited-angle computed tomography: we test two different implementations of our network on simulated data from limited-angle geometry, achieving promising results.

Tatiana A. Bubba
University of Bath
Department of Mathematical Sciences
tab73@bath.ac.uk

Mathilde Galinier
University of Modena and Reggio Emilia
mathildeemmanuelle.galinier@unimore.it

Matti Lassas
University of Helsinki
matti.lassas@helsinki.fi

Marco Prato
Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Università di Modena e Reggio Emilia
marco.prato@unimore.it

Luca Ratti
University of Genoa
luca.ratti@unige.it

Samuli Siltanen
University of Helsinki
Finland
samuli.siltanen@helsinki.fi

MS72

Multi-Probabilistic Guarantees for Detecting and Correcting Non-Conformal Data Samples

Trained neural networks are often expected to generalize beyond the training data. For deep neural networks it has been observed many times that this is the case, however, the reasons are not fully understood. Furthermore, a small modification to the input of a deep neural network may drastically change the output. A possible explanation of this phenomenon is offered by the calibration hypothesis, which states that the probabilistic predictions of typical ML models are mis-calibrated, causing the classifier to be confident in often completely erroneous predictions. Inductive Venn-Abers predictors (IVAP) have the theoretical guarantee of being perfectly calibrated and they can be used to improve the predictions of binary classification models. The key component of IVAP is a non-conformity measure, which assesses how much a given test sample differs from a given set of training samples. Therefore, IVAP can also be used to detect non-conformal data samples. In our recent work, we have presented multIVAP as an extension of IVAP to multi-class classification, with similar theoretical guarantees to the IVAP. In this talk, we will

explain with some practical use cases, how multIVAP can be used to detect erroneous samples on the one hand, and data samples that are outliers of the training data set on the other hand (in which case user interaction could be required). Therefore, multIVAP has applications in both improving the data annotation process as well as in online learning.

Jonathan Peck
Ghent University
jonathan.peck@ugent.be

Bart Goossens
Ghent University-imec
Research group on Image Processing and Interpretation
bart.goossens@ugent.be

Yvan Saeys
Ghent University
yvan.saeys@ugent.be

MS72

Image Inpainting Using from Sparsity Inducing CNNs

Image inpainting aims at recovering missing blocks of data in images or videos. To understand the fundamental challenges of the inpainting problem and derive theoretical performance guarantees, I assume a simplified image model consisting of distributions supported on curvilinear singularities and consider the problem of recovering a missing edge fragment. Using a sparse representation approach, I will show that the theoretical performance of image inpainting critically depends on the microlocal properties of the image representation system, namely, exact image recovery is achieved if the size of the missing edge fragment is smaller than the size of the structure elements of the representation system. As a result, an image inpainting algorithm that exploits the microlocal properties of shearlets significantly outperforms a similar approach based on more traditional multiscale methods. Based on this theoretical observation, I will next propose an improved method for blind image inpainting that combines Convolutional Neural Networks and sparse representations. This modified neural network approach provides state-of-the-art performance on classical benchmark image datasets.

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

MS72

Image Analysis for Satellite Imagery

The co-registration of multi-sensor imagery is a critical step in remote sensing applications. Yet, in the absence of a precise shared coordinate system through georeferencing this remains a challenging problem. In general, area-based nonconvex optimization requires initialization close to the optimum nor can one depend on the presence and extraction of scene specific invariants for feature matching. In this setting, we present a broad exploratory analysis of Fourier-Accelerated template matching for coarse initial alignment. Light Detection and Ranging (LiDAR), Optical, and Hyperspectral (HSI) data are collected at varying levels of detail, have non-linear radiometric differences between their intensities, and are corrupted by different sources of noise. First, We examine the robustness of Normalized Gradient Corss Correlation (NGCC), Gradient Phase Correlation (GPC), and Gradient Fields (GF) to these modalities under affine distortions. Second, we provide a distributed GPU accelerated framework to solve their multi-scale alignment.

William L. Ruys, George Biros
University of Texas at Austin
will@oden.utexas.edu, biros@oden.utexas.edu

MS73

A Spectral Exterior Calculus for Graphs

We present a framework for an exterior calculus on graphs based on the Spectral Exterior Calculus (SEC). The SEC is a global approach to the exterior calculus on manifolds based on representing differential forms globally in a frame (overcomplete basis). The overcomplete basis used in the SEC is built entirely from the eigenvalues and eigenfunctions of the Laplace-Beltrami operator on the manifold. This choice allows various formulas of the classical exterior calculus to be computed analytically. Moreover, the SEC approach was motivated by manifold learning theory which shows that the Laplace-Beltrami operator can be estimated with a graph Laplacian using data sampled from the manifold. By applying the formulas from the SEC to the eigenvalues and eigenvectors of a graph Laplacian, we have a natural notion of an exterior calculus on graphs. In this talk we explore the connection between this spectral exterior calculus and various graph theoretic properties.

Tyrus Berry
George Mason University
tyrus.berry@gmail.com

MS73

Multiscale Basis Dictionaries on Simplicial Complexes

Our previous multiscale graph basis dictionaries (e.g., Generalized Haar-Walsh Transform [GHWT], Hierarchical Graph Laplacian Eigen Transform [HGLET], Natural Graph Wavelet Packets [NGWPs], and their relatives) were for analyzing data recorded on nodes of a given graph. In this work, we propose their generalization for analyzing data recorded on edges or on triangles of a simplicial complex (e.g., a triangle mesh of a manifold). The key idea is to use the so-called Hodge Laplacians and their variants for hierarchical partitioning of edges or triangles, and then build localized basis functions on those subsets. We plan to demonstrate their usefulness for data approximation on a triangular mesh and analysis and clustering of directed graphs.

Naoki Saito
Department of Mathematics
University of California, Davis
saito@math.ucdavis.edu

Stefan Schonsheck
UC Davis
scschonsheck@ucdavis.edu

Eugene Shvarts
UC Davis
Applied Mathematics

eshvarts@math.ucdavis.edu

MS73

Refined Clustering via T-Sne Vectorfields

Modern methods in dimensionality reduction are dominated by nonlinear attraction-repulsion force-based methods (this includes t-SNE, UMAP, ForceAtlas2, LargeVis, and many more). The purpose of this paper is to demonstrate that all such methods, by design, come with an additional feature that is being automatically computed along the way, namely the vector field associated with these forces. We show how this vector field gives additional high-quality information and propose a general refinement strategy based on ideas from Morse theory. The efficiency of these ideas is illustrated specifically using t-SNE on synthetic and real-life data sets.

Yulan Zhang
Yale
yulan.zhang@yale.edu

Anna C. Gilbert
Yale University
anna.gilbert@yale.edu

Stefan Steinerberger
University of Washington
steinerb@uw.edu

MS73

Equivariant Machine Learning and Graph Neural Networks

There has been enormous progress in the last few years in designing neural networks that respect the fundamental symmetries and coordinate freedoms of physical law. Some of these frameworks make use of irreducible representations, some make use of high-order tensor objects, and some apply symmetry-enforcing constraints. Different physical laws obey different combinations of fundamental symmetries, but a large fraction (possibly all) of classical physics is equivariant to translation, rotation, reflection (parity), boost (relativity), and permutations. Here we show that it is simple to parameterize universally approximating polynomial functions that are equivariant under these symmetries, or under the Euclidean, Lorentz, and Poincar groups, at any dimensionality d. The key observation is that nonlinear O(d)-equivariant (and related-group-equivariant) functions can be universally expressed in terms of a lightweight collection of scalars – scalar products and scalar contractions of the scalar, vector, and tensor inputs. We complement our theory with numerical examples that show that the scalar-based method is simple, efficient, and scalable.

Soledad Villar
John Hopkins University
svillar3@jhu.edu

MS74

Metric Hypertransformers are Universal Adapted Maps

We introduce a universal class of geometric deep learning models, called metric hypertransformers (MHTs), capable of approximating any adapted map $F : \mathcal{X}^{\mathbb{Z}} \to \mathcal{Y}^{\mathbb{Z}}$ with approximable complexity, where $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y}$ is any suitable metric space, and $\mathcal{X}^{\mathbb{Z}}$ (resp. $\mathcal{Y}^{\mathbb{Z}}$) capture all discrete-time paths on $\mathcal{X}$ (resp. $\mathcal{Y}$). Suitable spaces $\mathcal{Y}$ include various adapted Wasserstein spaces, all Frchet spaces admitting a Schauder basis, and a variety of Riemannian manifolds arising from information geometry. Even in the "static case", where $f : \mathcal{X} \to \mathcal{Y}$ is a Hlder map, our results provide the first (quantitative) universal approximation theorem compatible with any such $\mathcal{X}$ and $\mathcal{Y}$. Our universal approximation theorems are quantitative, and they depend on the regularity of F, the choice of activation function, the metric entropy and diameter of $\mathcal{X}$ and on the regularity of the compact set of paths whereon the approximation is performed. Our guiding examples originate from stochastic analysis. Notably, the MHT models introduced here are able to approximate a broad range of stochastic processes' kernels, including solutions to SDEs and many processes with an arbitrarily long memory such as certain Volterra processes.

Anastasis Kratsios
ETH Zürich
anastasis.kratsios@unibas.ch

MS74

Fitting a Manifold of Large Reach to Noisy Data

We give a solution to the following question from manifold learning. Suppose data belonging to a high dimensional Euclidean space is sampled independently, identically at random, from a measure supported on a d dimensional twice differentiable embedded manifold M, and corrupted by Gaussian noise with small standard deviation sigma. How can we produce a manifold M_o whose Hausdorff distance to M is small and whose reach (normal injectivity radius) is not much smaller than the reach of M? We show how to produce a manifold within $O(\sigma^2)$ of M in Hausdorff distance, whose reach is smaller than that of M by a factor of no more than $O(d^6)$. This is joint work with Charles Fefferman, Sergei Ivanov and Matti Lassas.

Hariharan Narayanan
Tata Institute for Fundamental Research, India
hariharan.narayanan@tifr.res.in

MS74

Universal Manifold Approximation for Inverse Problems via Simple Injective Flows

We study approximations of one-chart n-manifolds embedded in $\mathbb{R}^m$ by injective flows. We show that in general, neural networks using injective flows universally approximate measures supported on images of extendable embeddings. If the embedding dimension is sufficiently large, $m \geq 3n+1$, we use an argument from algebraic topology known as the clean trick to prove that injective flows universally approximate any differentiable embedding. We believe that our work can have applications to inverse problem, whereby a proxy to the inverse operator can be obtained, by modeling the forward operator with a neural network which is guaranteed to be invertible.

Michael Puthawala
Rice University
michael.puthawala@gmail.com

MS75

Functional Dimension and Moduli Spaces of ReLU

Neural Network Functions

It is well-known that the parameterized family of functions $\mathbb{R}^n \to \mathbb{R}$ representable by neural networks with ReLU activation function is precisely the class of piecewise linear (PL) functions $\mathbb{R}^n \to \mathbb{R}$ with finitely many pieces. It is less well-known that for every fixed architecture of ReLU neural network, the parameter space admits positive-dimensional spaces of symmetries, and hence the local functional dimension near any given parameter is lower than the parametric dimension. In this talk I will explain exactly what I mean by local functional dimension, convince you that it is far from constant across parameter space, and speculate about the implications for training a ReLU neural network via stochastic gradient descent. This is joint work with Kathryn Lindsey, Rob Meyerhoff, and Chenxi Wu.

Elisenda Grigsby
Boston College Math Department
egrigsby@gmail.com

MS75

The Geometry of Memoryless Stochastic Policy Optimization in Infinite-Horizon POMDPs

We study the reward maximization problem in partially observable Markov decision processes with stationary stochastic policies, which constitute an important model in sequential decision making under uncertainty. We obtain a description of the problem as the optimization of a linear objective over the probability simplex subject to polynomial constraints that we characterize explicitly. This allows us to describe the combinatorial and algebraic complexity of the problem and we obtain bounds on its algebraic degree. Further, we conduct experiments using tools from constrained optimization, which demonstrate stability improvements of this approach to several baselines. This is joint work with Guido Montfar as well as Mareike Dressler, Marina Garotte-Lpez and Kemal Rose.

Johannes Müller
Max Planck Institute for Mathematics in the Sciences
johannes.mueller@mis.mpg.de

MS75

Consistency of Infinitely Wide and Deep Neural Network Classifiers

While neural networks are used for classification tasks across domains, a long-standing open problem in machine learning is determining whether neural networks trained using standard procedures are Bayes optimal for classification. In this work, we identify and construct a set of neural network classifiers that achieve Bayes optimality. Since effective neural networks in practice are typically both wide and deep, we analyze infinitely wide networks that are also infinitely deep. Using the recent connection between infinitely wide neural networks and Neural Tangent Kernels, we provide explicit activation functions that can be used to construct networks that achieve optimality. Interestingly, these activation functions are simple and easy to implement, yet differ from commonly used activations such as ReLU or sigmoid. More generally, we create a taxonomy of infinitely wide and deep networks and show that these models implement one of three well-known classifiers depending on the activation function used: (1) 1-nearest neighbor (model predictions are given by the label of the nearest training example); (2) majority vote (model predictions are given by the label of the class with greatest representation in the training set); or (3) singular kernel classifiers (a set of classifiers containing those that achieve optimality). Our results highlight the benefit of using deep networks for classification tasks, in contrast to regression tasks, where excessive depth is harmful.

Adityanarayanan Radhakrishnan
Massachusetts Institute of Technology
aradha@mit.edu

MS76

Optimal Transport and the Geometry of Collider Data

Collider experiments probe physics at the shortest distances by smashing protons together and measuring the debris produced by the collisions. However, it is challenging to infer the detailed short-distance physics from the distribution of debris measured by the detector. Over the past two years, optimal transport has emerged as a promising tool for classifying collider events based on these distributions. In this talk, we will describe how optimal transport distances endow collider data with a geometric structure that can be used for ML-based event classification, and we will show how selecting optimal transport distances with good geometric properties can dramatically reduce computational effort.

Katy Craig
University of California, Santa Barbara
kcraig@math.ucsb.edu

Nathaniel Craig, Tianji Cai
UC Santa Barbara
ncraig@physics.ucsb.edu, tianji_cai@ucsb.edu

MS76

Wasserstein Distance and Transport Transforms for Classification and Parametric Signals Estimation

We will introduce a mathematical signal transform based on Optimal Transport Theory. We provide a measure theoretic framework to extend the existing Cumulative Distribution Transform by Park, Kolouri, Kundu, and Rohde to arbitrary signals in one dimension. We will present both forward (analysis) and inverse (synthesis) formulas for the transform, and describe several of its properties including translation, scaling, convexity, and linear separability. We will describe a metric in transform space, and demonstrate the application of the transform in classifying (detecting) signals under random displacements, and parameter estimation problems. Collaborators: Akram Aldroubi, Ivan Medri, Gustavo K. Rohde, and Sumati Thareja.

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

MS77

Significance Tests for Neural Networks

We develop a pivotal test to assess the statistical significance of the feature variables in a single-layer feedforward neural network regression model. We propose a gradient-based test statistic and study its asymptotics using nonparametric techniques. Under technical conditions, the limiting distribution is given by a mixture of chi-square dis-

tributions. The tests enable one to discern the impact of individual variables on the prediction of a neural network. The test statistic can be used to rank variables according to their influence. Simulation results illustrate the computational efficiency and the performance of the test. An empirical application to house price valuation highlights the behavior of the test using actual data. Joint work with Enguerrand Horel, Stanford

Kay Giesecke
Stanford University
Dept. of Management Science and Engineering
giesecke@stanford.edu

MS78

Reduced Order Models for Multi-Material ALE Hydrodynamics

Although many model reduction schemes have been developed to reduce the computational cost of simulations while minimizing the error introduced in the reduction process, there are challenges especially in nonlinear advection-dominated problems such as sharp gradients, moving shock fronts, and turbulence, which hinder those model reduction schemes from being practical. In this talk, we will present recent developments in reduced order models for such simulations, with which we will demonstrate both good accuracy and considerable speed-up, enabling faster simulation and optimization solvers.

Dylan M. Copeland
Lawrence Livermore National Laboratory, US
copeland11@llnl.gov

MS78

Machine Learning Constitutive Models of Inelastic Materials with Microstructure

Traditional simulations of complex physical processes are both crucial technologically and expensive computationally. Furthermore the development of physical models via traditional methods typically is particularly time-consuming in human terms. Developing comparably accurate models directly from data can enable rapid development of accurate models as well as more robust design, uncertainty quantification, and exhaustive structure-property exploration. We have been developing neural network models that are guided by traditional constitutive theory, such as tensor function representation theorems to embed symmetries, and also exploiting deep learning to infer intrinsic microstructural features. Neural networks are flexible since sub-components of their graph-like structure can be arranged to suit particular tasks, such as image processing and time integration, and flow of information. This talk will describe the architectures and demonstrate the efficacy of neural networks designed to model the response of complex history-dependent materials with pores, inclusions or grains based solely on observable data. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engi- neering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Depart- ment of Energys National Nuclear Security Administration under contract DE-NA-0003525.

Reese Jones, Ari Frankel, Cosmin Safta, Kyle Johnson
Sandia National Laboratories
rjones@sandia.gov, alfrank@sandia.gov, csafta@sandia.gov, kyljohn@sandia,gov

MS78

Stabilized Neural Ordinary Differential Equations for Learning Chaotic Dynamical Systems

This talk will showcase recent research that builds fast and accurate reduced-order models (ROMs) for various high-dimensional dynamical systems. ROMs are computationally efficient approximations of their full-order counterparts and are tasked with making accurate forecasts in time, given parameters and/or varying initial and boundary conditions. Our ROMs are constructed with several scientific machine learning strategies, specifically, those that rely on the coupling of deep learning and differential equations, to dramatically improve accuracy and time-to-solution for extended computational campaigns. Furthermore, in addition to canonical experiments, this talk will present results for real-world applications of strategic importance. Some examples are building ROMs for geophysical forecasting from ship and satellite observation data and wind-turbine wake predictions from meteorological and LIDAR measurements.

Romit Maulik
Argonne National Laboratory
rmaulik@anl.gov

MS78

Interlaced Characterization and Calibration of Elastoplastic Constitutive Models

Calibration of constitutive models in solid mechanics remains both challenging and costly, despite the widespread availability of high-quality full-field deformation measurement techniques (e.g. digital image/volume correlation). Furthermore, the process of collecting data from a mechanical test and its ultimate utilization within a computational model calibration workflow are often loosely coupled. To address these issues, we propose an Interlaced Characterization and Calibration (ICC) paradigm that leverages concepts from Bayesian optimal experimental design (OED) to actively drive an experiment that is tailored to the accurate calibration of a given material. In this talk we will provide an overview of the ICC concept and its application to elastoplastic constitutive models and share technical progress on a few of its key components. These topics will include the construction of surrogate model graphs that replace expensive computational simulations of the experiment and the "step selection" OED problem, i.e. given a set of choices for the next material state and current posterior distribution, which one should we choose? SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Daniel T. Seidl, Denielle Ricciardi, Brian Lester, Amanda Jones, Elizabeth Jones
Sandia National Laboratories
dtseidl@sandia.gov, dericci@sandia.gov, btleste@sandia.gov, ajones1@sandia.gov, emjones@sandia.gov

MS79

Asymptotics of Unrolled Convex Optimization Al-

gorithms

In this talk, we study the approach of learning parameters of a convex optimization problem from data by unrolling a fixed number of L iterations of an associated iterative algorithm. We provide a rigorous analysis of the asymptotic behavior of the resulting parameter gradients in the limiting case $L \to \infty$. Finally, we derive a truncated backpropagation scheme to tackle the underlying bi-level optimization problem and present related numerical results from image and speech processing.

Christoph Brauer
TU Braunschweig
Institute of Analysis and Algebra
ch.brauer@tu-bs.de

Dirk Lorenz
TU Braunschweig
d.lorenz@tu-braunschweig.de

MS79

Unrolling Versus Bilevel Optimization in the Context of Learning Variational Models

In image and signal processing and beyond, quantities of interest are often reconstructed from potentially noisy observations by means of suitably parameterized energy minimization problems. An unknown ground truth is approximately recovered in terms of an optimization problem where the energy to be minimized depends on an observation and parameters. While purely model based approaches rely on the assumption of a priori availability of a suitable energy function and associated parameters, data driven approaches are commonly used when choosing an appropriate energy is not obvious or infeasible. Then, the parameters or parts of them shall be learned from data. This gives rise to bilevel optimization problems in which the original convex problem (hereafter referred to as the lower level problem) appears as a constraint. Those bilevel problems are hard to solve. Moreover, in many cases one cannot even assume that the lower level problem has a unique solution. An approach to avoid solving the lower level problem explicitly is the so-called unrolling. Thereby, the minimizer is replaced with an iterate of an algorithm known to be capable of solving the problem. In this contribution we consider the approach to unroll a fixed number of iterations of gradient descent to the lower level problem of a tractable toy model. We compare this approach to solving the lower level problem explicitly by investigating the expressivity and computing the corresponding risks in a fixed framework.

Niklas Breustedt
TU Braunschweig
n.breustedt@tu-braunschweig.de

Christoph Brauer
TU Braunschweig
Institute of Analysis and Algebra
ch.brauer@tu-bs.de

Timo de Wolff
Technische Universität Braunschweig
t.de-wolff@tu-braunschweig.de

Dirk Lorenz
TU Braunschweig
d.lorenz@tu-braunschweig.de

MS79

Algorithm Unrolling - Key for Physics-Based Learning in Magnetic Resonance Image Reconstruction

During the past years, deep learning has evolved tremendously in the research field of MR image reconstruction. While purely data-driven methods are typically implemented in the image domain and discard the MR acquisition physics entirely, physics-driven methods are the key to the success of learning-based MRI reconstruction. In this talk, I will guide you through the developments of physics-based learning, and introduce you to learned unrolled algorithms. We will connect learned unrolled algorithms to advanced regularization schemes and generative models, and discuss the challenges and caveats of deep learning for image reconstruction. I will showcase real-world examples ranging from 2D musculoskeletal imaging to higher-dimensional cardiac imaging that will underpin the vast potential of learned unrolled algorithms to shape the future of fast MR image acquisition and reconstruction.

Kerstin Hammernik
Technical University of Munich
k.hammernik@tum.de

Wenqi Huang
Technical University of Munich, Germany
wenqi.huang@tum.de

MS79

Hyperparameter Tuning Is All You Need for Lista

Learned Iterative Shrinkage-Thresholding Algorithm (LISTA) introduces the concept of unrolling an iterative algorithm and training it like a neural network. It has had great success on sparse recovery. In this paper, we show that adding momentum to intermediate variables in the LISTA network achieves a better convergence rate and, in particular, the network with instance-optimal parameters is superlinearly convergent. Moreover, our new theoretical results lead to a practical approach of automatically and adaptively calculating the parameters of a LISTA network layer based on its previous layers. Perhaps most surprisingly, such an adaptive-parameter procedure reduces the training of LISTA to tuning only three hyperparameters from data: a new record set in the context of the recent advances on trimming down LISTA complexity. We call this new ultra-light weight network HyperLISTA. Compared to state-of-the-art LISTA models, HyperLISTA achieves almost the same performance on seen data distributions and performs better when tested on unseen distributions (speci?cally, those with different sparsity levels and nonzero magnitudes).

Jialin Liu
Alibaba DAMO academy
jialin.liu@alibaba-inc.com

MS80

Towards a Theory of Non-Log-Concave Sampling: First-Order Stationarity Guarantees for Langevin Monte Carlo

Although many recent works explore the interplay between

optimization and sampling, thus far there has not been a sampling analogue of the approximate first-order stationarity guarantees from non-convex optimization. Indeed, despite practical interest in non-log-concave sampling, most papers adopt additional assumptions (e.g. dissipativity conditions or functional inequalities) and do not consider a truly general framework for this problem. In this work, we propose using bounds on the relative Fisher information as a notion of approximate first-order stationarity in sampling; this perspective is justified by the interpretation of the Langevin dynamics as a gradient flow on the space of probability measures. We then show that Langevin Monte Carlo achieves approximate first-order stationarity in polynomially many steps, under the sole assumption that the target distribution is log-smooth. This is joint work with Krishnakumar Balasubramanian, Murat A. Erdogdu, Adil Salim, and Matthew Zhang.

Sinho Chewi
Massachusetts Institute of Technology
schewi@mit.edu

MS80

A Proximal Algorithm for Sampling

We consider sampling problems with possibly non-smooth potentials. In particular, we study two specific settings of sampling where the convex potential is either semi-smooth or in composite form as the sum of a smooth component and a semi-smooth component. To overcome the challenges caused by the non-smoothness, we propose a Markov chain Monte Carlo algorithm that resembles proximal methods in optimization for these sampling tasks. The key component of our method is a sampling scheme for a quadratically regularized target potential. This scheme relies on rejection sampling with a carefully designed Gaussian proposal whose center is an approximate minimizer of the regularized potential. We develop a novel technique (a modified Gaussian integral) to bound the complexity of this rejection sampling scheme in spite of the non-smoothness in the potentials. We then combine this scheme with the alternating sampling framework (ASF), which uses Gibbs sampling on an augmented distribution, to accomplish the two settings of sampling tasks we consider. Furthermore, by combining the complexity bound of the rejection sampling we develop and the remarkable convergence properties of ASF discovered recently, we are able to establish several non-asymptotic complexity bounds for our algorithm, in terms of the total number of queries of subgradient of the target potential. Our algorithm achieves state-of-the-art complexity bounds compared with all existing methods in the same settings.

Jiaming Liang
Yale University
jiaming.liang@yale.edu

Yongxin Chen
Georgia tech
yongchen@gatech.edu

MS80

Structured Logconcave Sampling with a Restricted Gaussian Oracle

We give algorithms for sampling several structured logconcave families to high accuracy. We further develop a reduction framework, inspired by proximal point methods in convex optimization, which bootstraps samplers for regularized densities to generically improve dependences on problem condition number from polynomial to linear. A key ingredient in our framework is the notion of a "restricted Gaussian oracle' (RGO) for function g, which is a sampler for distributions whose negative log-likelihood sums a quadratic (in a multiple of the identity) and g. By combining our reduction framework with our new samplers, we obtain the state-of-the-art bounds for sampling composite densities and finite-sum densities to high accuracy in total variation distance.

Ruoqi Shen
University of Washington, U.S.
shenr3@cs.washington.edu

Kevin Tian
Stanford University
tiankevin@microsoft.com

MS80

On Forward and Backward Methods for Sampling and Optimization

Sampling and optimization are intricately connected, and there has been a productive exchange of ideas between them. In particular, the perspective of sampling as optimization in the space of distributions allows a translation of techniques from optimization to sampling by applying them in the space of distributions. At the continuous-time level, various dynamics for optimization give rise to new dynamics for sampling with fast convergence guarantees, and a natural class of them can be implemented as stochastic processes in space. However, in discrete time, different choices of discretization methods give rise to algorithms with different convergence properties, with the typical difference between the forward method which is explicit and the backward method which is implicit but stable. Basic discretization schemes for sampling suffer from the issue of asymptotic bias, which is typically not present in optimization. The size of the bias becomes a bottleneck in the complexity in discrete time, and the question of understanding the bias becomes more intricate for constrained sampling with non-Euclidean geometry. In this talk, we will survey some dynamics and algorithms for sampling derived from optimization. We will describe their convergence properties under log-concavity or isoperimetry, and highlight their differences in behavior based on the forward or backward methods.

Andre Wibisono
Yale University
andre.wibisono@yale.edu

MS81

Rigorous Data-Driven Computation of Spectral Properties of Koopman Operators in Dynamical Systems

Koopman operators are infinite-dimensional operators that globally linearise nonlinear dynamical systems, making their spectral information valuable for understanding dynamics. Their increasing popularity in data-driven methods, dubbed Koopmania, includes 10,000s of articles over the last decade. However, Koopman operators can have continuous spectra and lack finite-dimensional invariant subspaces, making computing their spectral properties a considerable challenge. This talk describes data-driven algorithms with rigorous convergence guarantees for computing spectral properties of Koopman operators from trajec-

tory data. We introduce residual dynamic mode decomposition (ResDMD), the first scheme for computing the spectra and pseudospectra of general Koopman operators from trajectory data without spectral pollution. By combining ResDMD and the resolvent, we compute smoothed approximations of spectral measures associated with measure-preserving dynamical systems. When computing the continuous and discrete spectrum, explicit convergence theorems provide high-order convergence, even for chaotic systems. Kernelized variants of our algorithms allow for dynamical systems with a high-dimensional state-space. We end with data-driven computations of spectral measures of a protein molecule (20,046-dimensional state-space) and nonlinear Koopman modes with error bounds for chaotic turbulent flow past aerofoils with Reynolds number $> 10^5$ (295,122-dimensional state-space).

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

Alex J. Townsend
Cornell University
townsend@cornell.edu

MS81

Neural Stochastic Partial Differential Equations: Resolution-Invariant Learning of Continuous Spatiotemporal Dynamics

Stochastic partial differential equations (SPDEs) are the mathematical tool of choice for modelling spatiotemporal PDE-dynamics under the influence of randomness. Based on the notion of mild solution of an SPDE, we introduce a neural architecture to learn solution operators of PDEs with (possibly stochastic) forcing from partially observed data. The proposed Neural SPDE model provides an extension to two popular classes of physics-inspired architectures. On the one hand, it extends Neural CDEs and variants – continuous-time analogues of RNNs – in that it is capable of processing incoming sequential information arriving irregularly in time and observed at arbitrary spatial resolutions. On the other hand, it extends Neural Operators – generalizations of neural networks to model mappings between spaces of functions – in that it can parameterize solution operators of SPDEs depending simultaneously on the initial condition and a realization of the driving noise. By performing operations in the spectral domain, we show how a Neural SPDE can be evaluated in two ways, either by calling an ODE solver (emulating a spectral Galerkin scheme), or by solving a fixed point problem. Experiments on various semilinear SPDEs, including the stochastic Navier-Stokes equations, demonstrate how the Neural SPDE model is capable of learning complex spatiotemporal dynamics in a resolution-invariant way, with better accuracy and lighter training data requirements compared to alternative models.

Maud Lemercier
University of Warwick
maud.lemercier@maths.ox.ac.uk

MS81

Learning Partial Differential Equations in Reproducing Kernel Hilbert Spaces

The rapid development of data-driven scientific discovery holds the promise of new and faster methods to understand and predict complex phenomena whose physical laws are still beyond our grasp. Of central interest to this development is the ability to efficiently solve partial differential equations (PDEs) for a variety of physical inputs: initial conditions, boundary conditions, and forcings. In recent years, neural networks and kernel methods have accelerated our ability to efficiently simulate solutions of PDEs but statistical guarantees for these approaches are difficult to show. In this work, we focus on estimating solution operators of linear PDEs and estimate the PDE solution operator in a reproducing kernel Hilbert space. In this setting, we show rigorous statistical guarantees on the prediction error of our estimator even in the misspecified case when the underlying PDE is nonlinear. This approach allows us to regularize the smoothness of the PDE solution operator and impose various structural constraints such as symmetry and time-causality which appear in many real-world systems. We conclude by demonstrating applications of our method to several linear PDEs including the Poisson, Helmholtz, Schrdinger, Fokker-Planck, and heat equation and highlight its ability to extrapolate to more finely sampled data without any additional training.

George Stepaniants
MIT
gstepan@mit.edu

MS81

Learning Constitutive Relations in Multiscale Viscoelastic Materials

The macroscopic behavior of materials is governed in part by small-scale rapidly-varying material properties. Fully resolving these features within the balance laws thus involves expensive fine-scale computations which need to be conducted on macroscopic scales. The theory of homogenization provides an approach to derive effective macroscopic equations which eliminates the small scales by exploiting scale separation. An accurate homogenized model avoids the computationally-expensive task of numerically solving the underlying balance laws at a fine scale. In simple settings the homogenization produces an explicit formula for a macroscopic constitutive model, but in more complex settings it may only define the constitutive model implicitly. In these complex settings machine learning can be used to learn the constitutive model from localized fine-scale simulations. In the case of one-dimensional viscoelasticity, the linearity of the model allows for a complete analysis. For this case, we derive a homogenized constitutive model and develop a theory to prove that the model may be approximated by a recurrent neural network (RNN) model that captures the memory; this may be thought of as discovering appropriate internal variables. Simulations are presented which validate the theory, and additional numerical experiments demonstrate extension of the methodology to higher dimensions and to nonlinear viscoplasticity.

Margaret K. Trautner
California Institute of Technology
mtrautne@caltech.edu

Andrew Stuart
Computing + Mathematical Sciences
California Institute of Technology
astuart@caltech.edu

Burigede Liu
University of Cambridge
BL377@cam.ac.uk

Kaushik Bhattacharya
Howell N. Tyson Sr. Professor of Mechanics
California Institute of Technology
bhatta@caltech.edu

MS82

Data-Driven Modeling of Diffusive Closures in Geophysical Flows

Hybrid models, which combine traditional partial differential equation (PDE) solvers with machine learning-based closures, have the potential to provide more accurate and reliable predictions. In the present work, we study two aspects of hybrid models i) how to improve the overall numerical stability; ii) how to train these machine learning-based closures with indirect (observation) data, which generally do not provide direct information about small-scale processes. We focus on diffusive closures, which are commonly used for modeling fluid flows featuring mixing phenomena with a wide range of eddy sizes. We present a kernel-smoothed neural network architecture, which improves stability and captures nonlocal effects, and corresponding training strategies based on the Kalman methodology. We test our framework by studying turbulence modeling in geophysics flows. The limitations of our framework are also discussed.

Daniel Huang, Zhaoyi Shen, Jinlong Wu, Hongkai Zheng
California Institute of Technology
dzhuang@caltech.edu, zhaoyi@caltech.edu,
jinlong@caltech.edu, hzzheng@caltech.edu

Anima Anandkumar
Amazon and California Institute of Technology
anima@caltech.edu

Schneider Tapio
California Institute of Technology
tapio@caltech.edu

MS82

Non-Intrusive Parametric Reduced-Order Modeling via Operator Inference

We formulate a new approach to reduced modeling of parameterized, time-dependent partial differential equations (PDEs). The method employs Operator Inference, a scientific machine learning framework combining data-driven learning and physics-based modeling. The parametric structure of the governing equations is embedded directly into the reduced-order model, and parameterized reduced-order operators are learned via a data-driven linear regression problem. The result is a reduced-order model that can be solved rapidly to map parameter values to approximate PDE solutions. Such parameterized reduced-order models may be used as physics-based surrogates for uncertainty quantification and inverse problems that require many forward solves of parametric PDEs. Numerical issues such as well-posedness and the need for appropriate regularization in the learning problem are considered, and an algorithm for hyperparameter selection is presented. The method is illustrated for a parametric heat equation and demonstrated for the FitzHugh-Nagumo neuron model.

Shane A. McQuarrie
Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin
shanemcq@utexas.edu

Parisa Khodabakhshi, Karen E. Willcox
UT Austin
parisa@austin.utexas.edu, kwillcox@oden.utexas.edu

MS82

Deep Operator Networks Inspired by Approximations to Variational Formulations

Deep Operator Networks (DeepONets) approximate operators that map functions from one Banach space to another. A DeepONet accommodates this very large class of operators by allowing significant latitude in the network architecture. In this talk we consider the specific case where a DeepONet approximates operators that emerge from the solution of PDEs and ask the question whether this knowledge leads to preferred network architectures with significant computational benefits. Inspired by the variational formulation for solving PDEs (which forms the basis for finite element and spectral methods) we conclude that this is indeed the case, and propose efficient architectures and training paradigms.

Dhruv V. Patel
Stanford University
dvpatel@stanford.edu

Michael Abdelmalik
Eindhoven University of Technology
michael.abdelmalik@austin.utexas.edu

Deep Ray
University of Southern California
deepray@usc.edu

Thomas Hughes
ICES
University of Texas at Austin
hughes@ices.utexas.edu

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

MS82

Extrapolation of Deep Operator Networks via Physics or Sparse Observations

Neural operators can learn nonlinear mappings between function spaces and offer a new simulation paradigm for real-time prediction of complex dynamics in realistic diverse applications as well as for system identification in science and engineering. Herein, we investigate one promising operator, deep operator network (DeepONet). Most of the current research on neural operators assumes that training and testing data have the same distribution, which is not always true in practice. Interpolation of the training data typically has good performance, while extrapolation of the training data often suffers from high error. In this study, five methods were proposed to reduce the error and increase accuracy in extrapolation with only a few observations. The first method is to fine-tune the pre-trained DeepONet via the physics-informed neural networks (PINN) loss. The second method is to fine-tune the pre-trained DeepONet using the observation data via the mean square error (MSE) loss. To avoid overfitting caused by only using the observation data, a third method was proposed, which is to fine-tune the pre-trained DeepONet with both observation data and previous DeepONet train-

ing data. The fourth and fifth methods are multi-fidelity models via gaussian process and neural networks, which employ data generated by DeepONet as low-fidelity data, and the observation data as high-fidelity data.

Min Zhu, Lu Lu, Handi Zhang
University of Pennsylvania
zhum20@seas.upenn.edu, lulu1@seas.upenn.edu, handi@sas.upenn.edu

MS83

Modeling Fluid Flows via Sparse Physically Informed Discovery of Empirical Relations (SPIDER)

We introduce a general approach to data-driven modeling of spatially extended systems called sparse physics-informed discovery of empirical relations (SPIDER). The models take the easily interpretable form of a system of algebraic or partial differential equations and are constructed using sparse regression applied to the weak form of these equations. A key challenge is constructing a family of candidate models that is both sufficiently comprehensive to include the relevant physics and compact enough for the computation to remain tractable. We address this challenge using a tensor-based decomposition, yielding a set of term libraries corresponding to different irreducible representations of the complete symmetry group of the physical problem. The approach is illustrated using representative examples of turbulent fluid flow and active matter whose models are identified from experimental data. We also demonstrate the utility of SPIDER for validating and debugging results of numerical simulations. In particular, we show that it enables cheap and efficient diagnosis of a range of issues, from inaccurate solutions of particular equations or the use of incorrect (or incorrectly implemented) boundary conditions to symmetry breaking resulting from discretization.

Roman Grigoriev
Georgia Institute of Technology
Center for Nonlinear Science and School of Physics
roman.grigoriev@physics.gatech.edu

Matthew Golden
Georgia Institute of Technology
mgolden30@gatech.edu

Daniel Gurevich
Princeton University
dgurevich@princeton.edu

MS83

Using Wsindy to Learn Anisotropic Interaction Rules from Individual Trajectories in a Heterogeneous Cellular Population

Interacting particle system (IPS) models have proven to be highly successful for describing the spatial movement of organisms. However, it is challenging to infer the interaction rules directly from data. In the field of equation discovery, the Weak form Sparse Identification of Nonlinear Dynamics (WSINDy) methodology has been shown to be computationally efficient for identifying the governing equations of complex systems from noisy data. Motivated by the success of IPS models to describe the spatial movement of organisms, we develop WSINDy for second order IPS to learn equations for communities of cells. Our approach learns the directional interaction rules for each individual cell that in aggregate, govern the dynamics of a heterogeneous population of migrating cells. To sort a cell according to the active classes present in its model, we also develop a novel ad hoc classification scheme (which accounts for the fact that some cells do not have enough evidence to accurately infer a model). Aggregated models are then constructed hierarchically to simultaneously identify different species of cells present in the population and determine best-fit models for each species. We demonstrate the efficiency and proficiency of the method on several test scenarios, motivated by common cell migration experiments.

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

David M. Bortz
University of Colorado
Department of Applied Mathematics
dmbortz@colorado.edu

MS83

Identifying Differential Equation: Weak Form Approach

We consider identifying differential equation using numerical techniques (IDENT) from one set of noisy observation. We assume that the governing PDE can be expressed as a linear combination of different linear and nonlinear differential terms. In this talk, extending from IDENT and robust IDENT, we will discuss using weak form for differential equation identification. We consider both ODE and PDE models. Numerical results show robustness against higher level of noise and higher order derivative in underlying equation.

Sung Ha Kang
Georgia Institute of Technology
Mathematics
kang@math.gatech.edu

Wenjing Liao
Georgia Institute of Technology
wliao60@gatech.edu

Mengyi Tang, Rachel Kuske
Georgia Tech
tangmengyi@gatech.edu, rachel@math.gatech.edu

MS83

How Much Can One Learn a PDE from its Solution Data?

In this work we study a few basic questions for PDE learning from observed solution data. Using various types of PDEs, we show 1) how the richness of the data space spanned by all snapshots along a solution trajectory depends on the differential operator and initial data, 2) how much can one learn from a single solution, and 3)identifiability of a differential operator from solution data on local patches. Then we propose a consistent and sparse local regression method for general PDE identification with guarantee. Our method requires minimal amount of local measurements in space and time from a single solution trajectory by enforcing global consistency and sparsity.

Hongkai Zhao

Duke University
hongkai.zhao@duke.edu

Yuchen He
Shanghai Jiao Tong University
yuchenroy@sjtu.edu.cn

Yimin Zhong
Duke University
yz678@math.duke.edu

MS84

Parameter Estimation with Convolutional Neural Networks Applied to An ODE and a Maximum Likelihood Problem

Many inverse problems include poorly behaved objective functions or computationally infeasible likelihood functions such as for multidimensional extremes, thus making traditional approaches for computational recovery of model parameters intractable. We propose a deep learning framework to estimate parameters of two models respectively governed by ordinary differential equations and statistical models for multidimensional extremes. In both cases, we use data from model simulations as input to train deep neural networks and learn the output ODE/statistical parameters. Our neural network-based method provides a competitive alternative to existing methods such as pairwise likelihood for multidimensional extremes, as demonstrated by considerable accuracy and computational time improvements.

Julie Bessac
Argonne National Laboratory
jbessac@anl.gov

MS84

Learning Stochastic Parametric Differentiable Predictive Control Policies

The problem of synthesizing stochastic explicit model predictive control policies is known to be quickly intractable even for systems of modest complexity when using classical control-theoretic methods. To address this challenge, we present a scalable alternative called stochastic parametric differentiable predictive control (SP-DPC) for unsupervised learning of neural control policies governing stochastic linear systems subject to nonlinear chance constraints. SP-DPC is formulated as a deterministic approximation to the stochastic parametric constrained optimal control problem. This formulation allows us to directly compute the policy gradients via automatic differentiation of the problem's value function, evaluated over sampled parameters and uncertainties. In particular, the computed expectation of the SP-DPC problem's value function is backpropagated through the closed-loop system rollouts parametrized by a known nominal system dynamics model and neural control policy which allows for direct model-based policy optimization. We provide theoretical probabilistic guarantees for policies learned via the SP-DPC method on closed-loop stability and chance constraints satisfaction. Furthermore, we demonstrate the computational efficiency and scalability of the proposed policy optimization algorithm in three numerical examples, including systems with a large number of states or subject to nonlinear constraints.

Jan Drgona
Pacific Northwest National Laboratory
jan.drgona@pnnl.gov

MS84

A Method for Cost-Aware, Adaptive, Multifidelity, Efficient Reliability Analysis

Estimating probability of failure in aerospace systems is a critical requirement for flight certification and qualification. We propose a method to use models of multiple fidelities that trade accuracy for computational efficiency. Specifically, we propose the use of multifidelity Gaussian process models to efficiently fuse models at multiple fidelity, thereby offering a cheap surrogate model that emulates the original model at all fidelities. Furthermore, we propose a novel sequential *acquisition function*-based experiment design framework that can automatically select samples from appropriate fidelity models to make predictions about quantities of interest in the highest fidelity. We use our proposed approach in an importance sampling setting and demonstrate our method on the failure level set estimation and probability estimation on synthetic test functions as well as two real-world applications, namely, the reliability analysis of a gas turbine engine blade using a finite element method and a transonic aerodynamic wing test case using Reynolds-averaged Navier–Stokes equations. We demonstrate that our adaptively constructed multifidelity surrogate model is guaranteed to predict the true failure level set with high probability.

Ashwin Renganathan
University of Utah
ashwin.renganathan@utah.edu

MS84

Bayesian Inference with Gaussian Process Models for Multiscale Data and ModelIntegration

The effective utilization of environmental datasets is often hindered by their sparseness in space and time. Although there are significant advances recently in non-invasive spatially extensive technologies such as remote sensing and geophysics, it is still challenging to integrate these multiscale and multi-type datasets with different resolutions, coverage and accuracy. This talk will present the two applications of Gaussian Process Models (GPMs) to environmental datasets to interpolate sparse datasets and to integrate different data types with uncertainty quantified. First, we have developed a multi-scale data integration method based on the Bayesian hierarchal model and GPMs to integrate different radiation surveys walk, car and airborne surveys in the Fukushima region. Different spatial resolutions are explicitly represented using the weighted spatial average based on radiation transport simulations. Second, we have developed a machine learning (ML) framework for groundwater contamination monitoring, including QA/QC, temporal gap filling, interpolation and sensor placement optimization. In particular, GPMs are used to improve the interpolation of contaminant concentrations between sparse wells by integrating (1) spatially extensive predictors such as topographic metrics, (2) the correlations between contaminant concentrations and in situ sensor-measured variables, and (3) the model simulated plume distribution.

Haruko M. Wainwright
Massachusetts Institute of Technology

hmwainw@mit.edu

MS85

Handling Infinitely Many Inequality Constraints in Function Optimization Problems Using Kernel Methods

Many optimization problems over function spaces involve an infinite number of pointwise inequality constraints. This side information can originate from both physical and theoretical constraints on the model such as "stay within boundaries" in path-planning or "be nonnegative and integrate to one" in density estimation. On the other hand, reproducing kernels are propitious for pointwise evaluations and some kernels encode very rich classes of functions. However, representer theorems, which ensure the numerical applicability of kernels, cannot be applied for an infinite number of evaluations. Through constructive algebraic and geometric arguments, I will present how to tackle this question by perturbing the constraints, through coverings in infinite dimensions and through kernel Sum-Of-Squares. Both schemes entail an extra computational price, involving second-order conic or SDP constraints, but assessing the amount of perturbation enables to prove rates on the convergence of the schemes. Tightening guarantees the satisfaction of the constraints and was used in state-constrained optimal control and density estimation. Relaxation instead gives better approximation rates for global optimization.

Pierre-Cyril Aubin-Frankowski
INRIA Paris
pierrecyrilaubin@gmail.com

MS85

Scalable Uncertainty Quantification with Random Features

The focus of this talk is on ways random feature methods (RFs) can enhance Gaussian process (GP) regression. RFs were originally developed as a tool to scale up kernel methods in supervised machine learning to large datasets. The idea is to make a relatively sparse, parametric, and statistically founded Monte Carlo approximation to a positive definite kernel function. While RFs have been widely adopted in the kernel methods literature, there has been significantly less attention on the GP aspects of RFs leading to uncertainty quantification (UQ). Under mild assumptions, we show that RFs enjoy superior computational complexity compared to the full GP posterior while still maintaining the hallmarks of GPs, namely, natural interpretability and automatic uncertainty estimates. Additionally, we develop a nonintrusive, derivative-free approach to learn RF or GP model hyperparameters under a noisy optimization landscape using ideas from ensemble Kalman methods and cross-validation. We apply our RF-UQ framework to problems ranging from dynamical systems to the climate sciences.

Nicholas H. Nelsen
California Institute of Technology
nnelsen@caltech.edu

MS85

Theoretical Guarantees for the Statistical Finite Element Method

The statistical finite element method (StatFEM) is an emerging probabilistic method that allows observations of a physical system to be synthesised with the numerical solution of a PDE intended to describe it in a coherent statistical framework, to compensate for model error. This work presents a new theoretical analysis of the statistical finite element method demonstrating that it has similar convergence properties to the finite element method on which it is based. Our results constitute a bound on the Wasserstein-2 distance between the ideal prior and posterior and the StatFEM approximation thereof, and show that this distance converges at the same mesh-dependent rate as finite element solutions converge to the true solution. Several numerical examples are presented to demonstrate our theory, including an example which tests the robustness of StatFEM when extended to nonlinear quantities of interest.

Yanni Papandreo
Imperial College London
john.papandreou18@imperial.ac.uk

MS86

Accelerating Learning Neural ODEs via Proximal Algorithms

Learning neural ODEs often requires solving very stiff ODE systems, primarily using explicit adaptive ODE solvers. These solvers are computationally expensive, requiring the use of tiny step sizes for numerical stability and accuracy guarantees. We consider learning neural ODEs using implicit ODE solvers of different orders leveraging proximal operators. The proximal implicit solver consists of inner-outer iterations: the inner iterations approximate each implicit update step using a fast optimization algorithm, and the outer iterations solve the ODE system over time. The proximal implicit ODE solver guarantees superiority over explicit solvers in numerical stability and computational efficiency.

Bao Wang
Data Science Research Lab in the DEPARTMENT OF MATHEMATICS a
bwang@math.utah.edu

Justin Baker
University of Utah, U.S.
u0726589@utah.edu

MS86

Learning to Solve Parametric PDEs

Deep learning continues to dominate machine learning and has been successful in computer vision, natural language processing, etc. Its impact has now expanded to many research areas in science and engineering. In this talk, I will present our recent work on combining wisdom from numerical PDEs (e.g., multigrid method) and machine learning to design data-driven solvers for parametric PDEs and their applications in electromagnetic simulation.

Bin Dong
Peking University
dongbin@math.pku.edu.cn

MS86

PDE-GCN: Novel Architectures for Graph Neural Networks Motivated by Partial Differential Equa-

tions

Graph neural networks are increasingly becoming the go-to approach in various fields such as computer vision, computational biology and chemistry, where data are naturally explained by graphs. However, unlike traditional convolutional neural networks, deep graph networks do not necessarily yield better performance than shallow graph networks. This behavior usually stems from the over-smoothing phenomenon. In this work, we propose a family of architectures to control this behavior by design. Our networks are motivated by numerical methods for solving Partial Differential Equations (PDEs) on manifolds, and as such, their behavior can be explained by similar analysis. Moreover, as we demonstrate using an extensive set of experiments, our PDE-motivated networks can generalize and be effective for various types of problems from different fields. Our architectures obtain better or on par with the current state-of-the-art results for problems that are typically approached using different architectures.

Moshe Eliasof
Ben Gurion University of the Negev
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

MS86

Controlling and Efficiently Computing Regularized Systems of Conservation Laws via Primal-Dual Methods

We propose, study, and compute solutions to a class of optimal control problems for hyperbolic systems of conservation laws and their viscous regularization Barotropic compressible Navier-Stokes equations are taken as a canonical example. We design a finite difference scheme for a variational system. The numerical approximations of conservation laws are implicit in time. We solve the variational problem with an algorithm inspired by the primal dual hybrid gradient method. This includes a new method for solving implicit time approximations for conservation laws, which seems to be unconditionally stable. Several numerical examples are presented to demonstrate the effectiveness of the proposed algorithm.

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

MS87

Hierarchical Nonnegative Tensor Factorizations and Applications

Nonnegative matrix factorization (NMF) has found many applications including topic modeling and document analysis. Hierarchical NMF (HNMF) variants are able to learn topics at various levels of granularity and illustrate their hierarchical relationship. Recently, nonnegative tensor factorization (NTF) methods have been applied in a similar fashion in order to handle data sets with complex, multi-modal structure. Hierarchical NTF (HNTF) methods have been proposed, however these methods do not naturally generalize their matrix-based counterparts. Here, we propose a new HNTF model which directly generalizes a HNMF model special case, and provide a supervised extension. Our experimental results show that this model more naturally illuminates the topic hierarchy than previous HNMF and HNTF methods.

Jamie Haddock
Harvey Mudd College
320 E Foothill Blvd, Claremont, CA 91711
jhaddock@g.hmc.edu

MS87

Optimal Iteratively Reweighted Least Squares Algorithms for Spectral Function Minimization

Low-rank matrix optimization problems have emerged as a powerful modeling tool within a variety of scientific communities ranging from data science, scientific computing to signal and image processing and beyond. It is well-understood that in many cases, a rank objective can be replaced by a non-convex surrogate such as a log-determinant or a Schatten-p quasinorm without any or only with little loss of modeling accuracy. However, the optimization of such non-convex spectral functions has remained challenging. Around 2011, Mohan and Fazel and Fornasier, Rauhut and Ward proposed iteratively reweighted least squares (IRLS) methods optimizing such spectral functions. While these methods have not been amenable for fast convergence, a scalable IRLS method with provable superlinear convergence has been proposed in [Kmmerle, Mayrink Verdun, "A scalable second order method for ill-conditioned matrix completion from few samples", ICML 2021]. In this talk, we present the optimality of this and related IRLS methods for the minimization of non-convex spectral functions, and show that the resulting quadratic models do not only include second-order information, but are also optimal in the sense that any smaller quadratic model does not retain a majorization property which is essential to ensure monotonicity of the algorithms. We illustrate with data efficiency of the methods with applications to matrix completion and phase retrieval.

Christian Kuemmerle
Johns Hopkins University
kuemmerle@jhu.edu

Felix Krahmer
Technical University of Munich
felix.krahmer@tum.de

Tomas Masák
Ecole Polytechnique Fédérale de Lausanne
tomas.masak@epfl.ch

Claudio Mayrink Verdun
Technical University of Munich
Department of Mathematics
verdun@ma.tum.de

MS87

Superiority of GNN over NN in Generalizing Bandlimited Functions

We constructively show, via rigorous mathematical arguments, that GNN architectures outperform those of NN in approximating bandlimited functions on compact d-dimensional Euclidean grids. We show that the former

only need $\mathcal{M}$ sampled functional values in order to achieve a uniform approximation error of $O_d(2^{-\mathcal{M}^{1/d}})$ and that this error rate is optimal, in the sense that, NNs might achieve worse.

Martina Neuman
Michigan State University
marsneuman.work@gmail.com

MS87

Multi-Dimensional Medical Imaging Data Reconstruction Using Kernel Low-Rank Minimization Scheme

Dynamic imaging is becoming more and more useful in clinics nowadays. For example, dynamic cardiac MRI and dynamic lung MRI is widely used to evaluate the heart and lung function in clinics. Multi-dimensional (spatial dimension, time dimension, respiratory dimension and etc.) data need to be acquired for getting desirable image quality. This requires long acquisition time to get the multi-dimensional data which satisfies the Nyquist sampling theorem for reconstruction. I this talk, we will talk about the usage of the kernel low-rank minimization scheme for the reconstructions of the medical images using highly undersampled multi-dimensional data.

Qing Zou
The University of Iowa
zou-qing@uiowa.edu

MS88

Challenges and Opportunities in Out-of-Distribution Detection

The real world is open and full of unknowns, presenting significant challenges for machine learning (ML) systems that must reliably handle diverse, and sometimes anomalous inputs. Out-of-distribution (OOD) uncertainty arises when a machine learning model sees a test-time input that differs from its training data, and thus should not be predicted by the model. As ML is used for more safety-critical domains, the ability to handle out-of-distribution data are central in building open-world learning systems. In this talk, I will talk about challenges, methods, and opportunities on uncovering the unknowns of deep neural networks for reliable predictions in an open world.

Sharon Li
University of Wisconsin-Madison
sharonli@cs.wisc.edu

MS88

A Data-Centric View on Robustness

Over the past few years, researchers have proposed many ways to measure the robustness of machine learning models. In the first part of the talk, we will survey the current robustness landscape based on two large-scale experimental studies involving more than 200 different models and test conditions. Despite the large variety of test conditions, common trends emerge: (i) robustness to natural distribution shift and synthetic perturbations are distinct phenomena, (ii) current algorithmic techniques have little effect on robustness to natural distribution shifts, (iii) training on more diverse datasets offers robustness gains on several natural distribution shifts. In the second part of the talk, we then leverage the aforementioned insights to improve OpenAIs CLIP model. CLIP achieved unprecedented robustness on several natural distribution shifts, but only when used as a zero-shot model. The zero-shot evaluation precludes the use of extra data for fine-tuning and hence leads to lower performance when there is a specific task of interest. To address this issue, we introduce a simple yet effective method for fine-tuning zero-shot models that leads to large robustness gains on several distribution shifts without reducing in-distribution performance.

Ludwig Schmidt
University of Washington
schmidt@cs.washington.edu

MS89

Hypegraph Diffusion Networks for Multi-Modal Data Classification

We address this problem as a multi-modal classification task. In this work, we introduce a novel semi-supervised hypergraph learning framework for Alzheimers disease diagnosis. Firstly, we introduce a dual embedding strategy for constructing a robust hypergraph that preserves the data semantics. This is achieved by enforcing perturbation invariance at the image and graph levels using a contrastive based mechanism. Secondly, we present a dynamically adjusted hypergraph diffusion model, via a semi-explicit flow, to improve the predictive uncertainty. We show, through our experiments, that our technique performs better than existing techniques.

Angelica I. Aviles-Rivero
University of Cambridge
Department of Pure Mathematics and Mathematical Statistics
ai323@cam.ac.uk

Christina Runkel
University of Cambridge, United Kingdom
cr661@cam.ac.uk

Nicolas Papadakis
CNRS/IMB
nicolas.papadakis@math.u-bordeaux.fr

Zoe Kourtzi
Faculty of Psychology
University of Cambridge
zk240@cam.ac.uk

Carola Bibiane Schoenlieb
University of Cambridge, United Kingdom
cbs31@cam.ac.uk

MS89

Lifted Bregman Networks

A combination of gradient-based optimisation method and gradient computation via backpropagation is often the method of choice when it comes to training deep neural networks. In this talk, we focus on alternative training strategies that are based on recent developments in distributed optimisation of deeply nested systems (Carreira-Perpinan Wang 2014). In particular, we propose a novel training approach that is based on proximal activation functions and generalised Bregman distances. The main advantages of this approach compared to previous approaches is that it enables (potentially distributed) training of neural network

parameters without having to differentiate activation functions. This work is joint work with Xiaoyu Wang from the University of Cambridge.

Martin Benning
Queen Mary University of London
m.benning@qmul.ac.uk

MS89

Big Data Inverse Problems

We consider the broad spectrum of linear inverse problems, where the aim is to reconstruct quantities with a sparse representation on some vector space; often solved using the (generalized) least absolute shrinkage and selection operator (lasso). We provide a new perspective on the underlying l1-regularized inverse problem by exploring the generalized lasso problem through variable projection methods. We arrive at our proposed variable projected augmented Lagrangian method. We analyze this method and provide an approach for automatic regularization parameter selection based on a degree of freedom argument. Further, we provide numerical examples demonstrating the computational efficiency for various imaging problems.

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

MS89

Fast Bi-Level Optimization for EPI-MRI Susceptibility Artifact Correction and DTI

We present a machine learning framework for removing susceptibility artifacts that limit the use of Echo Planar Magnetic Resonance Imaging (EPI-MRI), while simultaneously estimating the Diffusion Tensor (DT) that provides information on brain structure and connectivity. While EPI-MRI drastically reduces scan times, the varying magnetic susceptibility in different tissues causes artifacts comprised of geometric transformations and intensity modulations along the phase encoding direction. Correcting the artifacts in a post-processing step is possible by acquiring two diffusion-weighted EPI-MRI sequences acquired with reversed-phase encoding directions and approximately solving a large-scale, non-convex, optimization problem. By including DTI information, this becomes a four-dimensional bi-level optimization between the estimation of the Diffusion Tensor and EPI artifact correction. In this talk, we present a machine learning approach that improves the accuracy and efficiency of the artifact correction while simultaneously solving for the diffusion-weighted tensor information. First, we present a new choice of initialization for the non-convex optimization problem based on optimal mass transport. Second, we provide an implementation in the machine learning framework PyTorch, which provides GPU acceleration and allows us to train hyperparameters such as regularization operators. Third, we solve the bi-level optimization problem for the distortion correction and diffusion tensor.

Abigail Julian
Emory University
abigail.julian@emory.edu

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

MS90

Spectral Convergence of Symmetrized Graph Laplacian on Manifolds with Boundary

In this talk, I will discuss the spectral convergence of a symmetrized Graph Laplacian matrix induced by a Gaussian kernel evaluated on pairs of embedded data sampled from a smooth manifold with boundaries. Theoretically, we deduce the convergence rates for eigenpairs of the discrete Graph-Laplacian matrix to the eigensolutions of the Laplace-Beltrami operator on manifolds with Neumann and Dirichlet boundary conditions. If time permits, I will also discuss an application to a Bayesian inversion problem involving elliptic Dirichlet boundary value problems.

John Harlim
Pennsylvania State University
jharlim@psu.edu

MS90

Robust Inference of Manifold Density and Geometry by Doubly Stochastic Scaling

Many data analysis techniques assume that the data lies on or near a low-dimensional manifold. However, in numerous applications, the data can be corrupted by noise whose magnitude is non-negligible. In this talk, we consider a setting where data points are sampled from a low-dimensional manifold and corrupted by high-dimensional noise that is sufficiently de-localized across the coordinates. The noise is allowed to be quite general, supporting heteroskedastic noise and outliers, and its magnitude can be comparable or even larger than that of the signal. We establish that the doubly stochastic scaling of the Gaussian kernel is robust to such noise, and analyze the scaling factors and the resulting affinity matrix when the dimension and the number of samples are large. Utilizing these results, we can recover the sampling density on the manifold, the pointwise noise magnitudes in the ambient space, and the Laplace-Beltrami operator. In contrast, standard approaches for estimating these quantities can be highly inaccurate in the considered noise regime. We demonstrate our results in simulations as well as on real data.

Boris Landa
Yale University
Yale University
Boris.landa@yale.edu

Xiuyuan Cheng
Department of Mathematics
Duke University
xiuyuan.cheng@duke.edu

MS90

Balancing Density and Geometry in Data with Fermat Distances

Fermat distances (FD) are a class of data-driven metrics based on geodesic paths that penalize passing through low density regions. We show theoretically and empirically how FD interpolate between geometry-based metrics and density-based ones, making them suitable for unsupervised analysis of a wide range of data with intrinsically low-dimensional structure. When FD are defined on discrete data generated as i.i.d. samples from a measure supported

on a Riemannian manifold $\mathcal{M}$, we prove convergence in the large sample limit of the discrete FD to a particular geodesic defined on $\mathcal{M}$. Algorithmically, we provide provably fast algorithms for FD based on Euclidean nearest neighbors graphs, thereby demonstrating the practical applicability of FD to large datasets. The interplay between (1) graphs generated from samples and (2) manifolds on which the sampling measure is supported is emphasized throughout, and several lines of ongoing work and open questions will be discussed.

James Murphy
Tufts University
jmurphy@math.jhu.edu

MS90

A Spectral Algorithm for Joint Clustering and Synchronization

Community detection and orthogonal group synchronization are both fundamental problems with a variety of important applications in science and engineering. In this talk, we consider an extended stochastic block model, where the edge connection contains pairwise orthogonal transformation. We introduce a spectral algorithm to recover the community memberships and perform synchronization simultaneously. The algorithm consists of a spectral decomposition step followed by a blockwise column pivoted QR factorization (CPQR). We also leverage the recently developed 'leave-one-out' technique to establish a near-optimal guarantee for exact recovery of the cluster memberships and stable recovery of the orthogonal transforms.

Zhizhen Zhao
University of Illinois
zhizhenz@illinois.edu

Yifeng Fan
University of Illinois at Urbana-Champaign
yifengf2@illinois.edu

Yuehaw Khoo
University of Chicago
ykhoo@uchicago.edu

MS91

A Constrained Quadratic Optimization Approach to Data-Consistent Inversion

Variations in parameters of a physical model result in variations in quantities of interest that are associated with model outputs. Determining a probabilistic description of the variation in the parameters that results in a specific distribution of the quantities of interest, or observed distribution, is a stochastic inverse problem that can be solved using data-consistent inversion. The data-consistent solution results in a change-of-measure for the "forward" problem and is written in terms of Radon-Nikodym derivatives (densities) of distributions. In practice, these densities are often unknown and must be constructed from independent, identically distributed samples using density-estimation methods. However, density-estimation methods may fail when sufficient samples are not available, e.g., due to high-dimensional data spaces and/or a computationally expensive model. We present an alternative construction of the data-consistent solution based on empirical distribution functions. Previous research shows that computing a set of weights minimizing the L^2 difference between a weighted empirical distribution function and a target distribution function performs a change-of-measure for the "forward" problem. Applying these weights on the parameter space results in an approximation of the data-consistent solution that is not reliant on density estimation methods.

Kirana Bergstrom
University of Colorado - Denver
kirana.bergstrom@ucdenver.edu

Troy Butler
University of Colorado Denver
troy.butler@ucdenver.edu

Tim Wildey
Optimization and Uncertainty Quantification Dept.
Sandia National Laboratories
tmwilde@sandia.gov

MS91

Robust Initializations of Variational Inference with Gaussian Mixtures Through Global Optimization and Laplace Approximations

Mean-field Gaussian distributions are often used as approximate posteriors in variational inference (VI) as their parameters grow linearly with dimension. This can be too restrictive to provide useful approximations of intractable Bayesian posteriors exhibiting multimodal and non-Gaussian behavior. High-fidelity approximations can be obtained with Gaussian mixture models (GMMs) capable of capturing multiple modes. VI with GMMs that possess components with full covariance structure suffers from a quadratic growth in parameters. Coupled with nonconvex trends in the loss functions associated with VI, these challenges motivate the need for robust initialization procedures to improve performance. In this work, we propose a method for constructing an initial GMM approximation that can be used to warm-start VI. The procedure begins with a global optimization stage in which multiple, gradient-based searches determine a set of local maxima, which we take to approximate the mixture component centers. Around each mode, a local Gaussian approximation is constructed via the Laplace approximation. The mixture weights are determined through constrained least squares regression. The procedure is subjected to a variance-based sensitivity analysis to investigate its robustness across various aspects of the underlying posterior distribution. The proposed methodology is also analyzed as an initialization procedure for VI on a high-dimensional, multimodal posterior.

Wyatt H. Bridgman, Reese Jones, Mohammad Khalil
Sandia National Laboratories
whbridg@sandia.gov, rjones@sandia.gov, mkhalil@sandia.gov

MS91

TNet: A Tikhonov Neural Network Approach to Inverse Problems

Deep Learning (DL) by design is purely data-driven and in general does not require physics. This is the strength of DL but also one of its key limitations when applied to science and engineering problems in which underlying physical properties (such as stability, conservation, and positivity) and desired accuracy need to be achieved. DL methods in their original forms are not capable of respecting the underlying mathematical models or achieving desired ac-

curacy even in big-data regimes. On the other hand, many data-driven science and engineering problems, such as inverse problems, typically have limited experimental or observational data, and DL would overfit the data in this case. Leveraging information encoded in the underlying mathematical models, we argue, not only compensates missing information in low data regimes but also provides opportunities to equip DL methods with the underlying physics and hence obtaining higher accuracy. This talk introduces a Tikhonov Network (TNet) that is capable of learning Tikhonov regularized inverse problems. We present and provide intuitions for our formulations for general nonlinear problems. We rigorously show that our TNet approach can learn information encoded in the underlying mathematical models, and thus can produce consistent or equivalent inverse solutions, while naive purely data-based counterparts cannot. Furthermore, we theoretically study the error estimate between TNet and Tikhhonov inverse solutions and under which conditions they are the same. Extension to statistical inverse problems will also be presented.

Tan Bui-Thanh, Hai Nguyen
The University of Texas at Austin
tanbui@ices.utexas.edu, hainguyen@utexas.edu

MS91

RealMVP: A Change of Variables Method For Rectangular Matrix-Vector Products

Rectangular matrix-vector products (MVPs) are used extensively throughout machine learning and are fundamental to neural networks such as multi-layer perceptrons. However, rectangular MVPs are notably missing not used as normalizing flow transforms. This paper identifies this methodological gap and plugs it with a tall and wide MVP change of variables formula. Our theory builds up to a scalable algorithm that envelops existing dimensionality increasing flow methods such as augmented flows (Huang et al., 2020). We show that tall MVPs are closely related to the stochastic inverse of wide MVPs and empirically demonstrate that they improve density estimation over existing dimension changing methods.

Madalina Fiterau
University of Massachusetts, Amherst
mfiterau@cs.umass.edu

Edmond Cunningham
University of Massachusetts Amherst
edmondcunnin@cs.umass.edu

MS92

Creating An Atlas of Approximate Straight-Line Linkages Using Homotopy Continuation and Machine Learning

A classical problem in mechanical engineering is to develop linkages which approximately trace along a straight line. For example, James Watt described such a four-bar linkage in his 1784 patent of the Watt steam engine. This talk will describe an approach which finds approximate straight-line four-bar linkages by computing critical points of an objective function using homotopy continuation. After sampling many approximate straight-line linkages, a dimension reduction algorithm was used to create a visualization of the computed straight-line linkages which provided clusters yielding a visual atlas. This is joint work with Aravind Baskar and Mark Plecnik.

Jonathan Hauenstein
University of Notre Dame
Dept. of App. Comp. Math. & Stats.
hauenstein@nd.edu

MS92

Machine Learning the Dimension of Fano Varieties

In this talk I shall discuss recent work with Tom Coates and Sara Veneziale in which we successfully recover the dimension of a Fano variety X directly from the regularized quantum period of X via machine learning. Fano varieties are basic building blocks in geometry - they are "atomic pieces" of mathematical shapes. Recent progress in the classification of Fano varieties involves analysing the regularized quantum period of a Fano variety X. This is a power series determined by X whose coefficients, which are a sequence of integers, give a numerical signature for X. We apply machine learning to the question: does the quantum period of a Fano variety X know the dimension of X? Note that there is as yet no theoretical understanding of this. We show that machine learning techniques can recover the dimension with ¿ 80% accuracy, demonstrating that machine learning can pick out structure from complex mathematical data in situations where we lack a theoretical understanding. It also gives positive evidence for the assertion (which is proven for smooth varieties in low dimensions, but unknown in general) that the quantum period of a Fano variety determines that variety.

Alexander Kaspryzyk
Nottingham University
a.m.kasprzyk@nottingham.ac.uk

MS92

Machine Learning the Real Discriminant Locus

Parameterized systems of polynomial equations arise in many applications in science and engineering with the real solutions describing, for example, equilibria of a dynamical system, linkages satisfying design constraints, and scene reconstruction in computer vision. Since different parameter values can have a different number of real solutions, the parameter space is decomposed into regions whose boundary forms the real discriminant locus. In this talk, I will discuss a novel sampling method for multidimensional parameter spaces and how it is used in various machine learning algorithms to locate the real discriminant locus as a supervised classification problem, where the classes are the number of real solutions. Examples such as the Kuramoto model will be used to show the efficacy of the methods. Finally, an application to real parameter homotopy methods will be presented. This project is joint work with Edgar Bernal, Jonathan Hauenstein, Dhagash Mehta, and Tingting Tang.

Margaret Regan
Duke University
Durham, NC USA
mregan@math.duke.edu

MS93

Graphical Optimal Transport and Its Applications

Multi-marginal optimal transport (MOT) is a generalization of optimal transport theory to settings with possibly

more than two marginals. The computation of the solutions to MOT problems has been a longstanding challenge. In this talk, we introduce graphical optimal transport, a special class of MOT problems. We consider MOT problems from a probabilistic graphical model perspective and point out an elegant connection between the two when the underlying cost for optimal transport allows a graph structure. In particular, an entropy regularized MOT is equivalent to a Bayesian marginal inference problem for probabilistic graphical models with the additional requirement that some of the marginal distributions are specified. This relation on the one hand extends the optimal transport as well as the probabilistic graphical model theories, and on the other hand leads to fast algorithms for MOT by leveraging the well-developed algorithms in Bayesian inference. We will cover recent developments of graphical optimal transport in theory and algorithms. We will also go over several applications in aggregate filtering and mean field games.

Yongxin Chen
Georgia tech
yongchen@gatech.edu

MS93

K-Mixup, K-Variance and Ot-Based Regularization

Mixup is a popular regularization technique for training deep neural networks that can improve generalization and increase adversarial robustness. It perturbs input training data in the direction of other randomly-chosen instances in the training set. To better leverage the structure of the data, we generalize this to perturbation of sets of k samples via displacement interpolation, calling it k-mixup. Empirical and theoretical evidence shows that the method better preserves cluster and manifold structures and improves generalization and robustness on benchmark datasets. This method connects with k-variance, a recently proposed generalization of variance, that measures the expected cost of matching two sets of k samples from a distribution to each other. In certain regimes, it reflects local rather than global information about the measure. We examine the currently known properties of this generalization for several key cases: one-dimensional measures, clustered measures, and low-dimensional measures. We will also discuss some open problems related to this concept and implications for the k-mixup method.

Edward Chien
Boston University
edchien@bu.edu

MS93

Robust Metrics by Integration

This talk will describe the construction of a certain class of metrics between functions, formed by integration against a class of test functions. These metrics are provably robust to perturbations of the inputs and additive noise, and can be rapidly evaluated. In addition to their theoretical properties, numerical results illustrating the behavior will also be shown.

William Leeb
School of Mathematics
University of Minnesota, Twin Cities
wleeb@umn.edu

MS93

Sinkhorn Distributionally Robust Optimization

We study distributionally robust optimization with Sinkorn distance – a variant of Wasserstein distance based on entropic regularization. We derive convex programming dual reformulations when the nominal distribution is an empirical distribution and a general distribution, respectively. Compared with Wasserstein DRO, it is computationally tractable for a larger class of loss functions, and its worst-case distribution is more reasonable. To solve the dual reformulation, we propose an efficient batch gradient descent with a bisection search algorithm. Finally, we provide various numerical examples using both synthetic and real data to demonstrate its competitive performance.

Yao Xie, Jie Wang
Georgia Institute of Technology
yao.xie@isye.gatech.edu, jwang3163@gatech.edu

MS94

Reinforcement Hedging

TBA

Heekyung Ahn, Kiseop Lee
Purdue University
ahn104@purdue.edu, kiseop@purdue.edu

MS94

Stochastic Differential Models for Ranked and Unranked Principal Eigenportfolios

We model ranked and unranked equity returns data with a system of stochastic differential equations that admit stationary solutions. We show how returns covariances for ranked and unranked data can be related using rank-switching frequency information that can be obtained from the SDEs. We will briefly discuss how this model is used with real equities data. This is joint work with B. Healy, A. Papanicolaou and J. Sasson.

George Papanicolaou
Department of Mathematics
Stanford University
papanicolaou@stanford.edu

MS95

Global Field Reconstruction from Sparse Flow Observations using Deep Learning

An improved understanding of the fundamental processes influencing near-surface wind flow over complex surface features and roughness elements is critical for accurate prediction of wind and dust emission events, and to guide the development of future equipment capable of withstanding extreme environments. A comprehensive exploration of near-surface turbulent flow phenomena involves the investigation of multi-fidelity data obtained from controlled-laboratory (wind tunnel) settings and numerical simulations. The goal of this study is to leverage emerging machine learning (ML) techniques to develop a framework for global reconstruction of system variables from potentially scarce and noisy, multi-fidelity observations. A Voronoi tessellation [Fukami *et al.*, 2021] is used to obtain

a structured-grid representation from sparse sensor observations enabling the use of convolutional neural networks (CNN) for space-time global field estimation. An ensemble-based approach is developed for efficient uncertainty quantification of predicted field values. Numerical results are presented using both benchmark turbulence datasets and wind tunnel observations, comparing the performance of CNN-based models with generative adversarial network (GAN) models.

Adam Collins
US Army Engineer Research and Development Center
adam.m.collins@erdc.dren.mil

Peter Rivera-Casillas
US Army Corps of Engineers ERDC
Vicksburg, MS
peter.g.rivera-casillas@erdc.dren.mil

Andrew Trautz, Orie Cecil
US Army Engineer Research and Development Center
andrew.c.trautz@erdc.dren.mil, orie.m.cecil@erdc.dren.mil

Sourav Dutta
US Army Corps of Engineers ERDC
Vicksburg, MS
sourav.dutta@erdc.dren.mil

Matthew Farthing
US Army Engineer Research and Development Center
matthew.w.farthing@erdc.dren.mil

MS95

Bayesian Neural Networks for Weak Solution of Pdes with Uncertainty Quantification

The solution of partial differential equations (PDEs) is canonical in mathematical physics. However, large scale solutions of PDEs using state of the art discretization techniques remains an expensive proposition. We propose a new physics-constrained neural network (NN) approach to solve PDEs without labels to enable high-throughput solutions in design and decision-making. Distinct from existing physics informed NN approaches, where the strong form or weak form of PDEs are used to construct the loss function, we write the loss function of NNs based on the discretized residual of PDEs through an efficient, convolutional operator-based, and vectorized implementation. We explore an encoder-decoder NN structure for both deterministic and probabilistic models, with Bayesian NNs (BNNs) for the latter, which allow us to incorporate both epistemic uncertainty from the model and aleatoric uncertainty from noise in the data. In our approach, both deterministic and probabilistic convolutional layers are used to learn the applied boundary conditions (BCs) and to detect the problem domain. The trained surrogate PDE solvers can solve for similar physics, but with different BCs and on a number of problem domains, and also extrapolate predictions to BCs that they were not exposed to during training. We demonstrate the performance of the proposed framework on steady-state and equilibrium boundary value problems spanning diffusion, linear elasticity, and nonlinear elasticity.

Krishna Garikipati, Xiaoxuan Zhang
Mechanical Engineering
University of Michigan
krishna@umich.edu, xxzh@umich.edu

MS95

A Model-Constrained Tangent Manifold Learning Approach for Dynamical Systems

Real-time accurate solutions of large-scale complex dynamical systems are in critical need for control, optimization, uncertainty quantification, and decision-making in practical engineering and science applications. This paper contributes in this direction a model-constrained tangent manifold learning (mcTangent) approach. At the heart of mcTangent is the synergy of several desirable strategies: i) a tangent manifold learning to take advantage of the neural network speed and the time-accurate nature of the method of lines; ii) a model-constrained approach to encode the neural network tangent with the underlying governing equations; iii) sequential learning strategies to promote long-time stability and accuracy; and iv) data randomization approach to implicitly enforce the smoothness of the neural network tangent and its likeliness to the truth tangent up second order derivatives in order to further enhance the stability and accuracy of `mcTangent` solutions. Both semi-heuristic and rigorous arguments are provided to analyze and justify the proposed approach. Several numerical results for transport equation, viscous Burger's equation, and Navier-Stokes equation are presented to study and demonstrate the capability of the proposed mcTangent learning approach.

Hai Nguyen, Tan Bui-Thanh
The University of Texas at Austin
hainguyen@utexas.edu, tanbui@ices.utexas.edu

MS96

Two-stage Stochastic Programs with Nonconvex Recourse

We study the decomposition methods for solving a class of nonconvex and nonsmooth two-stage stochastic programs, where both the objective and constraints of the second-stage problem are nonlinearly parameterized by the first-stage variable. Due to the failure of the Clarke-regularity of the resulting nonconvex recourse function, classical decomposition approaches such as Benders decomposition and (augmented) Lagrangian-based algorithms cannot be directly generalized to solve such models. By exploring an implicitly convex-concave structure of the recourse function, we introduce a novel surrogate decomposition framework based on the so-called partial Moreau envelope. Convergence for both fixed scenarios and interior sampling strategy is established. Numerical experiments are conducted to demonstrate the effectiveness of the proposed algorithm.

Ying Cui
University of Minnesota
yingcui@umn.edu

MS96

Lyapunov Stability of the Subgradient Method with Constant Step Size

We consider the subgradient method with constant step size for minimizing locally Lipschitz functions definable in an o-minimal expansion of the real field. In order to analyze the behavior of its iterates in the vicinity of a local minimum, we introduce a notion of discrete Lyapunov sta-

bility. On the one hand, we prove that strict local minima are stable using the theory of differential inclusions. On the other hand, the existence of a Lyapunov function in the neighborhood of a non-strict local minimum guarantees instability. This is true of several spurious local minima arising in robust principal component analysis and deep neural networks. When analyzing the landscape of a nonconvex objective function, it is thus crucial not only to determine whether spurious local minima exist, but also whether they are stable.

Cedric Josz, Lexiao Lai
Columbia University
cj2638@columbia.edu, ll3352@columbia.edu

MS96

The Multi-Objective Polynomial Optimization

The multi-objective optimization is to optimize several objective functions over a common feasible set. Since the objectives usually do not share a common optimizer, people often consider (weakly) Pareto points. This paper studies multi-objective optimization problems that are given by polynomial functions. First, we study the geometry for (weakly) Pareto values and represent Pareto front as the boundary of a convex set. Linear scalarization problems (LSPs) and Chebyshev scalarization problems (CSPs) are typical approaches for getting (weakly) Pareto points. For LSPs, we show how to use tight relaxations to solve them, how to detect existence or nonexistence of proper weights. For CSPs, we show how to solve them by moment relaxations. Moreover, we show how to check if a given point is a (weakly) Pareto point or not and how to detect existence or nonexistence of (weakly) Pareto points. We also study how to detect unboundedness of polynomial optimization, which is used to detect nonexistence of proper weights or (weakly) Pareto points.

Jiawang Nie
University of California, San Diego
njw@math.ucsd.edu

Zi Yang
University of California, Santa Barbara
ziy@ucsb.edu

MS96

Loss Functions for Finite Sets

This talk introduces a novel kind of loss functions for finite sets. For a given finite set S, we give sum-of-square type loss functions of minimum degree. When S is the vertex set of a standard simplex, we show such loss functions have no spurious minimizers (i.e., every local minimizer is a global one). Up to transformations, we give similar loss functions without spurious minimizers for general finite sets. When S is approximately given by a sample set T, we show how to get loss functions by solving a quadratic optimization problem. Numerical experiments and applications are given to show the efficiency of these loss functions.

Jiawang Nie, Suhan Zhong
University of California, San Diego
njw@math.ucsd.edu, suzhong@ucsd.edu

MS97

Signal Processing with the Dirac Operator

Topological signals associated not only to nodes but also to links and to the higher dimensional simplices of simplicial complexes are attracting increasing interest in signal processing, machine learning and network science. Typically, topological signals of a given dimension are investigated and filtered using the corresponding higher-order Laplacian. In this talk, I will introduce the topological Dirac operator that can be used to process simultaneously topological signals of different dimensions. I will discuss the main spectral properties of the Dirac operator defined on networks, simplicial complexes and multiplex networks, and their relation to higher-order Laplacians. Finally I will show how the Dirac operator allows to perform signal processing of coupled topological signals of different dimension.

Ginestra Bianconi
Queen Mary University of London
ginestra.bianconi@gmail.com

MS97

Simplicial Convolutional Neural Networks for Neural Spike Train Decoding

Artificial neural networks have been designed to handle a variety of data-types (e.g., vectors, matrices, graphs) as inputs, but only recently has work been done to allow for the input of simplicial complexes into neural networks for machine learning tasks. Simplicial complexes are topological spaces that use not only vertices and edges, but also higher-dimensional objects like triangles and tetrahedra to generalize graphs and capture more than just pairwise relationships. Thus, they naturally lend themselves to defining neural activity where ensembles of brain cells may work together to encode a single variable. We devise a Simplicial Convolutional Neural Network framework for decoding neural spike train activity. The effectiveness of the framework is then demonstrated by decoding Head Direction cells, which are brain cells that fire when the head is facing a certain direction.

Edward Mitchell
University of Tennessee Knoxville
emitch39@vols.utk.edu

MS97

A New Pipeline for Genomics Data Analysis Using Nonparametric Statistics and Topology

Genomic data has become increasingly prevalent in recent years, particularly due to the recent development of whole transcriptome sequencing (RNA-Seq) data. A number of statistical methods are typically used to reduce the dimensionality of such data and to search for global patterns within the data. Topological data analysis (TDA) is a mathematical approach that can help to gain a better understanding of the intrinsic shape of high-dimensional genomic data and to retain information that may be lost using standard dimension-reduction algorithms. The topological algorithm known as Mapper is frequently used as an exploratory tool to build a graphical representation of data. We propose a workflow to process and analyze RNA-Seq data from tumor and healthy subjects, using Mapper. To

showcase the workflow, we apply it to two distinct RNA-Seq data sets collected in a cross-sectional cancer study in The Cancer Genome Atlas (TCGA), in conjunction with RNA-Seq data from corresponding healthy tissue samples from the Genotype Tissue Expression project (GTEx). We show that our workflow, involving Gaussian mixture approximations of RNA-Seq data can be used to produce graphical structures that successfully separate tumor and healthy subjects, and are robust over parameter variation.

Farzana Nasrin
University of Hawaii
fnasrin@hawaii.edu

Kathleen Storey
Lafayette College
storeyk@lafayette.edu

Erik Amézquita
Michigan State University
amezqui3@msu.edu

Masato Yoshizawa
University of Hawaii
yoshizaw@hawaii.edu

MS97

Correcting the Bias in Laplacian Learning at Low Label Rates

Laplacian learning is a semi-supervised method that finds missing labels via minimising a Dirichlet energy. It is well known that Laplacian learning is (asymptotically) ill-posed at low labelling rates. In this talk I will identify the bias of Laplace learning and show how this can be corrected leading to significant improvement in performance. The correction in the bias leads one to a Poisson equation.

Jeff Calder
University of Minnesota
jcalder@umn.edu

Brendan Cook
University of Minnesota, U.S.
tbd@siam.org

Matthew Thorpe
University of Manchester
matthew.thorpe-2@manchester.ac.uk

Dejan Slepcev
Carnegie Mellon University
slepcev@math.cmu.edu

MS98

Physics-informed Deep Generative Model Based on Variational Autoencoder for Stochastic Systems

Despite of great progress over the last decades in simulating complex problems with the numerical discretization of (stochastic) partial differential equations (PDEs), solving high-dimensional problems governed by parameterized PDEs remains challenging. Machine learning has emerged as a promising alternative in scientific computing community by enforcing the physical laws. We propose physics-informed deep generative model based on variational autoencoder for data-driven stochastic systems. The governing equation is encoded into the model through the differential operator on the decoder. The proposed model is not only memory efficient but also offers huge computational savings. Numerical examples are provided to illustrate the proposed algorithm.

Minseok Choi
POSTECH Mathematics Institute, Korea
mchoi@postech.ac.kr

MS98

Unsupervised Legendre-Galerkin Neural Network for Stiff Partial Differential Equations

Modern machine learning research with deep neural networks has demonstrated outstanding performance in many scientific fields. Recently, machine learning approaches have been widely used for solving differential equations and dynamical systems. As a result, this trend opens up new research known as scientific machine learning. Since neural networks provide an approximation capability for nonlinear functions, the computational parameterization through machine learning and optimization methods is available with enhanced quality when solving various PDEs. In this paper, we provide a novel algorithm in the field of numerical methods for PDEs using techniques from machine learning and artificial intelligence. More precisely, we develop an unsupervised machine learning algorithm based on the Legendre-Galerkin neural network to find an accurate approximation to the solution of different types of the PDEs. The neural network is applied to not only the general 1D and 2D differential equations but also singularly perturbed PDEs, which possess boundary layer behavior.

YoungJoon Hong
Sungkyunkwan University
hongyj@g.skku.edu

MS98

A Neural Network Approach for Homogenization of Multiscale Problems

We propose a neural network-based approach to the homogenization of multiscale problems. The proposed method uses a derivative-free formulation of a training loss, which incorporates Brownian walkers to find the macroscopic description of a multiscale PDE solution. Compared with other network-based approaches for multiscale problems, the proposed method is free from the design of handcrafted neural network architecture and the cell problem to calculate the homogenization coefficient. The exploration neighborhood of the Brownian walkers affects the overall learning trajectory. We determine the bounds of micro- and macro-time steps that capture the local heterogeneous and global homogeneous solution behaviors, respectively, through a neural network. The bounds imply that the computational cost of the proposed method is independent of the microscale periodic structure for the standard periodic problems. We validate the efficiency and robustness of the proposed method through a suite of linear and nonlinear multiscale problems with periodic and random field coefficients.

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

Jihun Han
Dartmouth College

jihun.han@dartmouth.edu

MS98

A Norm-Minimization Based Training Method for Operator Networks

We propose a novel training method for operator networks (ONets). ONets are constructed by a sum of products two sub-networks, namely, branch and trunk networks. The goal is to effectively learn ONets that accurately approximate nonlinear operators from data. The standard approaches train the two sub-networks simultaneously via first-order optimization methods. The proposed method, however, trains trunk networks first based on norm-minimization, and then train branch networks in sequence. The norm-minimization provides a natural way of handling (sparse) noise in training data, as the user has flexibility to choose a suitable norm, e.g., l_1 norm. Several numerical examples including Darcy flow in heterogeneous porous media with and without noise are presented to illustrate the capabilities of the proposed training method.

Sanghyun Lee
Florida State University
slee17@fsu.edu

Yeonjong Shin
KAIST, Korea
yeonjong_shin@kaist.edu

MS99

Learning Stochastic Closures for Climate Models from Indirect Data Using Kalman Methods

Climate models are tools for learning about and predicting the climate response to increasing greenhouse gas emissions. They simulate a wide range of scales, spanning large-scale circulations to small-scale turbulence in clouds. The turbulent entrainment of environmental air into clouds and the detrainment of cloud air into their environment are among the most important uncertain processes in climate models. The lack of scale-separation between the turbulent entrainment and detrainment and larger-scale motions makes it difficult to justify traditional deterministic closures. To remedy these deficiencies, we construct a stochastic closure that complements existing physically motivated closures for entrainment and detrainment in an eddy-diffusivity mass-flux scheme for atmospheric turbulence and convection. To learn parameters of this stochastic closure model, we employ gradient-free Kalman inversion methods, which allow the model to be trained from indirect data such as time-averaged statistics. The performance of the proposed closure is then investigated on a series of test cases, with emphasis on improving statistical fits and characterizing predictive uncertainty. More broadly, this work represents a step toward designing climate models that provide better estimates for the uncertainty of future warming in response to increasing greenhouse gas emissions.

Haakon Ludvig Langeland Ervik, Tapio Schneider
California Institute of Technology
haakon@caltech.edu, tapio@caltech.edu

Andrew Stuart
Computing + Mathematical Sciences
California Institute of Technology
astuart@caltech.edu

Jinlong Wu
California Institute of Technology
jinlong@caltech.edu

MS99

Structure-Exploiting Projected Deep Neural Networks for Bayesian Inversion and Optimal Experimental Design

Solution of Bayesian inverse problems (BIPs) governed by large-scale complex models in high parameter dimensions (such as PDEs with discretized infinite-dimensional parameter fields) is often prohibitive. Efficient evaluation of the parameter-to-observable (p2o) map, which involves solution of the forward model, is key to making BIPs tractable. Surrogate approximations of p2o maps have the potential to greatly accelerate BIPs, provided the p2o map can be accurately approximated using relatively fewer forward model solves. Unfortunately, constructing such surrogates presents significant challenges when the parameter dimension is high and the forward model is expensive. Deep neural networks (DNNs) have emerged as leading contenders for overcoming these challenges. We demonstrate that black box application of DNNs for problems with infinite dimensional parameter fields leads to poor generalization when training data are limited in number. However, by constructing a network architecture that is adapted to the geometry and intrinsic low-dimensionality of the p2o map as revealed by adjoint-based Hessian actions, one can construct a parsimonious projected DNN surrogate with superior approximation properties using only limited training data. We employ this projected DNN surrogate to solve Bayesian optimal experimental design problems for finding sensor locations that maximize the expected information gain for inverse wave scattering.

Keyi Wu
University of Texas Austin
keyiwu@math.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, Omar Ghattas
The University of Texas at Austin
peng@oden.utexas.edu, omar@oden.utexas.edu

MS99

Learning Physics-Constrained, Transferable Subgrid-Scale Closure Models in the Small-Data Regime

Simulating high-dimensional, multi-scale systems often requires subgrid-scale (SGS) models to represent the effects of small-scale processes in a computationally affordable low-resolution solver. Data-driven SGS models, which are learned using high-resolution data, e.g., using deep neural networks, have shown promising results in recent years. However, the availability of such high-resolution data is often limited, i.e., training should be done in the small-data regime. Here, we use 2D turbulence as the test case and show that 1) by incorporating physics in the learning process, stable and accurate SGS models can be learned in the small-data regime, and 2) transfer learning enables out-of-distribution generalization (e.g., extrapolation to 100x higher Reynolds numbers). We assess the quality of the

data-driven SGS models using a number of a priori and a posteriori tests.

Pedram Hassanzadeh
Mechanical Engineering and Earth Science
Rice University
pedram@rice.edu

Yifei Guan, Ashesh K. Chattopadhyay
Rice University
yg62@rice.edu, akc6@rice.edu

Adam Subel
NYU
as15415@nyu.edu

MS99

Deep Uncertainty Quantification for Learning Spatiotemporal Dynamics

Applications such as public health, transportation, and climate science often require learning complex dynamics from large-scale spatiotemporal data. While deep learning has shown tremendous success in these domains, prior works have mostly focused on point estimates without quantifying the uncertainty of the predictions. In high-stakes domains, being able to generate probabilistic predictions with confidence intervals is critical to risk assessment and decision making. In this talk, I will present our efforts in uncertainty quantification (UQ) in learning spatiotemporal dynamics. I will discuss (1) a systematic study of UQ for deep spatiotemporal forecasting. We analyze UQ methods from both the Bayesian and the frequentist point of view, casting in a unified framework. We perform benchmark tests on COVID-19, traffic forecasting and air quality prediction tasks. (2) Spatiotemporal Neural Processes (STNP), a neural latent variable model to mimic the spatiotemporal dynamics of stochastic simulators, together with Bayesian active learning strategy to speed up training. We demonstrate our method on the use cases of COVID-19 forecasting and scenario creation.

Dongxia Wu
University of California San Diego
dowu@ucsd.edu

Rose Yu
University of California, San Diego, U.S.
roseyu@ucsd.edu

Yian Ma
University of California San Diego
Halicioglu Data Science Institute
yianma@ucsd.edu

Matteo Chinazzi
Northeastern University
m.chinazzi@northeastern.edu

Alessandro Vespignani
Northeastern University, U.S.
a.vespignani@northeastern.edu

Xinyue Xiong
Northeastern University
xiong.xin@northeastern.edu

Liyao Gao
Universty of Washington
marsgao@uw.edu

MS100

The Statistical Power of Weak Form Sparse Identification of Nonlinear Dynamics (WSINDy)

In recent years, data-driven modeling to infer governing equations directly from data has received considerable attention. In particular, the Sparse Identification of Nonlinear Dynamics (SINDy) has been widely used. While SINDy is highly appealing for its abilities, in its original form, it is know to perform poorly in the presence of noisy data. In many cases, 0.5% additive error will result in catastrophically incorrect learned equations. Within the last 3 years, the consensus has emerged that learning a weak form of the equation offers a superior strategy for learning equations in the presence of noise. In particular, the WSINDy algorithm can (in many cases) learn equations in the presence of greater than 100% noise. In this talk, we will discuss computational properties of the method as well as best practices for its implementation. Time permitting, we will illustrate the method with several examples.

David M. Bortz
University of Colorado
Department of Applied Mathematics
dmbortz@colorado.edu

Daniel A. Messenger
University of Colorado Boulder
daniel.messenger@colorado.edu

MS100

Bayesian Nonlocal Operator Regression (BNOR): A Data-Driven Learning Framework of Nonlocal Models with Uncertainty Quantification

We consider the problem of modeling heterogeneous materials where small-scale dynamics and interactions affect the global behavior. The material's microstructure, properties and interfacial conditions cause variability in the material's response; hence it is often non-practical, if not impossible, to provide quantitative characterization for each sample. The goal of this work is to develop a Bayesian framework to characterize the uncertainty of material response when using a nonlocal model to describe wave propagation through heterogeneous materials. Our approach is based on the nonlocal operator regression (NOR) technique, and Bayesian inference. Specifically, we use a Markov chain Monte Carlo (MCMC) method to predict the probability distribution of the nonlocal constitutive law that embeds the material's properties. As an application, we consider the wave propagation problem in a heterogeneous bar with randomly generated microstructure layers. In particular, we apply the proposed approach to model the stress wave propagation and provide a characterization of the uncertainty originated from the material microstructure. With several numerical tests, we illustrate the effectiveness of our approach in predicting the posterior distribution of the optimal nonlocal model and characterizing its uncertainty in further prediction tasks.

Yiming Fan
Lehigh University

yif319@lehigh.edu

MS100

Asymptotic Theory of L1-Regularized PDE Identification from Single Noisy Trajectory

We provide a formal theoretical analysis on the PDE identification via the ℓ_1-regularized pseudo least square method from the statistical point of view. In this article, we assume that the differential equation governing the dynamic system can be represented as a linear combination of various linear and nonlinear differential terms. Under noisy observations, we employ local-polynomial fitting for estimating state variables and apply the ℓ_1-penalty for model selection. Our theory proves that the classical mutual incoherence condition on the feature matrix F and the $\beta^\star_{\min}$-condition for the ground-truth signal $\beta^\star$ are sufficient for the signed-support recovery of the ℓ_1-PsLS method. We run numerical experiments on two popular PDE models, the viscous Burgers and the Kortewegde Vries(KdV) equations, and the results from the experiments corroborate our theoretical predictions.

Namjoon Suh
Georgia Institute of Technology
namjsuh@gatech.edu

MS101

Learning from Sparse, Mixed, And/or Multi-Fidelity Data via Latent Map Gaussian Processes

Materials modeling and design is typically hampered by two major uncertainty sources: lack of data and model form discrepancies. In this talk, we will present a novel approach based on nonlinear manifold learning that addresses these (and more) uncertainty sources. Our approach is based on latent map Gaussian processes (LMGPs) and aims to leverage multiple data sources to quantify uncertainty sources and, more importantly, provide visually interpretable diagnostic measures that indicate the extent to which different data sources (e.g., experiments, simulations, etc.) agree with one another. We will introduce our approach in both Bayesian and frequentist settings and demonstrate that both perform well in a wide range of applications.

Ramin Bostanabad
Northwestern University
raminb@uci.edu

MS101

Bayesian Inference with Latent Hamiltonian Neural Networks (L-HNNs)

When sampling for Bayesian inference, one popular approach is to use Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS). However, HMC and NUTS can require numerous numerical gradients of the target density and can prove slow in practice. We propose Hamiltonian neural networks (HNNs) with HMC and NUTS for solving Bayesian inference problems. Once trained, HNNs do not require gradients of the target density while sampling. Moreover, they satisfy important properties such as perfect time reversibility and Hamiltonian conservation, making them well suited for use within HMC and NUTS because stationarity can be shown. We also propose an HNN extension called latent HNNs (L-HNNs), which predict latent variable outputs. Compared to HNNs, L-HNNs offer improved expressivity and a reduction in integration errors. Finally, we propose employing L-HNNs in NUTS with an online error monitoring scheme to prevent degeneracy of the sampling in regions of low probability density. We demonstrate L-HNNs in NUTS with online error monitoring by using several example cases involving complex, heavy-tailed, and high local curvature probability densities. Overall, L-HNNs in NUTS with online error monitoring satisfactorily inferred these probability densities. Compared to traditional NUTS, L-HNNs in NUTS with online error monitoring improved the effective sample size (ESS) per gradient by an order of magnitude.

Som Dhulipala
TBD
Som.Dhulipala@inl.gov

Yifeng Che
Idaho National Laboratory
yifeng.che@inl.gov

Michael D. Shields
Johns Hopkins University
michael.shields@jhu.edu

MS101

Learning to Accelerate Large-Scale Physical Simulations in Fluid and Plasma Physics

Simulating the time evolution of large-scale physical systems is crucial in many scientific and engineering domains. Recently, deep learning surrogate modeling has emerged as a useful tool to speed up the compute-intensive domain-specific classical solvers, but scaling it to simulate larger systems still faces several challenges: stable long-term evolution and modeling multi-scale dynamics. In this talk, I will present our works in two important domains to address these challenges. In a subsurface fluid system that consists of millions of cells per time step, we introduce a Hybrid Graph Network Simulator to model the complex and heterogeneous dynamics in the subsurface. We introduce a multi-step based objective to improve its long-term rollout, and sector-based training to make training such system possible. We demonstrate that our method is able to scale up to millions of cells per time step, and achieves up to 18x speedup compared to the classical solvers. In a laser-plasma interaction system whose dynamics is multi-scale, we introduce a hybrid particle + fluid method, which uses the classical solver to model kinetics of the few but highly energetic particles, and CNNs to model the evolution of the moments of the dominant thermal particles in each cell, and their injection from fluid state to particle state. We show that this significantly speeds up the simulation.

Tailin Wu
Stanford University
tailin@cs.stanford.edu

MS102

Lifting the Curse of Dimensionality in Nonlinear Operator Learning with Operator Adapted GPs and Wavelets

We consider the problem of learning the mapping of the right-hand side of a nonlinear PDE (possibly with rough coefficients) to the solution from examples. For linear PDEs, when represented in a basis of operator-adapted wavelet, the solution operator has a sparse structure across

scales and across space, enabling approximation without the curse of dimensionality. We present a method to learn the solution operator of nonlinear PDEs using kernel methods by encoding the localization of the problem in scales and space into operator-valued kernels. The kernels are designed in such a way to take advantage of the near sparsity of the solution operator in appropriate operator-adapted wavelet bases. This allows us to lift the curse of dimensionality, much like the case of linear PDEs, when learning the solution operator of nonlinear PDEs. This is in contrast to existing methods in the literature that suffer from the curse of dimensionality by learning the solution map directly in the physical or frequency domains.

Pau Batlle
Caltech
pbatllef@caltech.edu

Houman Owhadi
California Institute of Technology
owhadi@caltech.edu

Bamdad Hosseini
University of Washington
bamdadh@uw.edu

Matthieu Darcy
Caltech
mdarcy@caltech.edu

MS102

Gaussian Processes for Solving and Learning Nonlinear PDEs

Partial differential equations (PDEs) are fundamental in today's applied and computational science. This talk presents a conceptually simple and algorithmically automatic framework for solving, learning, and quantifying uncertainties of these PDEs. The framework assigns Gaussian processes (GPs) priors to the unknowns and finite-dimensionalizes the PDE into a generally nonlinear combination of linear differential measurements of the GPs at sampled collocation points. Performing statistical inference or machine learning on these GPs then addresses the computational problems associated with the PDEs. We will present the theory and numerical experiments of this method in a wide range of nonlinear PDE problems. A near-linear complexity multiscale implementation for computing with large kernel matrices will also be discussed.

Yifan Chen, Pau Batlle
Caltech
yifanc@caltech.edu, pbatllef@caltech.edu

Bamdad Hosseini
University of Washington
bamdadh@uw.edu

MS102

Sparse Cholesky Factorization by Greedy Conditional Selection

Dense kernel matrices resulting from pointwise evaluations of a kernel function arise naturally in machine learning and statistics. Previous work in constructing sparse transport maps or sparse approximate inverse Cholesky factors for such matrices by minimizing Kullback-Leibler divergence recovers the Vecchia approximation for Gaussian processes. However, this method for Cholesky factorization relies only on geometry to construct the sparsity pattern, ignoring the conditional effect of adding an entry. In this work, we construct the sparsity pattern by leveraging a greedy selection algorithm which selects points that maximize mutual information with target points, conditional on all points selected previously. For selecting k points out of N, the naive time complexity is $\mathcal{O}(Nk^4)$, but by maintaining a partial Cholesky factor we reduce this to $\mathcal{O}(Nk^2)$. Furthermore, for multiple (m) targets we achieve a time complexity of $\mathcal{O}(Nk^2 + Nm^2 + m^3)$ which is maintained in the setting of Cholesky factorization where a selected point need not condition every target. We directly apply the selection algorithm to image classification, Gaussian process regression, and recovery of sparse Cholesky factors, improving upon k-th nearest neighbors in every case. Through Kullback-Leibler minimization we apply the algorithm to Cholesky factorization, improving in high dimensional geometries as well as when preconditioning with the conjugate gradient.

Stephen Huan
Georgia Tech
shuan@gatech.edu

Joe Guinness
Cornell University
guinness@cornell.edu

Matthias Katzfuss
Texas A&M University
katzfuss@gmail.com

Houman Owhadi
Applied Mathematics
Caltech
owhadi@caltech.edu

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

MS102

Finite Element and Graphical Representations of Gaussian Processes

Gaussian processes (GPs) are popular models for random functions in computational and applied mathematics, statistics, machine learning and data science. However, GP methodology scales poorly to large data-sets due to the need to factorize a dense covariance matrix. In spatial statistics, a standard approach to surmount this challenge is to represent Matrn GPs using finite elements, obtaining an approximation with a sparse precision matrix. The first part of the talk will give new understanding of this approach for regression and classification with large data-sets, showing that under mild smoothness assumptions the dimension of the matrices that need to be factorized can be reduced without hindering the estimation accuracy. The analysis balances finite element and statistical errors to show that there is a threshold beyond which further refining of the discretization increases the computational cost without improving the estimation accuracy. In the second part of the talk, I will introduce graphical representations of GPs to model random functions on high-dimensional point clouds, greatly expanding the important but limited scope of the finite element approach. I will show error bounds on the graphical representations, and study the associated posterior contraction in a semi-supervised learning

problem.

Daniel Sanz-Alonso, Ruiyi Yang
University of Chicago
sanzalonso@uchicago.edu, yry@uchicago.edu

MS103

On the Complexity of Hypermatrix Equivalence

Equivalence of matrices defined relative to matrix group actions are crucial to many practical numerical matrix algorithms. In this talk we discuss how equivalence classes of hypermatrices defined relative to hypermatrix analog of group actions can serve as the basis for measuring the complexity of hypermatrix algorithms.

Edinah Gnang
Johns Hopkins University
egnang1@jhu.edu

MS103

Approximate Tensor BM Product Decomposition for Temporal Analysis of Third-Order Data

Given matrices A, B, C of size $m \times n$, $m \times p$, and $p \times n$, respectively, and oriented as third order tensors in the appropriate dimensions, their Bhattacharya-Mesner (BM) product will result in a third order tensor of dimension $m \times p \times n$ and BM-rank of 1. In this talk, we give an iterative algorithm for approximately decomposing a third order tensor into a small sum of BM-rank 1 tensors. We provide some analysis as to the features captured by this decomposition and give its relation to other tensor methods. Moreover, we demonstrate the ability of our decomposition to extract important temporal information from third order blood platelet data and video image data.

Fan Tian
Tufts University
fan.tian@tufts.edu

Misha E. Kilmer
Mathematics Department
Tufts University
misha.kilmer@tufts.edu

Eric L. Miller, Abani Patra
Tufts University
elmiller@ece.tufts.edu, abani.patra@tufts.edu

Anna Konstorum
Department of Pathology
Yale School of Medicine
anna.konstorum@yale.edu

MS103

Practical Leverage-Based Sampling for Low-Rank Tensor Decomposition

The low-rank canonical polyadic tensor decomposition is useful in data analysis and can be computed by solving a sequence of overdetermined least squares subproblems. Motivated by consideration of sparse tensors, we propose sketching each subproblem using leverage scores to select a subset of the rows, with probabilistic guarantees on the solution accuracy. We randomly sample rows proportional to leverage score upper bounds that can be efficiently computed using the special Khatri-Rao subproblem structure inherent in tensor decomposition. Crucially, for a $(d+1)$-way tensor, the number of rows in the sketched system is $O(r^d/\epsilon)$ for a decomposition of rank r and ϵ-accuracy in the least squares solve, independent of both the size and the number of nonzeros in the tensor. Along the way, we provide a practical solution to the generic matrix sketching problem of sampling overabundance for high-leverage-score rows, proposing to include such rows deterministically and combine repeated samples in the sketched system; we conjecture that this can lead to improved theoretical bounds. Numerical results on real-world large-scale tensors show the method is significantly faster than deterministic methods at nearly the same level of accuracy.

Brett Larsen
Stanford University
bwlarsen@stanford.edu

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

MS103

Tensor Learning for Large-Scale Spatiotemporal Data

Multivariate spatiotemporal data is ubiquitous in science and engineering, from sports analytics to neuroscience. Such data can be naturally represented as a multiway tensor. Tensor latent factor models provide a powerful tool for reducing the dimensionality and discovering the higher-order latent structures from data. However, existing tensor models are often slow or fail to yield latent factors that are easy to interpret by domain experts. In this talk, I will demonstrate advances in tensor methods to generate interpretable latent factors for high-dimensional spatiotemporal data. In particular, I will discuss (1) a multiresolution tensor learning algorithm, that can leverage the multiscale property of high-resolution spatial data, to speed up training and learn interpretable patterns. (2) a tensor latent feature learning algorithm that can learn binary representations of data that are both memory efficient and easy to interpret. We provide theoretical guarantees for our optimization algorithms and demonstrate their applications to real-world data from basketball plays and neuroscience.

Rose Yu
University of California, San Diego, U.S.
roseyu@ucsd.edu

MS104

Multiscale DNN for Oscillatory Navier-Stokes Flows and Causality in DNN for Dynamic Systems

We present linearized learning methods to accelerate the convergence of trainings for a stationary nonlinear Navier-Stokes (NS) equation. Different linearizations of the nonlinear convection term in the NS equation are integrated into the training process of a multi-scale deep neural network approximation of oscillatory NS solutions in complex domains. Also, we will present a DeepONet structure with built-in causality to represent linear operators between Banach spaces of time-dependent signals. The theorem of Universal Approximations to nonlinear operators is extended to operators with causalities and the proposed Causality-DeepONet framework embodies the causality naturally. As an application, we utilize the Causality-DeepONet to learn accurately the retarded operator that maps seismic excitations to the responses of a building at

different locations.

Lizuo Liu
Southern Methodist University
lizuol@mail.smu.edu

MS104

Physics-Informed Neural Networks with Hard Constraints for Inverse Design

Inverse design arises in a variety of areas in engineering such as acoustic, mechanics, thermal/electronic transport, electromagnetism, and optics. Topology optimization is an important form of inverse design, where one optimizes a designed geometry to achieve targeted properties parameterized by the materials at every point in a design region. This optimization is challenging, because it has a very high dimensionality and is usually constrained by partial differential equations (PDEs) and additional inequalities. Here, we propose a new deep learning method—physics-informed neural networks with hard constraints (hPINNs)—for solving topology optimization. hPINN leverages the recent development of PINNs for solving PDEs, and thus does not require a large dataset (generated by numerical PDE solvers) for training. However, all the constraints in PINNs are soft constraints, and hence we impose hard constraints by using the penalty method and the augmented Lagrangian method. We demonstrate the effectiveness of hPINN for a holography problem in optics and a fluid problem of Stokes flow. We achieve the same objective as conventional PDE-constrained optimization methods based on adjoint methods and numerical PDE solvers, but find that the design obtained from hPINN is often smoother for problems whose solution is not unique. Moreover, the implementation of inverse design with hPINN can be easier than that of conventional methods because it exploits the extensive deep-learning software infrastructure.

Lu Lu
University of Pennsylvania
lulu1@seas.upenn.edu

MS104

Active Neuron Least Squares: A Training Method for Multivariate Rectified Neural Networks

In this talk, we will present the Active Neuron Least Squares (ANLS), an efficient training algorithm for neural networks (NNs). ANLS is designed from the insight gained from the analysis of gradient descent training of NNs, particularly, the analysis of Plateau Phenomenon. The core mechanism is the option to perform the explicit adjustment of the activation pattern at each step, which is designed to enable a quick exit from a plateau. The performance of ANLS will be demonstrated and compared with existing popular methods in various learning tasks ranging from function approximation to solving PDEs.

Yeonjong Shin
KAIST, Korea
yeonjong_shin@kaist.edu

MS104

Heavy-Ball Neural ODEs

We will present a new class of continuous-depth deep neural networks that were motivated by the ODE limit of the classical momentum method, named heavy-ball neural ODEs (HBNODEs). HBNODEs enjoy two properties that imply practical advantages over NODEs: (i) The adjoint state of an HBNODE also satisfies an HBNODE, accelerating both forward and backward ODE solvers, thus significantly accelerate learning and improve the utility of the trained models. (ii) The spectrum of HBNODEs is well structured, enabling effective learning of long-term dependencies from complex sequential data.

Bao Wang
Data Science Research Lab in the DEPARTMENT OF MATHEMATICS a
bwang@math.utah.edu

MS105

Hybrid Projection Methods for Solution Decomposition in Large-Scale Bayesian Inverse Problems

In this work, we develop hybrid iterative projection methods for large-scale inverse problems, where the solution represents a sum of different stochastic components. Such scenarios arise in many imaging applications (e.g., anomaly detection in atmospheric emissions tomography) where the reconstructed image can be represented as a combination of two or more images and each image contains different smoothness or stochastic properties. In an inversion or inverse modeling framework, these assumptions correspond to different regularization terms for each image in the sum, although the desired image is the combination of images. We describe hybrid projection methods for computing solution estimates that are based on a combined flexible and generalized Golub-Kahan process and provide numerical examples from atmospheric imaging.

Julianne Chung
Department of Mathematics
Virginia Tech
jmchung@emory.edu

Jiahua Jiang
University of Birmingham
j.jiang.3@bham.ac.uk

Scot Miller
Johns Hopkins University
smill191@jhu.edu

Arvind Saibaba
North Carolina State University
asaibab@ncsu.edu

MS105

Adaptive Modular-Proximal Gradient Algorithms in $L^{p(\cdot)}$ for Imaging Problems

Solving inverse problems with a variational approach often leads to the minimization of structured functionals, defined in terms of the sum of a smooth and convex function (the data-fidelity term), and a proper, l.s.c., convex (typically non-smooth) one (the regularization term). We study optimization problems in this form, defined on non-standard variable exponent Lebesgue spaces $L^{p(\cdot)}$. The pointwise variable exponent map induces a specific shift-variant norm and intrinsic space-variant properties, which makes interesting the choice of the Banach space $L^{p(\cdot)}$ as a solution space, as, implicitly, it favors a locally varying smoothness of the solution and enhanced sparsity-promoting properties. For the algorithmic optimization, we propose a prox-

imal gradient algorithm where the proximal step, rather than depending on the natural but non-separable $L^{p(\cdot)}$ norm, is defined in terms of its separable modular function, which allows for the efficient computation of algorithmic iterates, and by means of a Bregman-like distance. We analyze the algorithms convergence in function values, showing its dependence on problem/space smoothness. Some numerical tests highlighting the flexibility of the modeling and the effectiveness of the proposed solution schemes for exemplar deconvolution and mixed noise removal problems are presented.

Marta Lazzaretti
University of Genova
Université Cote d'Azur
lazzaretti@dima.unige.it

Luca Calatroni
CNRS, UCA, INRIA
calatroni@i3s.unice.fr

Claudio Estatico
Department of Mathematics, University of Genova, Italy
estatico@dima.unige.it

MS105

Dual Forward-Backward Unfolded Network for Flexible Plug-and-Play for Image Recovery

Proximal methods have been extensively used to find maximum a posteriori (MAP) estimates of unknown images from degraded measurement. Recently, they have been mixed with neural networks (NN) to further improve the reconstruction quality. Two approaches can be distinguished: unfolded NNs, implementing a given iteration number of an optimisation algorithm, and plug-and-play (PnP) algorithms, incorporating NNs in existing optimisation algorithms. Unfolded NNs usually incorporate the measurement operator in the learning process, which can be prohibitive for applications with non-fixed measurement operators. PnP do not have this drawback, but involved NNs still depend on the underlying statistical models (e.g., higher noise level on the measurements requires stronger denoisers). In this work, we propose a PnP algorithm based on forward-backward (FB) iterations, where the learned denoiser is an unfolded NN based on dual-FB iterations. This NN is built to mimic a Gaussian denoiser from a MAP viewpoint. This allows us to introduce a regularisation parameter in the model to tune the regularization strength, similarly to standard variational approaches. This has the advantage of making the learned NN more adaptive to a variety of inverse problem statistical models, without requiring to train the NN for different noise levels.

Audrey Repetti, Matthieu Terris
Heriot-Watt University
a.repetti@hw.ac.uk, mt114@hw.ac.uk

Jean-Christophe Pesquet
Centrale Supélec - Paris
jean-christophe@pesquet.eu

MS105

Neural Fixed-Point Network for Some Inverse Problems

A growing trend in deep learning replaces fixed depth models with approximations of the limit as network depth approaches infinity. This approach uses a portion of network weights to prescribe behavior by defining a limit condition. This makes network depth implicit, varying based on the provided data and error tolerance. Moreover, existing implicit models can be implemented and trained with fixed memory costs in exchange for additional computational costs. In particular, backpropagation through implicit depth models requires solving a Jacobian-based equation arising from the implicit function theorem. This talk first presents fixed-point networks (FPNs) with a new Jacobian-free backpropagation (JFB) scheme that circumvents the need to solve Jacobian-based equations. This makes FPNs much cheaper to train and easy to implement. Then, we describe its applications in imaging and game inverse problems.

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

Howard Heaton
University of California, Los Angeles
hheaton@ucla.edu

Qiuwei Li
Alibaba Group US, DAMO Academy
li.qiuwei@alibaba-inc.com

Daniel McKenzie
University of California, Los Angeles, U.S.
mckenzie@math.ucla.edu

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

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

MS106

Learning Closures of Dynamical Systems with Quantum Mechanics

When modelling chaotic dynamical systems, notably in climate prediction models, a common obstacle is the issue of approximating the dynamics of dimensions of the state space for which dynamical laws are unknown. Common approaches to solving this issue include parametrization schemes which employ stochasticity or which are exclusively dependent on the internal state of the system at a given time. We propose a novel parametrization scheme based on the mathematical framework of quantum mechanics and Koopman operator theory. Given a system in which some components of the state are unknown, this method involves defining the surrogate system as being in a time-dependent quantum-state which influences the choice of the unknown component of the classical state at each timestep. The quantum state is an operator on the space of classical observables and evolves over time under an action by the Koopman operator. The quantum state also updates with new data-points according to a quantum Bayes law, and evolves under the action of data-driven operators. Kernel functions are utilized to allow the quantum Bayes law to be implemented numerically. We analyze the results of two different modalities of this methodology applied to the Lorenz 63 and Lorenz 96 systems, and show how this approach preserves important statistical and qual-

itative properties of the underlying chaotic systems.

David Freeman, Dimitrios Giannakis
Dartmouth College
david.c.freeman.gr@dartmouth.edu, Dimitrios.Giannakis@dartmouth.edu

MS106

Data Assimilation in Non-Abelian Operator Algebras

We present an algebraic framework for sequential data assimilation of partially observed dynamical systems is developed. In this framework, the Bayesian formulation of data assimilation is embedded in a non-abelian operator algebra, which provides a representation of observables by multiplication operators and probability densities by density operators (quantum states). In the algebraic approach, the forecast step of data assimilation is represented by a quantum channel induced by the Koopman operator of the dynamical system. Moreover, the analysis step is described by a quantum effect, which generalizes the Bayesian observational update rule. Projecting the infinite-dimensional operator algebra to a finite-dimensional matrix algebra leads to new computational data assimilation schemes, which are positivity-preserving and have consistent data-driven implementations using kernel methods for machine learning. We illustrate these methods with applications to data assimilation of the Lorenz 96 multiscale system and the El Nino Southern Oscillation of the climate system.

Dimitrios Giannakis
Dartmouth College
Dimitrios.Giannakis@dartmouth.edu

Joanna M. Slawinska
Finnish Center for Artificial Intelligence
Department of Computer Science, University of Helsinki
joanna.m.slawinska@dartmouth.edu

David Freeman
Dartmouth College
david.c.freeman.gr@dartmouth.edu

MS106

Radial Basis Function Interpolation for Manifold Learning: Revisited

In the past, algorithms involving radial basis functions have been used to estimate the Laplace Beltrami operator; however, many such results take place in a deterministic setting. The purpose of this talk is to revisit such a method of approximating the Laplace Beltrami operator and adapt it to the probabilistic setting. First, we present new results regarding the spectral convergence of an estimator which makes use of radial basis function interpolation. We then discuss numerical results demonstrating this method on some examples. Finally, we adapt this construction to prove analogous results for other Laplacians, such as the connection Laplacian. This is joint work with John Harlim and Shixiao Jiang.

Wilson Peoples
Pennsylvania State University
jwp5828@psu.edu

MS106

Spectral Approximation of Linear Derivations

With the advent of data-driven methods such as Diffusion maps, spectral approximation has become a useful tool in dynamical systems. For example, the Spectral Exterior Calculus (Berry, Giannakis 2020) is a framework whereby many constructions in classical Riemannian geometry are represented in terms of Laplacian eigenfunctions. This approach has several advantages, one of which being the decoupling of memory requirements from the number of data points. In this work, we adopt a similar approach and vector fields on Riemannian manifolds via approximation of C^∞-derivations.

Ryan Vaughn
New York University
rv2152@nyu.edu

Dimitrios Giannakis
Dartmouth College
Dimitrios.Giannakis@dartmouth.edu

MS107

Geometric Ergodicity of Slice Sampling

For approximate sampling of a partially known distribution, e.g., posterior distributions in Bayesian inverse problems, the slice sampling methodology provides a machinery for the design and simulation of a suitable Markov chain without the necessity to tune any parameters as in many Metropolis-Hastings algorithms. In the machine learning community slice sampling is a frequently used approach, which appears not only there as standard sampling tool. In particular, the elliptical slice sampler attracted in the last decade considerable attention as a dimension-robust algorithm. However, from a theoretical point of view it is not well understood. In this talk, we show the geometric ergodicity of Markov chains generated by elliptical slice sampling as well as simple slice sampling with particular emphasis on their (in)dependence on the state space dimension.

Björn Sprungk
TU Freiburg, Germany
bjoern.sprungk@math.tu-freiberg.de

MS107

Ensemble Kalman Methods for Machine Learning Applications

The Ensemble Kalman approach for solving inverse problems (EKI) has attracted a lot of attention due to its convergence properties, which allow for exact approximation for linear and Gaussian settings, its interpretation as a derivative-free optimization method (which implicitly computes preconditioned finite differences using the ensemble's pointwise evaluations), and its unreasonable effectiveness even in settings where it is not exact, e.g. in nonlinear settings. We will discuss some recent results of EKI's properties, both for linear and nolinear problems, in its deterministic and stochastic form, and both for its original discrete formulation and continuous-time extension. We will also discuss some ways in which EKI can be employed for the

practical solution of inverse problems.

Philipp Wacker
Friedrich-Alexander Universitaet Erlangen-Nuernberg, Germany
wacker@math.fau.de

MS107

Expression Rate Bounds for Neural Operators

Approximation rates are analyzed for deep surrogates of maps between infinite-dimensional function spaces, which arise e.g. as data-to-solution maps for linear and nonlinear partial differential equations. Such surrogates may be used to speed up computations in parameter estimation problems in engineering. We study in particular deep neural surrogate operators for holomorphic maps between separable Hilbert spaces. The operator inputs are parametrized by stable, affine representation systems such as frames. Additionally, we discuss an interpolation based alternative to this framework, that allows for a deterministic construction and therefore does not require training of the network weights. Algebraic and dimension independent convergence rates are established in both cases.

Jakob Zech
Heidelberg University
jakob.zech@uni-heidelberg.de

Lukas Herrmann
RICAM, Linz, Austria
lukas.herrmann@sam.math.ethz.ch

Christoph Schwab
ETH-Zurich
schwab@sam.math.ethz.ch

MS107

Sequential Decision Support

In many application areas there is a need to determine a control variable that optimizes a prespecified objective. This problem is particularly challenging when knowledge on the underlying dynamics is subject to various sources of uncertainty and access to observations is limited is sequentially becoming available. A scenario such as that for example arises in the context of therapy individualization to improve the efficacy and safety of medical treatment. We present a learning scheme that combines a Sequential Monte Carlo algorithm with a Tree Search and an Upper Confidence Bounds exploration strategy to iteratively estimate unknown parameters and states and an associated optimal control. In particular we discuss how the underlying uncertainties affect the stability and accuracy of the approximative control variable.

Jana de Wiljes
University of Potsdam
wiljes@uni-potsdam.de

MS108

Spectral-Inspired Graph Neural Networks

Graph Neural Networks (GNNs) are powerful deep learning methods for Non-Euclidean data. Most popular GNNs are message passing algorithms (MPNNs) that aggregate and combine signals in a local graph neighborhood. However, shallow MPNNs tend to miss global information or long-range signals, while deep MPNNs can suffer from issues like over-smoothing or over-squashing. To mitigate such issues, existing works typically borrow normalization techniques from training neural networks on Euclidean data, or modify the graph structures. But these approaches are usually not completely understood theoretically, and could increase the overall computational complexity. In this work, we draw inspirations from spectral graph embedding and propose two simple techniques: (1) a layer-wise normalization method `PowerEmbed` to inject global information in local MPNNs; (2) a graph decomposition procedure to exploit the advantages of global and local methods. We show `PowerEmbed` can provably express the top-k leading eigenvectors of the graph operator, which avoids over-smoothing and over-squashing, with provable guarantees for certain stochastic block models. We also provide sufficient conditions for the graph decomposition strategy to outperform pure global and local methods. We apply our techniques in a wide range of simulated and real graphs and demonstrate their superior performance, particularly for dense graphs.

Teresa Huang, Carey Priebe
Johns Hopkins University
nhuang19@jhu.edu, cep@jhu.edu

Soledad Villar
John Hopkins University
svillar3@jhu.edu

Da Zheng
Amazon
dzzhen@amazon.com

MS108

Learning with Graphs: Graph Neural Networks and the Weisfeiler-Leman Algorithm

Graph-structured data is ubiquitous across domains ranging from chemo- and bioinformatics to image and social network analysis. To develop successful machine learning models in these domains, we need techniques mapping the graph's structure to a vectorial representation in a meaningful way – so-called graph embeddings. Starting from the 1960s in chemoinformatics, different research communities have worked in the area under various guises, often leading to recurring ideas. Moreover, triggered by the resurgence of (deep) neural networks, there is an ongoing trend in the machine learning community to design permutation-invariant or -equivariant neural architectures capable of dealing with graph input, often denoted as neural graph networks (GNNs). However, although often successful in practice, GNN's capabilities and limits are understood to a lesser extent. In this talk, we overview some results shedding some light on the limitations and capabilities of GNNs by leveraging tools from graph theory and related areas.

Christopher Morris
McGill University
christopher.morris@tu-dortmund.de

MS108

Graph Classification Using the Interleaving Distance

Reeb graphs and other related graphical signatures have a long history in topology and topological data analysis, but only recently has there been intense interest in finding metrics for these objects. The idea is that graphical signatures such as Reeb graphs, merge trees, and contour

trees encode data in both a space and a real valued function, and we want to build metrics that are sensitive to this information. In this talk, we will focus on a particular metric for comparing Reeb graphs known as the interleaving distance which is a categorical reformulation of the eponymous metric from persistence modules arising in Topological Data Analysis, and show how it can be used as input to statistical and machine learning problems.

Elizabeth Munch
Michigan State University
muncheli@msu.edu

MS108

Revisiting Local Neighborhood Definition from a Sparse Signal Approximation Perspective

Several machine learning methods leverage the idea of locality by using k-nearest neighbors or ϵ-neighborhood techniques to design pattern recognition models. However, the choice of parameters k/ϵ in these methods is often ad-hoc and lacks a clear interpretation. We pose the problem of neighborhood definition as that of sparse signal representation. We show that, under this view, earlier methods correspond to sub-optimal representation based on thresholding and propose an improved approach, Non-Negative Kernel regression (NNK). NNK formulates neighborhood selection as a non-negative basis pursuit problem and is adaptive to the local distribution of samples near the data point of interest. NNK neighbors are geometric, robust, and exhibit superior performance in neighborhood-based machine learning.

Sarath Shekkizhar
University of Southern California
shekkizh@usc.edu

MS109

Reduced Order Models for Spatio-Temporal Chaos

In this talk we explore how methods of machine learning can be used to understand, forecast, and classify chaotic systems through an exploration of the Kuramoto-Sivashinsky PDE in different parameter regimes. We couple an autoencoder neural network architecture with model discovery to produce low-dimensional, simple models that govern the dynamics of the system in an appropriate Poincare section. The underlying mathematical theory of the method is based on topological conjugacies which provide an equivalence relation between chaotic maps and acts as the critical neural network regularization for learning the coordinate and dynamics pairing. We show that the network can be used to project the dynamics down onto the chaotic attractor, discover normal forms, and provide evidence for shadowing of critical unstable periodic orbits in the flow. It is through these applications that we show how machine learning can compliment and extend existing analysis of spatio-temporal chaos, while also providing new reduced-order models that describe the dynamics of high-dimensional systems.

Jason J. Bramburger
George Mason Universityjbrambur@gmu.edu
jbrambur@gmu.edu

MS109

An Opinion Dynamics Model on Hypergraphs and its Mean-field Limit

The perspectives and opinions of people change and spread through social interactions on a daily basis. The study of opinion dynamics gives a quantitative approach to examine how opinions evolve as dynamical processes on networks. In this talk, I will focus on one type of opinion dynamics model bounded-confidence model. I will derive a density-based opinion model on hypergraphs, and extend the model to its mean-field limit, where the mean-field density follows a kinetic equation of Kac type. We prove that the solution of the governing equation is a probability density and converges to a sum of Dirac delta measures as time goes to infinity. Each single delta function represents an opinion cluster at steady states. We also examine the steady-state opinion clusters show nice bifurcation patterns when we increase the variance of initial distributions.

Weiqi Chu
University of California, Los Angeles
chu@math.ucla.edu

MS109

A Deep Learning-Based Non-Newtonian Hydrodynamic Model

A long-standing problem in the modeling of non-Newtonian hydrodynamics of polymeric flows is the availability of reliable and interpretable hydrodynamic models that faithfully encode the underlying micro-scale polymer dynamics. The main complication arises from the long polymer relaxation time, the complex molecular structure, and heterogeneous interaction. We developed a deep learning-based non-Newtonian hydrodynamic model, DeePN^2, that enables us to systematically pass the micro-scale structural mechanics information to the macro-scale hydrodynamics for polymer suspensions. The model retains a multi-scaled nature with clear physical interpretation, and strictly preserves the frame-indifference constraints. We show that DeePN^2 can faithfully capture the broadly overlooked viscoelastic differences arising from the specific molecular structural mechanics without human intervention.

Huan Lei
Michigan State University
leihuan@msu.edu

MS109

Stochastic Model Reduction Near the Unknown Invariant Manifold

We introduce a nonlinear stochastic model reduction technique for high-dimensional stochastic dynamical systems that have a low-dimensional invariant effective manifold with slow dynamics, and high-dimensional, large fast modes. Given only access to a black box simulator from which short bursts of simulation can be obtained, we design an algorithm that outputs an estimate of the invariant manifold, a process of the effective stochastic dynamics on it, which has averaged out the fast modes, and a simulator thereof. This simulator is efficient in that it exploits the low dimension of the invariant manifold, and takes time steps of size dependent on the regularity of the effective process, and therefore typically much larger than that of the original simulator, which had to resolve the fast modes. The algorithm and the estimation can be performed on the fly, leading to efficient exploration of the effective state space, without losing consistency with the underlying dynamics. This construction enables fast and efficient simulation of

paths of the effective dynamics, together with the estimation of crucial features and observables of such dynamics, including the stationary distribution, identification of metastable states, residence times, and transition rates between them. I will talk about the possible extension in the talk.

Felix Ye
Johns Hopkins University
SUNY Albany
xye2@albany.edu

MS110

Efficient Bayesian Inference for Mechanistic Modeling with High-Throughput Data

Bayesian methods are routinely used to combine experimental data with detailed mathematical models to obtain insights into physical phenomena. However, the computational cost of Bayesian computation with detailed models has been a notorious problem. While high-throughput data presents opportunities to calibrate sophisticated models, comparing large amounts of data with model simulations quickly becomes computationally prohibitive. Inspired by the method of stochastic gradient descent, we propose a minibatch approach to approximate Bayesian computation. Through a case study of a high-throughput scratch assay experiment, we show that reliable inference can be performed at a fraction of the computational cost of a traditional approximate Bayesian inference scheme. By applying a detailed mathematical model of cell movement, proliferation, and death to a wide range of gene knockdowns, we characterise the relative contributions of local cell density-dependent and -independent mechanisms of cell movement and proliferation. Within a screen of 118 gene knockdowns, we characterise functional subgroups of gene knockdowns, each displaying its own typical combination of local cell density-dependent and independent movement and proliferation patterns. By comparing these patterns to experimental measurements of cell counts and wound closure, we find that density-dependent interactions play a crucial role in the outcome of the scratch assay.

Simon Martina Perez, Heba Sailem
University of Oxford
martinaperez@maths.ox.ac.uk, heba.sailem@eng.ox.ac.uk

Ruth E. Baker
Mathematical Institute
University of Oxford
ruth.baker@maths.ox.ac.uk

MS110

A Framework for Bayesian Parameter Estimation and Uncertainty Quantification in Systems Biology

We propose a Bayesian parameter estimation framework to facilitate model calibration and complete uncertainty analysis for dynamical models of biological systems. The available experimental data for these systems are generally sparse and noisy, which introduces uncertainties into the model. Despite this, systems biologists often overlook rigorous uncertainty quantification. We present a new modeling paradigm that bridges the gap between standard practices in systems biology and more thorough uncertainty quantification. Our proposed method accounts for uncertainties in the data and the model structure by adapting an approximate marginal Markov chain Monte Carlo (MCMC) method for Bayesian parameter estimation. This method enables us to characterize the parameter uncertainty and analyze the distribution of possible model outputs. We use this method to study the effects of experimental assumptions on parameter inference for a model of a well-known biological system; the mitogen-activated protein kinase signaling pathway. Our work demonstrates uncertainty quantification in systems biology and outlines the considerations and necessary steps to apply this method to future modeling efforts.

Nathaniel J. Linden
University of California San Diego
nlinden@ucsd.edu

Padmini Rangamani
Department of Mechanical and Aerospace Engineering
University of California San Diego
padmini.rangamani@eng.ucsd.edu

Boris Kramer
University of California San Diego
bmkramer@ucsd.edu

MS110

Highly Reduced Models of Random ODEs: Some Examples and Analysis

Consider a modeler's job when tasked to describe a system of interacting species, whether for an ecological environment, an epidemic, or even a chemical reaction: In many cases, interaction coefficients are unknown, relevant data are unavailable, inclusion of all possible species is computationally infeasible, etc. Thus, the modeler may have to model the system with a small subset of the species truly active in the system. This would seem to make the modeler's job extremely difficult, or even perhaps a lost cause. Surprisingly, though, we have examined several cases of random nonlinear ordinary differential equations where this type of extreme reduction, along with a simple, data-driven, embedded discrepancy model, still results in descriptive and predictive models. In this talk, I'll show a few specific numerical examples, and give some analysis that supports such reductions.

Rebecca Morrison
CU Boulder
rebeccam@colorado.edu

MS111

Inferring Topological Transitions in Pattern-Forming Processes Via Self-Supervised Learning

The identification and classification of transitions in topological and microstructural regimes in pattern-forming processes is critical for understanding and fabricating microstructurally precise novel materials in many application domains. Relevant microstructure transitions may depend on process parameters in subtle and complex ways that are not captured by the classic theory of phase transition. While supervised machine learning methods may be useful for identifying transition regimes, they need labels which require prior knowledge of order parameters or relevant structures. Motivated by the universality principle for dynamical systems, we instead use a self-supervised approach to solve the inverse problem of predicting process parameters from observed microstructures using neural networks. This approach does not require labeled data about the target task of predicting microstructure transitions. We show that the difficulty of performing this prediction task is re-

lated to the goal of discovering microstructure regimes, because qualitative changes in microstructural patterns correspond to changes in uncertainty for our self-supervised prediction problem. We demonstrate the value of our approach by automatically discovering transitions in microstructural regimes in two distinct pattern-forming processes: the spinodal decomposition of a two-phase mixture and the formation of concentration modulations of binary alloys during physical vapor deposition of thin films.

Remi Dingreville
Sandia National Laboratories
rdingre@sandia.gov

MS111

Convolutional Neural Networks for Data Compression and Reduced-Order Modeling

Convolutional neural networks are seen as a useful alternative to linear mappings for dimension reduction, but there is very little understanding of how their architecture affects performance on scientific tasks. We show that the convolutional operation employed for data compression and model reduction dramatically influences network effectiveness in these applications, and a novel graph convolutional autoencoder architecture is proposed which produces lower reconstruction error than conventional CNN when the domain is unstructured and the latent dimension is large enough.

Anthony Gruber
Sandia National Laboratories
adgrube@sandia.gov

MS111

Data-Driven Phase Field Model Discovery Using the Physics Informed Neural Networks

In this talk, we introduce a new deep learning framework for discovering the phase-field models from existing image data. The new framework embraces the approximation power of physics informed neural networks (PINNs) and the computational efficiency of the pseudo-spectral methods, which we named pseudo-spectral PINN or SPINN. Unlike the baseline PINN, the pseudo-spectral PINN has several advantages. First of all, it requires less training data. Secondly, it is computationally efficient, as the pseudo-spectral method is used for spatial discretization. Thirdly, it requires less trainable parameters compared with the baseline PINN, which significantly simplifies the training process and potentially assures fewer local minima or saddle points. We illustrate the effectiveness of pseudo-spectral PINN through several numerical examples. The newly proposed pseudo-spectral PINN is rather general, and it can be readily applied to discover other PDE-based models from image data.

Jia Zhao
Utah State University
jia.zhao@usu.edu

MS111

Learning Stochastic Behavior from Aggregate Data Using Optimal Transport Distance

Learning nonlinear dynamics from aggregate data is a challenging problem because the full trajectory of each individual is not available, namely, the individual observed at one time may not be observed at the next time point, or the identity of individual is unavailable. This is in sharp contrast to learning dynamics with full trajectory data, on which the majority of existing methods are based. We propose a novel method using the weak form of Fokker Planck Equation (FPE) a partial differential equation to describe the density evolution of data in a sampled form, which is then combined with Wasserstein generative adversarial network (WGAN) in the training process. In such a sample-based framework we are able to learn the nonlinear dynamics from aggregate data with-out explicitly solving FPE. We demonstrate our approach in the context of a series of synthetic and real-world data sets. This presentation is based on a recent joint work with Shaojun Ma (Math, GT), Shu Liu (Math GT) and Hongyuan Zha (CUHK-Shenzhen).

Haomin Zhou
Georgia Institute of Technology
School of Mathematics
hmzhou@math.gatech.edu

MS112

Model Based Deep Learning: Applications to Imaging and Communications

Deep neural networks provide unprecedented performance gains in many real-world problems in signal and image processing. Despite these gains, the future development and practical deployment of deep networks are hindered by their black-box nature, i.e., a lack of interpretability and the need for very large training sets. On the other hand, signal processing and communications have traditionally relied on classical statistical modeling techniques that utilize mathematical formulations representing the underlying physics, prior information and additional domain knowledge. Simple classical models are useful but sensitive to inaccuracies and may lead to poor performance when real systems display complex or dynamic behavior. Here we introduce various approaches to model based learning which merge parametric models with optimization tools leading to efficient, interpretable networks from reasonably sized training sets. We will consider examples of such model-based deep networks to image deblurring, image separation, super resolution in ultrasound and microscopy, efficient communications systems, and finally we will see how model-based methods can also be used for efficient diagnosis of COVID19 using X-ray and ultrasound.

Yonina Eldar
Weizmann Institute of Science
yonina.eldar@weizmann.ac.il

MS112

Interpretable RNNs via Algorithm Unrolling: Design, Applications and Generalization Analysis

Deep unfolding methods (a.k.a. algorithm unrolling) design deep neural networks (DNNs) as learned variations of iterative algorithms to solve various signal processing tasks, leading to models which are interpretable by design. Likewise, deep unfolding recurrent neural networks (RNNs) are obtained by unrolling iterative algorithms that are applied sequentially in time. In this presentation, we elaborate on a generic deep unfolding RNN architecture coined reweighted-RNN, originally designed for the task of video reconstruction. We investigate theoretical aspects of reweighted-RNN and similar models in terms of their generalization ability. Specifically, we derive generalization error bounds (GEB) via Rademacher complexity analy-

sis: to our knowledge, these are the first generalization bounds proposed for deep unfolding RNNs. We conduct a series of experiments to empirically evaluate the performance and the GEB of reweighted-RNN, both in regression (video reconstruction and super-resolution) and classification (language modelling) settings. Our results indicate that reweighted-RNN outperforms traditional RNNs in both settings. These experiments allow us to relate the empirical generalization error to the theoretical bounds. Furthermore, we show that reweighted-RNN achieves tight theoretical error bounds with minimal decrease in accuracy, when trained with explicit regularization and weight constraints.

Boris Joukovsky, Tanmoy Mukherjee
Vrije Universiteit Brussel
imec
bjoukovs@etrovub.be, tmukherj@etrovub.be

Huynh Van Luong, Nikos Deligiannis
Vrije Universiteit Brussel
huynhlvd@gmail.com, ndeligia@etrovub.be

MS113

Linear Convergence of the Subspace Constrained Mean Shift Algorithm: From Euclidean to Directional Data

Directional data consist of observations distributed on a (hyper)sphere, and appear in many applied fields, such as astronomy, ecology, and environmental science. This paper studies both statistical and computational problems of kernel smoothing for directional data. We generalize the classical mean shift algorithm to directional data, which allows us to identify local modes of the directional kernel density estimator (KDE). The statistical convergence rates of the directional KDE and its derivatives are derived, and the problem of mode estimation is examined. We also prove the ascending property of the directional mean shift algorithm and investigate a general problem of gradient ascent on the unit hypersphere. To demonstrate the applicability of the algorithm, we evaluate it as a mode clustering method on both simulated and real-world data sets.

Yen-Chi Chen
Statistics
University of Washington
yenchic@uw.edu

Yikun Zhang
University of Washington
yikun@uw.edu

MS113

Statistical Frameworks for Discovering Biophysical Signatures in 3D Shapes and Images

The recent curation of large-scale databases with 3D surface scans of shapes has motivated the development of tools that better detect global patterns in morphological variation. Studies which focus on identifying differences between shapes have been limited to simple pairwise comparisons and rely on pre-specified landmarks (that are often known). In this talk, we present SINATRA: a statistical pipeline for analyzing collections of shapes without requiring any correspondences. Our method takes in two classes of shapes and highlights the physical features that best describe the variation between them. We develop a rigorous simulation framework to assess our approach, which themselves are a novel contribution to 3D image and shape analyses. Lastly, as case studies with real data, we use SINATRA to (1) analyze mandibular molars from four different suborders of primates and (2) facilitate the visual identification of structural signatures differentiating between the trajectories of two protein ensembles resulting from molecular dynamics simulations.

Lorin Crawford
Brown university
lcrawford@microsoft.com

MS113

Geomstats: a Python Package for Geometry in Statistics and Machine Learning

We introduce Geomstats, an open-source Python package for computations and statistics on nonlinear manifolds that appear in machine learning applications, such as: hyperbolic spaces, spaces of symmetric positive definite matrices, Lie groups of transformations, and many more. We provide object-oriented and extensively unittested implementations. Manifolds come equipped with families of Riemannian metrics with associated exponential and logarithmic maps, geodesics, and parallel transport. Statistics and learning algorithms provide methods for estimation, clustering, and dimension reduction on manifolds. All associated operations provide support for different execution backends — namely NumPy, Autograd, PyTorch, and TensorFlow. Thistalkpresents the package, compares it with related libraries, and provides relevant examples. We show that Geomstats provides reliable building blocks to both foster research in differentialgeometryand statistics and democratize the use of Riemanniangeometryin statistics and machine learning. The source code is freely available under the MIT license athttps://github.com/geomstats/geomstats.

Nina Miolane
University of California Santa Barbara
ninamiolane@ucsb.edu

MS113

The Topology of Neuron Activations

Deep neural networks have achieved impressive performance in tasks such as image and text classification. To understand how such performance is achieved, we probe a trained deep neural network by studying neuron activations – the combination of neurons firing in response to a particular input – at various layers of the network. We are interested in understanding the topology of neuron activations using tools from topological data analysis and visualization. To that end, we present various exploration scenarios to provide topological insights into learned representations of neural networks.

Bei Wang
University of Utah
beiwang@sci.utah.edu

MS114

Machine Learning Methods for Output-Based Mesh Adaptation

We present two applications of machine learning to output-based mesh adaptation in computational fluid dynamics (CFD). First, we show how machine learning can assist

in anisotropic mesh optimization. Mesh anisotropy is required for many problems exhibiting singular perturbation characteristics, including high-Reynolds number models of turbulent flow. Predicting the proper amount of element anisotropy, which includes stretching and orientation, is not trivial, and existing optimization approaches are expensive and sometimes sensitive to local errors. We show that a cheaper machine-learning approach compares well with and even rivals the results of the more expensive technique. Second, we show how a machine learning model can simplify the challenging and expensive task of error estimation and mesh adaptation for unsteady turbulent flow. Such simulations do not admit stable unsteady adjoint solutions, and this prevents the direct application of unsteady output-based methods. We therefore instead use an augmented steady-state model that is trained, via machine-learning, to reproduce the statistical quantities of interest from the unsteady simulation. This augmented model provides a steady adjoint solution that is used to weight time-averaged residuals to produce the output error and drive mesh adaptation. Results show that accurate outputs can be obtained with only a few unsteady adaptive iterations, each requiring only a primal simulation.

Krzysztof Fidkowski
University of Michigan
kfid@umich.edu

MS114

Improving Multigrid Methods with Deep Neural Networks

Multigrid methods are one of the most efficient techniques for solving large sparse linear systems arising from Partial Differential Equations (PDEs) and graph Laplacians from machine learning applications. There are two key components of multigrid, smoothing and coarse grid correction. However, finding optimal smoothing algorithms is problem-dependent and can impose challenges for many problems. Meanwhile, as the multigrid hierarchy is formed, coarse-grid operators have significantly more nonzeros per row than the original fine-grid operator, which generates high parallel communication costs on coarse-levels. We first propose an efficient adaptive framework for learning optimal smoothers from operator stencils in the form of CNNs. The CNNs are trained on small-scale problems from a given type of PDEs based on a supervised loss function and can be applied to large-scale problems of the same class of PDEs. We also propose a deep learning framework for sparsifying coarse grid operators. Two neural networks are constructed to learn the sparsity pattern and the corresponding values, respectively. The learned sparser operator has the same interpolation accuracy on algebraic smooth basis. Numerical results on challenging anisotropic rotated Laplacian problems, variable coefficient diffusion problems and linear elasticity problems demonstrate the superior performance of the proposed framework over classical hand-crafted methods.

Ru Huang
Emory University
jerryhuangru@gmail.com

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

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

MS114

Marking Policies for Adaptive Mesh Refinement via Reinforcement Learning

In this work we revisit the marking decisions made in the prototypical adaptive finite element method (AFEM). Experience shows that a naive marking policy leads to inefficient use of adaptive mesh refinement (AMR). Thus AFEM, in practice, often involves time-consuming offline parameter tuning by an expert user. We choose to recast AMR as a partially-observed Markov decision process that can be optimized using methods from reinforcement learning. This recasting delivers a tractable optimization framework which eliminates the need for parameter tuning by expert users. We use the Poisson equation to demonstrate various applications of our framework and construct three representative AFEM simulations inspired by the literature: (1) h-refinement and (2) hp-refinement on non-convex polyhedra and (3) dynamic h-refinement and derefinement for a transient source term. Our experiments indicate that superior marking policies remain undiscovered for many canonical AFEM applications. An unexpected observation in this work is that marking policies trained on one family of PDEs can be robust enough to perform well on problems far outside the training family. For instance, we show that a simple hp-refinement policy optimized on 2D problems can be used for 3D problems without significant performance loss.

Socratis Petrides
Lawrence Livermore National Laboratory
petrides1@llnl.gov

Brendan Keith
Brown University
brendan_keith@brown.edu

Andrew Gillette
Lawrence Livermore National Laboratory
Center for Applied Scientific Computing
gillette7@llnl.gov

MS114

A Data-Driven Approach to Adaptive Mesh Refinement

Adaptive methods for a finite element mesh refinement have emerged as an indispensable tool in achieving accurate and efficiently computed solutions to problems which require impractically fine uniform meshes to obtain an accurate approximation. The adaptive method involves a recursive application of steps: SOLVE-ESTIMATE-MARK-REFINE where in particular the step 'ESTIMATE' requires computing a posteriori error estimator based *only* on the numerical solution and the data of the problem. Over the years, several a posteriori error estimators have been developed and successfully applied but oftentimes, the choice of estimator is ill-suited for the problem at hand. In this talk, we present a learning algorithm that optimizes the choice of a posteriori error estimator for the given problem. As a prototype, we apply this learning algorithm to linear elliptic boundary value problems with poorly resolved solutions and report on the performance of this algorithm for several benchmark problems.

Natasha S. Sharma

The University of Texas at El Paso
nssharma@utep.edu

Ramakrishna Tipireddy
Pacific Northwest National Laboratory
ramakrishna.tipireddy@pnnl.gov

Samrat Chatterjee
Applied Statistics and Computational Modeling
Pacific Northwest National Laboratory
samrat.chatterjee@pnnl.gov

Augustine Twumasi
University of Texas at El Paso
atwumasi@miners.utep.edu

MS115

Infer Explicit Numerical Schemes from Implicit Schemes

Efficient simulation of SDEs is essential in many applications, particularly for ergodic systems that demand efficient simulation of both short-time dynamics and large-time statistics. To achieve the efficiency, dimension reduction is often required in both space and time. In this talk, I will talk about our recent work on the temporal reductions. For temporal dimension reduction, we introduce a framework to construct inference-based schemes adaptive to large time-steps (ISALT) from data, achieving a reduction in time by several orders of magnitudes. The key is the statistical learning in a parametric form of an explicit scheme from the data generated by the stable implicit scheme. Numerical results show that ISALT can tolerate time-step magnitudes larger than plain numerical schemes. It reaches optimal accuracy in reproducing the invariant measure when the time-step is medium-large.

Xingjie Li
University of North Carolina, Charlotte
xli47@uncc.edu

Felix Ye
Johns Hopkins University
SUNY Albany
xye2@albany.edu

Fei Lu
John Hopkins University
flu15@jhu.edu

MS116

Broader Engagement (BE): Fundamentals of Accelerated Data Science with RAPIDS - Part I of II

Traditional CPU-driven data science workflows can be cumbersome but with the power of GPUs, data scientists and Ph.D. students can make sense of data quickly to drive research decisions. In this workshop, student or developers will learn how to build and execute end-to-end GPU-accelerated data science workflows that enable them to quickly explore, iterate, and get their work into production. Using the RAPIDS accelerated data science libraries, developers will apply a wide variety of GPU-accelerated machine learning algorithms, including XGBoost, cuGRAPHs single-source shortest path, and cuMLs KNN, DBSCAN, and logistic regression to perform data analysis at scale. All workshop attendees get access to fully configured, GPU-accelerated servers in the cloud, guidance from an instructor, and the opportunity to network with other developers, data scientists, and researchers. Attendees can earn a certificate to prove subject matter competency and support professional growth.

Xi Chen
University of Kentucky
billchenxi@uky.edu

MS117

The Hidden Convex Optimization Landscape of Deep Neural Networks

Since deep neural network training problems are inherently non-convex, their recent dramatic success largely relies on non-convex optimization heuristics and experimental findings. However, extensive research has failed to provide a solid understanding of the reasons behind the success of these highly complex non-convex optimization models. The choice of optimizer and its internal parameters such as initialization, mini-batching, and step sizes have a considerable effect on the quality of a non-convex model. This is in sharp contrast to convex optimization problems, where these optimization parameters have no effect, and globally optimal solutions can be obtained in a very robust, efficient, and reproducible manner. To this end, we introduce exact convex optimization formulations of ReLU network training problems. We show that two-layer networks can be globally trained via convex programs with the number of variables polynomial in the number of training samples, feature dimension, and the number of hidden neurons. Our results provide an equivalent characterization of neural networks as convex models where a mixture of locally linear models are fitted to the data with sparsity inducing convex regularization. We also discuss extensions to convolutional networks, batch normalization, and deeper architectures. Finally, we present numerical simulations illustrating that the proposed convex approach is faster and more reliable than standard local search heuristics such as SGD.

Tolga Ergen, Mert Pilanci
Department of Electrical Engineering
Stanford University
ergen@stanford.edu, pilanci@stanford.edu

MS117

Representation Costs of Linear Neural Networks: Analysis and Design

For different parameterizations (mappings from parameters to predictors), we study the regularization cost in predictor space induced by l_2 regularization on the parameters (weights). We focus on linear neural networks as parameterizations of linear predictors. We identify the representation cost of certain sparse linear ConvNets and residual networks. In order to get a better understanding of how the architecture and parameterization affect the representation cost, we also study the reverse problem, identifying which regularizers on linear predictors (e.g., l_p norms, group norms, the k-support-norm, elastic net) can be the representation cost induced by simple l_2 regularization, and designing the parameterizations that do so.

Zhen Dai
Computational and Applied Mathematics
University of Chicago
zhen9@uchicago.edu

Mina Karzand
University of Wisconsin Madison
mina.karzand@gmail.com

Nathan Srebro
Toyota Technical Institute - Chicago
nati@ttic.edu

MS117

Regularizing Neural Networks via Radon Domain Total Variation: Representer Theorems, Sparsity, and Splines

What kinds of functions do neural networks learn? Why can neural networks perform well in high dimensional settings? What is the right way to regularize a neural network? This talk will answer these questions and provide mathematical explanations of existing design and training strategies that have evolved largely through experiments. This includes new insights into the importance of weight decay, linear layers, and skip connections, as well as a deeper understanding of sparsity and the curse of dimensionality. Our main result is a representer theorem that states that neural networks are exact solutions to nonparametric learning problems in "mixed variation" function spaces. These results are inspired from classical results in spline theory, and in the univariate case these neural network solutions are exactly the locally adaptive splines of nonparametric statistics and the function spaces are related to classical bounded variation spaces. In the multivariate case these spaces are characterized by total variation in the Radon domain and include functions that are very regular in all but a small number of directions. Spatial inhomogeneity of this sort leads to a fundamental gap between the performance of neural networks and linear methods (which include kernel methods), explaining why neural networks can outperform classical methods for high-dimensional tasks. This theory suggests new neural network architectures that include linear layers and new regularization schemes.

Robert Nowak, Rahul Parhi
University of Wisconsin-Madison
nowak@ece.wisc.edu, rahul@ece.wisc.edu

MS117

The Role of Linear Layers in Nonlinear Interpolating Networks

This talk explores the implicit bias of overparameterized neural networks of depth greater than two layers. Our framework considers a family of networks of varying depth that all have the same capacity but different implicitly defined representation costs. The representation cost of a function induced by a neural network architecture is the minimum sum of squared weights needed for the network to represent the function; it reflects the function space bias associated with the architecture. Our results show that adding linear layers to a shallow ReLU network yields a representation cost that reflects a complex interplay between the alignment and sparsity of ReLU units. Specifically, the addition of linear layers yields an interpolating function that is constant in directions perpendicular to a low-dimensional subspace on which a parsimonious interpolant exists. We discuss the implications of this result for out-of-sample generalization.

Greg Ongie
Marquette University
gregory.ongie@marquette.edu

Rebecca Willett
University of Chicago
willett@uchicago.edu

MS118

Fast Algorithms for Elliptic PDEs with Gaussian Boundary Noise

Recent years have brought numerous advances in statistical modeling and data integration methods for problems involving PDE operators, relying primarily on finite element discretizations. At the same time, fast algorithms for integral equation representations of these PDE operators have demonstrated advantages over finite elements in some deterministic settings, namely homogeneous linear constant coefficient elliptic problems, for example the Laplace and Helmholtz equations. However, work at the intersection of these fields, which would allow statistical modeling using integral equation formulations of the PDE remain relatively undeveloped. We present work here for statistical modeling of homogeneous elliptic boundary value problems where the boundary data is assumed to follow a Gaussian process. We discuss fast algorithms for computing covariances, likelihoods, and conditional distributions of the solution under such models when observations are collected on the boundary or in the volume.

Paul Beckman
NYU - Courant
pgb8409@nyu.edu

Michael O'Neil
New York University
oneil@cims.nyu.edu

Antoine Cerfon
Courant Institute NYU
cerfon@cims.nyu.edu

MS118

Scalably Exploiting Screening in Gaussian Process Computations in the Presence of Measurement Noise

While hierarchical matrices and other methods that exploit "data sparsity" have been demonstrated to be effective in achieving scalability in Gaussian process computations, methods that instead focus on approximating precision matrices with sparse surrogates, such as Vecchia- and DAG-type approximations, are also very effective and offer some advantages over purely algebraic matrix approximations in some settings. Perhaps the principle drawback of these methods, however, is that they are primarily motivated by the screening effect, a phenomenon in which interpolants mostly depend on data locations nearest to the predictand, which can be severely reduced in the presence of additive white noise. While several strategies for dealing with this problem presently exist in the literature, we introduce in this work a new and very general alternative approach that can be employed even when only linear solves with covariance matrices are available. After demonstrating its effectiveness, we close by discussing further applications of our methodology to parameter estimation in other popular modeling paradigms for Gaussian processes.

Christopher Geoga, Michael Stein

Rutgers University
christopher.geoga@rutgers.edu, ms2870@stat.rutgers.edu

Mihai Anitescu
Argonne National Laboratory
Mathematics and Computer Science Division
anitescu@mcs.anl.gov

MS118

Efficient Fourier Methods for Gaussian Processes

In this talk, we describe recent work on a class of Fourier-based fast algorithms for computing with Gaussian processes. These algorithms involve representing a Gaussian process with complex exponentials with equispaced frequencies. This discretization results in a weight-space linear system with a matrix that can be applied in $O(m \log m)$ operations where m is the number of frequencies and can be solved efficiently with iterative methods. The efficient matrix-apply of the linear system results in computational costs that are not highly sensitive to discretizations that use large numbers of Fourier modes, enabling high performance for Gaussian processes in higher dimensions and kernels with fat-tailed spectral densities.

Philip Greengard
Columbia University
pgreenga@gmail.com

MS118

Spatial Factor Modeling: A Bayesian Matrix-Normal Approach for Massive Spatial Data with Missing Observations

The last decade has witnessed substantial developments in scalable models for univariate spatial processes, but such methods for multivariate spatial processes, especially when the number of outcomes is moderately large, are limited in comparison. In this work, we extend scalable modeling strategies for a single process to multivariate processes. We pursue Bayesian inference which is attractive for full uncertainty quantification of the latent spatial process. Our approach exploits distribution theory for the Matrix-Normal distribution, which we use to construct scalable versions of a hierarchical linear model of coregionalization (LMC) and spatial factor models that deliver inference over a high-dimensional parameter space including the latent spatial process.

Lu Zhang
Columbia University
lz2786@columbia.edu

Sudipto Banerjee
Department of Biostatistics
UCLA
sudipto@ucla.edu

MS119

Multi-Tensor Decompositions for Personalized Cancer Medicine

Starting with our invention of the "eigengene," I will describe the formulation of physics-inspired multi-tensor generalizations of the singular value decomposition to (*i*) compare and integrate any data types, of any number and dimensions, and (*ii*) scale with data sizes [Ponnapalli, Saunders, Van Loan and Alter, *PLoS One* 6, e28072 (2011)]. These models (*iii*) are interpretable in terms of known biology and batch effects and (*iv*) correctly predict previously unknown mechanisms [Alter and Golub, *PNAS* 101, 16577 (2004); Omberg, et int., Alter, *MSB* 5, 312 (2009)]. By validating a genome-wide pattern of DNA copy-number alterations in brain tumors as the best predictor of survival, our retrospective clinical trial proved that the models (*v*) discover accurate, precise, and actionable genotype-phenotype relationships, (*vi*) are relevant to populations based upon whole genomes of small cohorts, and (*vii*) can be validated [Ponnapalli, et int., Alter, *APL Bioeng* 4, 026106 (2020)]. We discovered this, and patterns in lung, nerve, ovarian, and uterine tumors, in public data [Bradley, et int., Alter, *APL Bioeng* 3, 036104 (2019)]. Such alterations were recognized in cancer, yet attempts to associate them with outcome failed, demonstrating that our algorithms are uniquely suited to personalized medicine.

Orly Alter
Scientific Computing and Imaging Institute
University of Utah
orly@sci.utah.edu

MS119

Sketching Low Rank Tensors

Machine learning tasks often utilize vast amounts of data where data can be represented in a variety of ways. Instead of representing data as a one or two dimensional arrays, one can instead utilize multi-dimensional arrays to represent more complex relationships between elements in a data set. Moving beyond the matrix or vector representation of data requires the use of tensors and tensor operations. In this talk, we focus on the use of the tensor-tensor t-product for recovering tensors of full and assumedly low tensor rank. In particular, we will explore the use of stochastic iterative methods and random sketches for recovering tensors.

Anna Ma
University of California - Irvine
anna.ma@uci.edu

Yizhe Zhu
UC Irvine
yizhez7@uci.edu

Dominik Stoeger
KU Eichstätt-Ingolstadt
dominik.stoeger@t-online.de

MS119

CP Decomposition via Alternating Least Squares with QR Decomposition

The CP tensor decomposition is used in applications such as machine learning and signal processing to discover latent low-rank structure in multidimensional data. Computing a CP decomposition via an alternating least squares (ALS) method reduces the problem to several linear least squares problems. The standard way to solve these linear least squares subproblems is to use the normal equations, which inherit special tensor structure that can be exploited for computational efficiency. However, the normal equations are sensitive to numerical ill-conditioning, which can compromise the results of the decomposition. In this talk, we develop versions of the CP-ALS algorithm using the QR decomposition and the singular value decomposition (SVD), which are more numerically stable than the normal equations, to solve the linear least squares problems. Our al-

gorithms utilize the tensor structure of the CP-ALS subproblems efficiently, have the same complexity as the standard CP-ALS algorithm when the rank is small, and are shown via examples to produce more stable results when ill-conditioning is present. Our MATLAB implementation achieves the same running time as the standard algorithm for small ranks, and we show that the new methods can obtain lower approximation error.

Rachel Minster
Wake Forest University
minsterr@wfu.edu

MS119

Dynamic Graph Representation Learning Using Tensor Algebra

In recent years, a variety of graph neural networks (GNNs) have been successfully applied for representation learning and prediction on graphs. In many of the real-world applications, the underlying graph changes over time, however, most of the existing GNNs are inadequate for handling such dynamic graphs. In this study, we explore a novel technique for learning embeddings of dynamic graphs using a tensor algebra framework. The tensor framework is based upon the notion of tensor-tensor product, an algebraic formalism to multiply tensors, which inherits mimetic matrix properties. We show that the proposed tensor formalism is a natural extension of popular graph convolutional network (GCN) to the dynamic graph setting. We present numerical experiments on real-world datasets to demonstrate the intrinsic advantages of the proposed tensorial architecture for prediction tasks on dynamic graphs. We also consider an application related to the COVID-19 pandemic and show how the tensor framework can be used for early infection detection and uncertainty quantification in testing from contact tracing data.

Shashanka Ubaru
IBM T. J. Watson Research Center
shashanka.ubaru@ibm.com

Misha E. Kilmer
Mathematics Department
Tufts University
misha.kilmer@tufts.edu

Lior Horesh
IBM Research
lhoresh@us.ibm.com

Osman Asif Malik
University of Colorado, Boulder
osman.malik@colorado.edu

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

MS120

Physics-Informed Neural Operator for Learning Partial Differential Equations

Machine learning methods have recently shown promise in solving partial differential equations (PDEs). They can be classified into two broad categories: approximating the solution function and learning the solution operator. The Physics-Informed Neural Network (PINN) is an example of the former while the Fourier neural operator (FNO) is an example of the latter. Both these approaches have shortcomings. The optimization in PINN is challenging and prone to failure, especially on multi-scale dynamic systems. FNO does not suffer from this optimization issue since it carries out supervised learning on a given dataset, but obtaining such data may be too expensive or infeasible. In this work, we propose the physics-informed neural operator (PINO), where we combine the operating-learning and function-optimization frameworks. This integrated approach improves convergence rates and accuracy over both PINN and FNO models. In the operator-learning phase, PINO learns the solution operator over multiple instances of the parametric PDE family. In the test-time optimization phase, PINO optimizes the pre-trained operator ansatz for the querying instance of the PDE. Experiments show PINO outperforms previous ML methods on many popular PDE families while retaining the extraordinary speed-up of FNO compared to solvers. In particular, PINO accurately solves challenging long temporal transient flows and Kolmogorov flows where other baseline ML methods fail to converge.

Zongyi Li
California Institute of Technology
zongyili@caltech.edu

MS120

Explicitly Antisymmetrized Neural Network Layers for Variational Monte Carlo Simulation

The combination of neural networks and quantum Monte Carlo methods has arisen as a promising path forward for highly accurate electronic structure calculations. Based on the FermiNet architecture, we introduce a generic antisymmetric (GA) neural network layer. The resulting FermiNet-GA architecture can yield very accurate ground state energy for small atoms and molecules. We will also discuss recent studies of the potential arise of sign problems in explicitly antisymmetrized two-layer neural networks.

Lin Lin
University of California, Berkeley
Lawrence Berkeley National Laboratory
linlin@math.berkeley.edu

MS121

Joint Image and Blood Flow Reconstruction for Magnetic Particle Imaging

Magnetic particle imaging (MPI) is a tracer-based medical imaging technique invented by Gleich and Weizenecker [Gleich Weizenecker, Nature, 2005]. It allows for the reconstruction of the spatial distribution of magnetic nanoparticles injected into the blood flow via exploiting their nonlinear magnetization response to changing magnetic fields. MPI has the potential to be a fast imaging technique which might be thus used for blood flow imaging and live instrument tracking. In this talk, we are not only interested in the image reconstruction of the particles from dynamic MPI data, but also in the estimation of the motion in between subsequent frames. Using an alternating scheme to jointly solve both subproblems is expected to be advantageous, as both processes endorse each other [Burger et al, SIAM Journal on Imaging Sciences, 2018]. We start from a variational problem linking images and motion and then exploit prior knowledge to adapt the problem to our application. We analyze the occurring subproblems in order to speed-up the numerical schemes adapted to the image

and motion reconstruction problem. We will illustrate our results by numerical tests computed from simulated as well as real MPI data.

Christina Brandt
Universität Hamburg
Fachbereich Mathematik
christina.brandt@uni-hamburg.de

Lena Zdun
Universität Hamburg
lena.zdun@uni-hamburg.de

MS121

A Variational Approach for Joint Image Recovery-Segmentation Based on Spatially Varying Generalized-Gaussian Models

The joint problem of reconstruction/feature extraction is a challenging task in image processing. It consists in performing, in a joint manner, the restoration of an image and the extraction of its features. In this work, we firstly propose a novel nonsmooth and nonconvex variational formulation of the problem. For this purpose, we introduce a versatile generalised Gaussian prior whose parameters, including its exponents, are space-variant. Secondly, we design an alternating proximal-based optimisation algorithm that efficiently exploits the structure of the proposed nonconvex objective function. We also analyze the convergence of this algorithm. As shown in numerical experiments conducted on joint segmentation/deblurring tasks, the proposed method provides high-quality results.

Emilie Chouzenoux
Inria Saclay
CVN, CentraleSupélec, Univ. Paris Saclay
emilie.chouzenoux@centralesupelec.fr

Gabriele Scrivanti
Central Supelec, France
gabrielescrivanti@gmail.com

Marie-Caroline Corbineau, Jean-Christophe Pesquet
Centrale Supélec - Paris
marie-caroline.corbineau@centralesupelec.fr, jean-christophe@pesquet.eu

MS121

Adaptive Primal-Dual Stochastic Algorithm for Inverse Problems

We present recent developments related to the Stochastic Primal-Dual Hybrid Gradient (SPDHG) algorithm, which builds on the Primal-Dual Hybrid Gradient (PDHG) algorithm used in convex, not necessarily smooth optimization (also known as Chambolle-Pock algorithm). SPDHGs added feature consists in its ability to randomly update partial variables, leading to an accelerated convergence. The admissible step-sizes include a free parameter controlling the primal-dual balance, whose value has been shown to impact the speed of convergence. We present an adaptive version of SPDHG allowing for online tuning of the primal dual balancing parameter. This new algorithm is provenly convergent and we illustrate its improved performance on a CT reconstruction example.

Antonin Chambolle
CEREMADE, CNRS, Paris-Dauphine PSL
chambolle@ceremade.dauphine.fr

Claire Delplancke
Bath University
claire.delplancke@edf.fr

Matthias Ehrhardt
Institute for Mathematical Innovation
University of Bath
m.ehrhardt@bath.ac.uk

Carola Bibiane Schoenlieb
University of Cambridge, United Kingdom
cbs31@cam.ac.uk

Junqi Tang
University of Cambridge
jt814@cam.ac.edu

MS121

Primal-Dual Proximal Methods with Bregman Distances

We discuss Bregman distance extensions of the primal-dual three-operator (PD3O) and Condat-Vu proximal algorithms. When used with standard proximal operators these algorithms include several important methods, including ADMM, as special cases. Extensions to generalized Bregman distances are attractive if the complexity per iteration can be reduced by matching the Bregman distance to the structure in the problem.

Xin Jiang
UCLA
jiangxjames@ucla.edu

MS122

Theory and Computation of Spectral Properties of Lie Derivatives in Dynamical Systems

Koopman operator methods along with the associated numerical algorithms have provided a powerful methodology for the data-driven study of nonlinear dynamical systems. In this talk, we will give a brief outline of how the Koopman group of operators can be generalized beyond function spaces to the space of sections of various vector bundles over the state space. We describe their relationship with the standard Koopman operator on functions as well as describe the new spectral invariants produced by these generalized operators. We then demonstrate how the recently developed spectral exterior calculus framework can be utilized to compute the spectral properties of the generator of the induced operator on sections of the cotangent bundle. We conclude with some applications of the algorithm to some well-known dynamical systems.

Allan Avila
AIMdyn Inc.
allan@aimdyn.com

Dimitrios Giannakis
Dartmouth College
Dimitrios.Giannakis@dartmouth.edu

Tyrus Berry
Department of Mathematical Sciences
George Mason University
tberry@gmu.edu

Igor Mezic

University of California Santa Barbara
mezic@ucsb.edu

MS122

Lawvere Metric Space Learning

Classical machine learning methods such as MDS, kernel methods, and diffusion maps rely on a symmetric distance function that separates unique points. However, many real-world problems such as transitioning between local minima in an energy landscape or measuring time in between states of a dynamical system posses naturally non-symmetric dissimilarity functions. Moreover, these examples of dissimilarities can fail to separate distinct points. Many ad hoc methods such as symmetrizations and removing repeats of data, can throw away potentially valuable information. Instead, we will approach these issues in the most general context of Lawvere metric spaces and consider new methods of machine learning that account for these important properties. Moreover, we will interpret kernel methods from a categorical perspective by considering kernel embeddings as instances of Yonedas lemma, allowing us to consider a much wider class of kernels that can be analyzed in a manifold learning context.

Yemeen Ayub
George Mason University
yayub@gmu.edu

MS122

Closure Detection in Agent Based Models

I will discuss the extraction of gray-box closures for agent based problems - in the form of black box, or grey box corrections to existing approximate closures. This will be demonstrated through both neural network and Gaussian Process Regression learning of the black/grey box closures; the illustrating examples will come from chemotactic as well as financial agent based models.

Ioannis G. Kevrekidis
John Hopkins University
yannisk@jhu.edu

MS122

Data-Driven Spectral Analysis of Koopman Operators Using Resolvent Compactification

Koopman operators transform general (possibly nonlinear) dynamics in phase space to linear dynamics on vector spaces of functions, enabling the use of spectral techniques without modeling constraints such as linearity. We will discuss the spectral decomposition of Koopman operators by an approach to approximate the Koopman generator from data via a compactification of its resolvent. The extraction of approximate Koopman eigenfunctions (and the associated eigenfrequencies) from an unknown system without spectral pollution can be a nontrivial task, particularly if the system has mixed or dense spectra. We will explore implementation of these techniques using data from several different example systems including Lorenz 63.

Claire Valva
New York University
cnv5172@nyu.edu

Dimitrios Giannakis
Dartmouth College
Dimitrios.Giannakis@dartmouth.edu

MS123

Learning Uncertain Quantities for Data-Consistent Inversion: A Data-to-Distribution Pipeline

A common challenge is to quantify uncertainties on model inputs (i.e., parameters) that correspond to a quantitative characterization of uncertainties on observable Quantities of Interest (QoI). In dynamical systems, model output data are often given as a series of state variable responses recorded over a particular time window. Consequently, the dimension of output data can easily exceed $\mathcal{O}(1E4)$ or more due to the frequency of observations, and the correct choice or construction of a QoI from this data is not self-evident. We present a new framework, Learning Uncertain Quantities (LUQ), that facilitates the tractable solution of data-consistent inverse problems for dynamical systems. LUQ enables a data-to-distribution pipeline utilizing routines for filtering data, learning the underlying dynamics in an unsupervised manner, classifying the observations, and performing feature extraction to learn the QoI map. Subsequently, data are transformed into samples coming from the underlying predicted and observed distributions associated with the QoI so that data-consistent solutions are computable. Following the introduction and demonstration of LUQ, numerical results from several SIPs are presented for a variety of dynamical systems arising in the life and physical sciences. In the interest of scientific reproducibility, we provide links to our Python implementation of LUQ, as well as all data and scripts required to reproduce the results in this presentation.

Troy Butler
University of Colorado Denver
butler.troy.d@gmail.com

Steven A. Mattis
Technical University of Munich
steve.a.mattis@gmail.com

Kyle R. Steffen
University of Texas at Austin
Oden Institute for Computational Engineering and Sciences
krsteffen@utexas.edu

Clint Dawson
University of Texas at Austin
clint.dawson@austin.utexas.edu

Donald Estep
Simon Fraser University
donald_estep@sfu.ca

MS123

Quantifying the Uncertainty in Solutions to Stochastic Inverse Problems Using Data-Driven Surrogate Models

Stochastic inverse problems have been attracting increasing attention in recent years due to recent advances in data acquisition techniques which enable utilizing tremendous amounts of data to construct data-informed physics-based computational models. At the same time, machine learning methods have also become much more prevalent in computational science, largely due to their ability to learn and exploit low-dimensional structure in high-dimensional

data. In this presentation, we will describe some of our recent work on using data-driven (machine learning) surrogate models to expedite the solution to both Bayesian and data-consistent inverse problems. On the theoretical side, we show that the universal approximation theorem allows for the use of such machine learning surrogate models to solve such problems. However, on the practical side, we demonstrate that the errors and uncertainties in these surrogate models can significantly impact the accuracy of the inferred probability distribution, and therefore on any subsequent predictions. This naturally leads to a probabilistic characterization of the solution to the stochastic inverse problem that appropriately characterizes the error/uncertainty coming from various sources, e.g., noisy data, architecture/model capacity, solver variability, etc., that impact the solution and predictions due to the use of the surrogate model.

Tim Wildey
Optimization and Uncertainty Quantification Dept.
Sandia National Laboratories
tmwilde@sandia.gov

MS123

Constructing Data-Consistent Solutions to Stochastic Inverse Problems with Sparse Observable Data

Data-consistent inversion is a measure-theoretic approach to solving stochastic inverse problems, i.e., for quantifying aleatoric uncertainties in input parameters to computational models. In the data-consistent approach, initial beliefs about model parameters are summarized with probability distributions. These initial beliefs are then updated using discrepancies between the predicted distribution of a quantity of interest and an observed distribution of the same quantity of interest. The resulting updated distribution is a solution to the inverse problem that is consistent with data in the sense that the push-forward of the updated probability distribution matches the observed distribution of data. While this approach has been successfully applied in a wide variety of contexts, its implementation relies on effective density estimates of the quantity of interest. In particular, the observed distribution is typically approximated using observed data samples. When the number of observable data samples is limited, estimating the density is non-trivial. In this talk, we show how to mitigate the issue of sparse observable data by utilizing structure induced by the push-forward samples from the model. Specifically, we demonstrate how to incorporate this information into flow-based machine learning methods for density estimation, improving the approximation of the observed density. Issues of induced bias and uncertainty due to such incorporations are discussed.

Tian Yu Yen
Sandia National Laboratories
tyen@sandia.gov

MS124

Geometric Scattering on Non-Euclidean Data

The scattering transform is a mathematical model for convolutional neural networks (CNNs), which enjoy several useful properties that make them particularly powerful when applied to data such as two-dimensional images. Because of the success of CNNs, there is great interest in designing similar architectures for non-Euclidean data. Several variants of the scattering transform have been proposed for special cases of such data, and it has been shown that these variants enjoy desirable properties such as stability and invariance. In this talk, we will outline a unifying framework for geometric scattering on generic measure spaces and showcase how the scattering transform can be implemented in a variety of non-Euclidean settings, such as manifolds with and without boundary and directed graphs. We also propose two methods for approximating the scattering transform using point cloud data sampled from an underlying manifold. Finally, we show that the scattering transform is an effective tool for signal, manifold, and node classification using several data formats, including spherical images, biomedical single-cell data, and directed graphs.

Joyce A. Chew
University of California Los Angeles
joycechew@math.ucla.edu

MS124

Persistence of Network Representations of Time Series for Dynamic State Detection

One of the most important problems in time series analysis is change point detection. These change are often associated with qualitative changes in the underlying dynamic skeleton as one or more of the system parameters vary. A standard approach to capture these changes is to study the structure or shape of the corresponding point cloud constructed using Takens embedding, for example using persistent homology from Topological Data Analysis (TDA). However, using persistence on the reconstructed state space can be limited by computation time since the simplicial complex generated in this case is large, and it possibly contains a great deal of redundant information. For this reason, we turn to a more recent method for encoding the structure of the attractor using network representations such as an Ordinal Partition Network (OPN). We show how to incorporate weight information, and we describe a pipeline to leverage TDA for analyzing the resulting weighted networks. Our results show that this framework is more resilient to noise in comparison to its unweighted counterpart, and it improves the accuracy of the dynamic state detection.

Firas A. Khasawneh
Michigan State University
khasawn3@egr.msu.edu

MS124

Advancing Neuroimaging with Graph Representation Learning

Recent work on neuroimaging has demonstrated significant benefits of using population graphs to capture non-imaging information in the prediction of nearedegenerative and neurodevelopmental disorders. This has been enabled by advances in the field of graph representation learning. The non-imaging attributes may contain demographic information about the individuals, but also the acquisition site, as imaging protocols and hardware might significantly differ across sites in large-scale studies. This talk will give an overview of the advances that graph representation learning has contributed to the fields of neuroimaging and connectomics in the recent years. It will also discuss fairness considerations that arise when these model leverage sensitive attributes.

Ira Ktena

DeepMind
iraktena@gmail.com

MS124

Compartmental Modeling to Forecast the 2022 U.S. Midterm Elections

Election dynamics are a rich complex system, and forecasting this fall's U.S. midterm elections is an exciting, high-stakes problem with many sources of subjectivity and uncertainty. In this talk, we take a dynamical-systems perspective on election forecasting, with the goal of helping to shed light on the forecast process and raising questions for future work. By adapting a well-studied model from epidemiology to account for interactions between voters in different states, we show how to combine a compartmental approach with polling data to produce forecasts of presidential, senatorial, and gubernatorial elections at the state level. Our results for the last two decades of U.S. elections are largely in agreement with those of popular analysts, and we apply our model to forecast the upcoming midterm elections. We also use our modeling framework to explore how weighting polling data by polling organization and accounting for uncertainty in different ways affects our forecasts.

Alexandria Volkening
Purdue University
avolkening@purdue.edu

MS125

Data-Driven Computation for Stochastic Dynamics: Fokker-Planck Solver and Beyond

The time evolution of stochastic differential equation is described by the Fokker-Planck equation. In this talk I will introduce our novel data-driven Fokker-Planck solver and its applications. The main idea of our solver is to use less accurate Monte Carlo simulation data to guide either the numerical PDE solver (low dimensional version) or the artificial neural network (high dimensional version) to find the numerical solution. I will also demonstrate some applications of our data-driven Fokker-Planck solver in importance sampling and identifying invariant manifolds of dynamical systems.

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

Jiayu Zhai
University of Massachusetts Amherst
zhai@math.umass.edu

MS125

HAL: Hyperactive Bayesian Learning for Molecular Force Fields

We present a novel method for the efficient generation of data sets of atomic configurations for the purpose of fitting an atomistic force-field model. The method combines elements of active learning and adaptive biasing, and is formulated in a Bayesian framework. We explain the underlying ideas for the design of the method in the context of a simple synthetic toy problem. We demonstrate the practical applicability of the method in real data examples, and show dramatic performance gain in the assembly of relevant data sets in material sciences applications.

Matthias Sachs
University of Birmingham
m.sachs@bham.ac.uk

Christoph Ortner
The University of British Columbia
ortner@math.ubc.ca

Cas van der Oord, Gabor Csanyi
University of Cambridge
casv2@eng.cam.ac.uk, gc121@cam.ac.uk

MS125

Data-Driven Learning and Prediction of Dynamics on Lie Groups

Incorporating domain knowledge and rigid structures into deep learning models provides powerful ways to improve their performance for specific classes of problems. For instance, previous works have considered the Hamiltonian / symplectic structure for learning the dynamics of physical systems. Although successful, they have mostly been limited to Euclidean spaces. In this talk, we extend previous approaches for dynamics prediction to the Lie group manifold and leverage structure-preserving integrators that guarantee the preservation of the manifold structure and symplecticity. Improved results show that structure-preservation plays a key role in being able to scale the learning to long time horizons.

Oswin So
Georgia Tech
oswinso@gatech.edu

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

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

MS125

A Generalized Ace Framework and Its Applications in Learning Equivariant Properties

Atomic Cluster Expansion(ACE) [Drautz, 2019] is a powerful tool/descriptor for approximating isometry invariant properties. There are also many other properties of interest that have different physical symmetries under certain group actions. In this talk, I will introduce a generalized ACE framework, which provides an efficient way to approximate those symmetric properties with a systematically controllable error, no matter they are scalars, vectors, matrices or even general tensors. As a concrete example, its implementations on Self-Consistent Hamiltonian learning will be introduced in detail.

Liwei Zhang
University of British Columbia
lzhang@math.ubc.ca

MS126

Sparse Inverse Problems with Generalized Gamma

Hyperpriors

Hierarchical models with generalized gamma hyperpriors provide a flexible framework to impose regularizations in Bayesian formulations to inverse problems. Despite the Bayesian motivation for these models, existing methodologies are limited to maximum a posteriori estimation. The potential to perform uncertainty quantification has not yet been realized. We introduce a variational iterative alternating scheme and regularized iterative ensemble Kalman filter for hierarchical inverse problems with generalized gamma hyperpriors. The proposed approaches yield accurate reconstruction and provide meaningful uncertainty quantification. We illustrate the performance of our methodology in several computed examples, including a compressed sensing problem, subsurface flow inverse problem, and sparse identification of dynamical systems from time-series data.

Hwanwoo Kim, Daniel Sanz-Alonso, Alexander Strang
University of Chicago
hwkim@uchicago.edu, sanzalonso@uchicago.edu, alexstrang@uchicago.edu

MS126

Hierarchical Bayesian Inference For Hematopoiesis

Hematopoiesis is the complex mechanism which describes how Hematopoietic stem cells (HSCs) produce diverse functional mature blood cells through undergoing multiple differentiation stages include intermediate heterogeneous multipotent progenitor cells (MPPs). Although a variety of Hematopoiesis model systems have been proposed before, few of them could take the advantage of fitting and interpreting experimental data through Bayesian framework inference method as well as integrate statistical with mechanistic models. In this paper, instead of employing the traditional nonlinear hill-equation Ordinary Differential Equation (ODE) model, we propose a new data-driven and explanatory chemical reaction ODE model with a deep insight into the mechanism and kinetics of feedback regulation for Hematopoiesis. The main objective for this paper is to probabilistically learn the division and feedback regulation rate values of various Hematopoietic cell compartments through Markov Chain Monte Carlo (MCMC) algorithm and a scalable, hierarchical Bayesian framework which takes the multi-individual heterogeneity into account. The performance of new model and framework is being tested and evaluated on both synthetic dataset and real experimental dataset to rationalize population variability within species as well as cross-dataset. Our scalable Bayesian framework and pipeline is broadly applicable to various biological process involves tumor growth modeling and tissue engineering.

Yushang Lai
Mathematical & Computational Systems Biology UC Irvine, CA
Center for Cancer Systems Biology UC Irvine, CA
yushanl2@uci.edu

Brianna Hoover, Gajalakshmi Ramanathan
Division of Hematology/Oncology UC Irvine, CA, USA
bcraver@uci.edu, gramanat@uci.edu

Angela Fleischman
Center For Cancer Systems Biology UC IRVINE, CA, USA
Division of Hematology/Oncology UC Irvine, CA, USA
agf@hs.uci.edu

Babak Shahbaba
University of California, Irvine
babaks@uci.edu

Vladimir Minin
UC Irvine
vminin@uci.edu

John Lowengrub
Center For Cancer Systems Biology UC IRVINE, CA, USA
Department of Mathematics, UC IRVINE, CA, USA
jslowengrub@gmail.com

MS126

Building and Constraining Kinetic Models for Spatially Organized Metabolic Pathways

Metabolic processes are exquisitely organized in space and time, and synthetic biologists are now engineering spatially-based strategies to enhance desired chemical reactions. A relatively ubiquitous strategy in bacteria is compartmentalization within 50-100 nm protein-based compartments. Evidence suggests, these compartments segregate a series of chemical reactions from the rest of the cellular metabolism, potentially enhancing flux through the compartmentalized pathway and avoiding intermediate loss due to leakage out of the cell or undesirable reactions. However, system functionality depends on uncharacterized parameters such as the permeability of the compartments to metabolites and the kinetic activity of enzymes. I will discuss our progress towards characterizing natural and engineered pathways using ODE models and data incorporation methods. Our current approaches include experimental design, model selection, Bayesian parameter estimation, and characterization of sensitive and insensitive parameter combinations using active subspaces. Our overall goal is to create methodologies to rapidly identify rate-limiting steps, estimate unknown parameters, and determine when models can be reduced or require additional complexity to capture natural behavior or produce desired engineering results.

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

Andre Archer, Sasha Shirman
Northwestern University
andrearcher2017@u.northwestern.edu, mathcone@gmail.com

MS126

Parameter Estimation for Compartmental Models in Physiology Combining Variational Inference and Autoregressive Transformations

Compartmental models of the human cardiovascular system are increasingly used either as low-fidelity surrogates or to provide boundary conditions in geometrically multiscale computational models at the organ level. Realistic models typically contain a large number of parameters. Additionally, the cost of inference through sampling-based approaches such as Markov chain Monte Carlo may still be significant even if a single model realization requires seconds to be evaluated. Non linearity and lack of parameter identifiability may also result in complicated and/or multimodal posterior distributions that are difficult to sam-

ple from. In this talk, I will describe a modular framework based on variational inference which simultaneously trains a collection of autoregressive transformations and a neural network surrogate through alternated gradient-based weight updates. Samples are adaptively collected either from a low-discrepancy Sobol' sequence (so-called "pre-grid", providing global surrogate accuracy) or from an adaptive, memory-aware, sampling strategy (increasing the surrogate accuracy in high posterior regions). Finally, I will discuss an adaptive annealing strategy, which makes it easier to sample from complicated posterior distributions, while reducing the overall computational cost with respect to widely used linear annealing schedulers.

Daniele E. Schiavazzi
University of Notre Dame
dschiavazzi@nd.edu

MS127

Compression of Streaming Scientific Data

This talk will discuss recent advancements in streaming compression methods for scientific computational and experimental data. We will exploit the fact that scientific data follow some underlying physical principle (e.g., experimental data) or some known, possibly complex model (e.g., computational simulation data). Such underlying properties allow for compression methods with unique designs capable of orders of magnitude reductions with prescribed accuracy in goal-oriented lossy and lossless compression. The key challenges that face streaming scientific data compression include: (1) the incorporation of known scientific information into compression; (2) the massive volume of data produced requiring methods that act on streaming data near the point of generation; (3) the uncertainty quantification for data with noise, error, or missing pieces; and (4) the ability to effectively use new computing architecture, both centralized and at the edge. We demonstrate that users can spend more computational cost in either compression or decompression, then greater accuracy and compression are possible. Studies for these methods demonstrate compression rates in the 10–1,000X range with high levels of compression increasing computational cost

Richard Archibald
Computational Mathematics Group
Oak Ridge National Labratory
archibaldrk@ornl.gov

MS127

Data-Driven Causal Model Discovery and Personalized Prediction in Alzheimers Disease

With the explosive growth of available biomarker data in Alzheimer's disease (AD) clinical trials, numerous mathematical models have been developed to characterize disease-relevant biomarker trajectories over time. While some of these models are purely empiric, others are causal, built upon various hypotheses of AD pathophysiology, a complex and incompletely understood area of research. One of the most challenging problems in computational causal modeling is to use a purely data-driven approach to derive not only the model's parameters, but the mathematical model itself, without any prior hypothesis bias. In this talk, we develop an innovative data-driven modeling approach to build and parameterize a causal model to characterize the trajectories of AD biomarkers. This approach integrates an understanding of AD pathophysiology, causal model learning and parameterization, personalized prediction, and parameter sensitivity analysis. By applying this integrated approach to a large multicenter database of AD biomarkers, the Alzheimers Disease Neuroimaging Initiative, several causal models for different AD stages are revealed. In addition, personalized models for each subject are calibrated and provide accurate predictions of future cognitive status.

Guang Lin, Haoyang Zheng
Purdue University
guanglin@purdue.edu, zheng528@purdue.edu

Jeffrey Petrella, P.Murali Doraiswamy
Duke University
jeffrey.petrella@duke.edu, murali.doraiswamy@duke.edu

Wenrui Hao
Pennsylvania State University
wxh64@psu.edu

MS127

Metanor: A Meta-Learnt Nonlocal Operator Regression Approach with Convergence Across Resolutions

We propose MetaNOR, a meta-learnt approach for transfer-learning operators based on the nonlocal operator regression. The overall goal is to efficiently provide surrogate models for new and unknown material-learning tasks with different microstructures. The algorithm consists of two phases: (1) learning a common nonlocal kernel representation from existing tasks; (2) transferring the learned knowledge and rapidly learning surrogate operators for unseen tasks with a different material, where only a few test samples are required. We apply MetaNOR to model the wave propagation within 1D metamaterials, showing substantial improvements in the sampling efficiency for new materials.

Lu Zhang
Lehigh University
luz319@lehigh.edu

MS127

Deep Network Approximation: Achieving Arbitrary Accuracy with Fixed Number of Neurons

This talk discusses a new type of simple feed-forward neural network that achieves the universal approximation property for all continuous functions with a fixed network size. This new type of neural network is simple because it is designed with a simple and computable continuous activation function sigma leveraging a triangular-wave function and a softsign function. First, we prove that s-activated networks with width $36d(2d+1)$ and depth 11 can approximate any continuous function on a d-dimensional hypercube with an arbitrarily small error. Next, we show that classification functions arising from image and signal classification can be exactly represented by sigma-activated networks with width $36d(2d+1)$ and depth 12, when there exist pairwise disjoint closed bounded subsets of R_d such that the samples of the same class are located in the same subset.

Shijun Zhang
National University of Singapore

zhangshijun@u.nus.edu

MS128

Learn Bifurcation Points Through Neural Networks

Nonlinear parametric systems have been widely used in modeling nonlinear dynamics in science and engineering. Bifurcation analysis of these nonlinear systems on the parameter space is usually used to study the solution structure, such as the number of solutions and the stability. In this talk, I will introduce a new machine learning approach to compute the bifurcations via so-called equation-driven neural networks (EDNNs). The EDNNs consist of a two-step optimization: the first step is to approximate the solution function of the parameter by training empirical solution data; the second step is to compute bifurcations using the approximated neural network obtained in the first step. Both theoretical convergence analysis and numerical implementation on several examples will be used to demonstrate the feasibility of the proposed method.

Wenrui Hao
Pennsylvania State University
wxh64@psu.edu

MS128

Generalized Nash Equilibrium Problems

This talk discuss generalized Nash equilibrium problems that are given by rational functions. Rational expressions for Lagrange multipliers and feasible extensions of KKT points are introduced to compute generalized Nash equilibria (GNEs). We give a hierarchy of rational optimization problems to solve rational generalized Nash equilibrium problems. The existence and computation of feasible extensions are studied. The Moment-SOS relaxations are applied to solve the rational optimization problems. Under some general assumptions, we show that the proposed hierarchy can compute a GNE if it exists or detect its nonexistence. Numerical experiments are given to show the efficiency of the proposed method.

Jiawang Nie
University of California, San Diego
njw@math.ucsd.edu

Xindong Tang
The Hong Kong Polytechnic University
xindong.tang@polyu.edu.hk

Suhan Zhong
University of California, San Diego
suzhong@ucsd.edu

MS128

Mirror-Prox Sliding Methods for Solving a Class of Generalized Monotone Variational Inequalities

We propose new algorithms for solving weak solutions to a class of generalized monotone variational inequalities with compact feasible sets. For problems involving gradient of a smooth convex function, we show that it is possible to skip computations of the gradients from time to time, while still maintaining the optimal iteration complexity for solving monotone variational inequalities. Specifically, for deterministic variational inequalities problems involving the sum of the gradient of a smooth convex function ∇G and a monotone operator H, we propose an algorithm, called the mirror-prox sliding method, that is able to compute an ε-approximate weak solution with at most $O((L/\varepsilon)^{1/2})$ evaluations of ∇G and $O((L/\varepsilon)^{1/2} + M/\varepsilon)$ evaluations of H, where L and M are Lipschitz constants of ∇G and H, respectively. We also propose a stochastic mirror-prox sliding method for the case when the operator H can only be accessed through its stochastic samples. It computes a stochastic ε-approximate weak solution with at most $O((L/\varepsilon)^{1/2})$ evaluations of ∇G and $O((L/\varepsilon)^{1/2} + M/\varepsilon + \sigma^2/\varepsilon^2)$ samples of H, where σ is the variance of the stochastic samples of H.

Guanghui Lan
Georgia Institute of Technology
george.lan@isye.gatech.edu

Yuyuan "Lance" Ouyang
Department of Mathematical Sciences
Clemson University
yuyuano@clemson.edu

MS129

Mapper-Style Interleavings for Geometric Graphs

The interleaving distance on Reeb graphs is a commonly used metric for comparison and analysis. In this talk, we present existing and ongoing work for computing a discretized interleaving distance on mapper graphs, a more commonly used structure which approximates the underlying Reeb graph but is easier to compute. In addition, we introduce some generalizations of the Reeb interleaving distance to geometric graphs, which are embedded in the Euclidean plane.

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

MS129

Box Filtration

We propose a new filtration for the topological data analysis of points in high-dimensional Euclidean space where the convex sets are hyperrectangles or boxes. Rather than center them at each point, we grow the boxes non-uniformly in each dimension or direction based on the distribution of points so as to better capture the topology of the point set. We present efficient algorithms to construct the box filtration. We also prove stability results for the boxes constructed in the filtration under small changes of input parameters. We compare topological summaries of standard data sets in the form of persistence diagrams produced by box filtration to those produced by the standard distance-to-measure (DTM) filtration. In a similar fashion, using boxes as cover elements in the Mapper framework allows the adaptive modifications of the covers.

Enrique Alvarado
University of California Davis

ealvarado@math.ucdavis.edu

Prashant Gupta
Natural Intelligence Systems
pragpt100@gmail.com

Bala Krishnamoorthy
Washington State University
kbala@wsu.edu

MS129

A Distance for Geometric Graphs via the Labeled Merge Tree Interleaving Distance

Geometric graphs appear in many real world data sets, such as road networks, sensor networks, and molecules. We investigate the notion of distance between graphs and present a stable metric to measure the distance between two geometric graphs via merge trees. In order to preserve as much useful information as possible from the original data, we introduce a way of rotating the sub-level set in obtaining the merge trees via the idea of the directional transform. We represent the merge tress by performing a surjective multi-labeling scheme and then compute the distance between two representative matrices. Our metric is not only reflective of the information from the input graphs, but also computationally efficient. We illustrate its utility by implementation and application on real data.

Elena Wang
Michigan State University, U.S.
wangx249@msu.edu

MS129

Interactive Exploration of High-Dimensional Point Clouds via the Mapper Graphs

High-dimensional point cloud data are becoming ubiquitous. However, the challenge is how to understand and interpret the underlying structures. This talk introduces an interactive toolbox, Mapper Interactive, for the visual exploration of high-dimensional point clouds via mapper graphs, as well as its domain-specific adaptation, PhenoMapper, for analyzing multidimensional phenomics data.

Youjia Zhou
University of Utah
zhou325@sci.utah.edu

MS130

Machine Learning Based Refinement Strategies for Polygonal and Polyhedral Grids with Applications to Virtual Element and Polyhedral Discontinuous Galerkin Methods

We propose new strategies based on Machine Learning techniques to handle polygonal/polyhedral grid refinement. We show that Convolutional Neural Networks can be successfully employed to correctly identify the shape of a polygonal/polyhedral element to design suitable refinement criteria to be possibly employed within adaptive refinement strategies. We test the proposed idea considering two families of finite element methods that support arbitrarily shaped polygonal and polyhedral elements, namely Polygonal Discontinuous Galerkin (PolyDG) methods and Virtual Element Methods (VEMs). We demonstrate that the proposed algorithms can greatly improve the performance of the discretization schemes in terms of accuracy and quality of the underlying grids. Moreover, since the training phase is performed off-line and independent of the differential model, overall computational costs are kept low.

Paola F. Antonietti
MOX-Dipartimento di Matematica
Politecnico di Milano
paola.antonietti@polimi.it

MS130

Recurrent Neural Networks in Adaptive Fem

We show that optimal mesh refinement algorithms for a large class of PDEs can be learned by a recurrent neural network with a fixed number of trainable parameters independent of the desired accuracy and the input size, i.e., number of elements of the mesh. This includes problems for which no optimal adaptive strategy is known yet. The proposed algorithm is problem independent in the sense that it only requires the current numerical approximation in order to optimally refine the mesh. Thus, the method is a provably optimal black-box mesh refinement tool for a wide variety of PDE problems.

Michael Feischl
TU Wien
michael.feischl@gmail.com

MS130

A Finite Element Based Deep Learning Solver for Parametric PDEs

We introduce a dynamic Deep Learning (DL) architecture based on the Finite Element Method (FEM) to solve linear parametric Partial Differential Equations (PDEs). The connections between neurons in the architecture mimic the Finite Element connectivity graph when applying mesh refinements. We select and discuss several losses employing preconditioners and different norms to enhance convergence. For simplicity, we implement the resulting DeepFEM in one spatial domain (1D), although its extension to 2D and 3D problems is straightforward. Extensive numerical experiments show in general good approximations for both symmetric positive definite (SPD) and indefinite problems in parametric and non-parametric problems. However, in some cases, lack of convexity prevents us from obtaining high-accuracy solutions.

Carlos Uriarte
Basque Center for Applied Mathematics
University of the Basque Country
carlos.uribar@gmail.com

David Pardo
Basque Center for Applied Mathematics
dzubiaur@gmail.com

Angel Javier Omella
Basque Center for Applied Mathematics (BCAM)
ajaome@gmail.com

MS130

Neural Network Guided Adjoint Computations in Dual Weighted Residual Error Estimation

In this presentation, we are concerned with neural net-

work guided goal-oriented a posteriori error estimation and adaptivity using the dual weighted residual method. The primal problem is solved using classical Galerkin finite elements. The adjoint problem is solved in strong form with a feedforward neural network using two or three hidden layers. The main objective of our approach is to explore alternatives for solving the adjoint problem with greater potential of a numerical cost reduction. The proposed algorithm is based on the general goal-oriented error estimation theorem including both linear and nonlinear stationary partial differential equations and goal functionals. Our developments are substantiated with some numerical experiments that include comparisons of neural network computed adjoints and classical finite element solutions of the adjoints. In the programming software, the open-source library deal.II is successfully coupled with LibTorch, the PyTorch C++ application programming interface.

Julian Roth, Max Schröder, Thomas Wick
Leibniz Universität Hannover
roth@ifam.uni-hannover.de,
max.felix.schroeder@outlook.com, thomas.wick@ifam.uni-hannover.de

MS131

Machine Learning Methods for Designing Thin Film Microstructures

Designing next generation thin films, tailor-made for specific applications, relies on the availability of robust relationships between processing conditions, resulting microstructure, and observed properties. Traditional mappings such as structure zone diagrams are limited to low-dimensional pictorial mappings, with machine-learning methods only recently attempting to relate multiple processing parameters to the final microstructure. Despite this progress, structure-processing relationships are unknown for processing conditions that vary during thin film deposition, limiting the range of microstructures and properties achievable. In this work, we explore a combination of machine learning and thin film deposition simulations to find novel processing protocols that achieve specific microstructures. We explore online learning, coupling the phase-field thin film deposition simulation with a genetic algorithm to find novel time-dependent protocols that achieve microstructures of interest. We then turn to offline learning for the same problem, replacing time-consuming deposition simulations with a generative model, using reinforcement learning to find a broader range of processing protocols. Additionally, we comment on the nature of these various machine-learning methods and their ability to represent the microstructures observed in this physical problem across various length and timescales.

Saaketh Desai
Sandia National Laboratories
saadesa@sandia.gov

MS131

Neural Network Approaches to Solving Low Dimensional Fokker Planck Equations

The Fokker-Planck equation describes the time evolution of the probability density of stochastic systems. Traditional PDE approaches are typically inadequate for solving even low dimensional systems due to the unbounded domain and the curse of dimensionality. Here we describe data-driven algorithms which combine sampling information with projection-based methods and neural network approaches for finding the target density.

Matthew Dobson
University of Massachusetts Amherst
dobson@math.umass.edu

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

Jiayu Zhai
University of Massachusetts Amherst
zhai@math.umass.edu

MS131

Learning Deep Neural Operators for Heterogeneous Material Modeling

Constitutive modeling based on continuum mechanics theory has been a classical approach for modeling the mechanical responses of materials. However, when constitutive laws are unknown or when defects and/or high degrees of heterogeneity are present, these classical models may become inaccurate. In this work, we propose to use data-driven modeling to predict a material's response without using conventional constitutive models. Specifically, the material response is modeled by learning the implicit mappings between loading conditions and the resultant displacement and/or damage fields, with the neural network serving as a surrogate for a solution operator. To model the complex responses due to material heterogeneity and defects, we develop a novel deep neural operator architecture, which we coin as the Implicit Fourier Neural Operator (IFNO). In the IFNO, the increment between layers is modeled as an integral operator to capture the long-range dependencies in the feature space. As the network gets deeper, the limit of IFNO becomes a fixed point equation that yields an implicit neural operator and naturally mimics the displacement/damage fields solving procedure in material modeling problems. As an application, we further employ the proposed approach to learn the material models directly from digital image correlation (DIC) tracking measurements, and show that the learned solution operators outperform the conventional constitutive models in predicting displacement fields.

Yue Yu
Department of Mathematics, Lehigh University
yuy214@lehigh.edu

MS132

Broader Engagement (BE): Fundamentals of Accelerated Data Science with RAPIDS - Part II of II

Traditional CPU-driven data science workflows can be cumbersome but with the power of GPUs, data scientists and Ph.D. students can make sense of data quickly to drive research decisions. In this workshop, student or developers will learn how to build and execute end-to-end GPU-accelerated data science workflows that enable them to quickly explore, iterate, and get their work into production. Using the RAPIDS accelerated data science libraries, developers will apply a wide variety of GPU-accelerated machine learning algorithms, including XGBoost, cuGRAPHs single-source shortest path, and cuMLs KNN, DBSCAN, and logistic regression to perform data analysis at scale. All workshop attendees get access to fully configured, GPU-

accelerated servers in the cloud, guidance from an instructor, and the opportunity to network with other developers, data scientists, and researchers. Attendees can earn a certificate to prove subject matter competency and support professional growth.

Xi Chen
University of Kentucky
billchenxi@uky.edu

MS133

Benign Overfitting Beyond the OLS

Benign overfitting, the phenomenon where interpolating models generalize well in the presence of noisy data, was first observed in neural network models trained with gradient descent. Recently, a large body of work has focussed on understanding this phenomenon in relatively simple linear models such as the minimum ℓ_2-norm interpolant (ordinary least squares (OLS) solution). In this talk I shall present two research vignettes that study benign overfitting in models beyond the OLS. First, to better understand the empirical observation that kicked off this line of research, we consider the generalization error of two-layer neural networks trained to interpolation by gradient descent on the logistic loss following random initialization. We assume the data comes from well-separated class-conditional log-concave distributions and allow for a constant fraction of the training labels to be corrupted by an adversary. We show that in this setting, neural networks exhibit benign overfitting: they can be driven to zero training error, perfectly fitting any noisy training labels, and simultaneously achieve test error close to the Bayes-optimal error. Second, we study the behaviour of sparse linear interpolators (such as the minimum ℓ_1-norm interpolant), and show that surprisingly the excess risk of sparse interpolators can be exponentially larger than the excess risk of the OLS (a dense interpolator), even when the optimal linear predictor is itself sparse.

Niladri Chatterji
Stanford University
niladri@cs.stanford.edu

MS133

Understanding Self-Supervised Learning: Analysis and Robustness to Imbalance Datasets

Self-supervised learning has made empirical breakthroughs in producing representations that can be applied to a wide range of downstream tasks. In this talk, I will primarily present a recent work that analyzes contrastive learning algorithms under realistic assumptions on the data distributions for vision applications. We show that contrastive learning can be viewed as a parametric version of spectral clustering on a so-called population augmentation graph, and analyze the linear separability of the learned representations and provide sample complexity bounds. I will also briefly another recent work that shows that self-supervised learning is more robust to dataset imbalance than supervised representations.

Tengyu Ma
Stanford University
tengyuma@stanford.edu

MS133

General Variational Formulation of Neural Networks with Universal Approximation Guarantees

A powerful framework for supervised learning is the minimization of a cost that consists of a data fidelity term plus a regularization functional. In this talk, I investigate the properties induced by a Radon-domain regularization functional (Ongie ICLR 2020; Parhi JMLR 2021), but which now depends on a more general (pseudo-)differential operator L. Under mild admissibility conditions on L, I prove that the solution set of the corresponding functional optimization problem is parameterized by a shallow neural network whose activation function is determined by L. I then show that the resulting architecture offers guarantees of universal approximation for a wide variety of activation functions, including sigmoids (Barron 1993) and cases such as ReLU where the activation increases polynomially. The proposed setting also explains the favorable role of bias and skip connections in neural architectures.

Michael Unser
EPFL
michael.unser@epfl.ch

MS133

Gradient Methods Provably Converge to Non-Robust Networks

Despite a great deal of research, it is still unclear why neural networks are so susceptible to adversarial examples. In this work, we identify natural settings where depth-2 ReLU networks trained with gradient flow are provably non-robust (susceptible to small adversarial l2-perturbations), even when robust networks that classify the training dataset correctly exist. Perhaps surprisingly, we show that the well-known implicit bias towards margin maximization induces bias towards non-robust networks, by proving that every network which satisfies the KKT conditions of the max-margin problem is non-robust.

Gal Vardi, Gilad Yehudai, Ohad Shamir
Weizmann Institute of Science
galvardi1@gmail.com, gilad.yehudai@weizmann.ac.il, ohad.shamir@weizmann.ac.il

MS134

A Factor Model of Multilayer Network Interdependence

Multilayer networks describe the rich ways in which nodes are related by expressing their connections through different types of relationships in separate layers. These multiple relationships are naturally represented by an adjacency tensor that can be interpreted using techniques from multilinear algebra. In this work we propose the use of the *nonnegative Tucker decomposition* (NNTuck) with KL-divergence as an expressive factor model for adjacency tensors. We discuss how the NNTuck generalizes existing methods for stochastic blockmodels of multilayer networks and provides an intuitive factor-based perspective on layer dependence. We propose a definition of layer dependence based on using a likelihood ratio test to evaluate NNTuck models corresponding to different layer dependence assumptions. Algorithmically, we find that using expectation maximization to maximize the log-likelihood under the NNTuck model is step-by-step equivalent to tensorial multiplicative updates for the NNTuck under a KL loss, extending a previously known equivalence from nonnegative matrices to nonnegative tensors. Using both synthetic and real-world data, we evaluate the use and interpretation

of the NNTuck as a generative model of multilayer networks. We propose a formal treatment of cross-validation for link-prediction tasks in multilayer settings by defining *tubular* and *independent* link-prediction tasks, and observe significant performance differences between the two.

Izabel Aguiar
Stanford University
izzya@stanford.edu

Dane Taylor
University at Buffalo, SUNY
danet@buffalo.edu

Johan Ugander
jugander@stanford.edu
Stanford University

MS134

Tensor Moments of Gaussian Mixture Models: Theory and Applications

Gaussian mixture models (GMMs) are fundamental tools in statistical and data sciences. We study the moments of multivariate Gaussians and GMMs. The d-th moment of an n-dimensional random variable is a symmetric d-way tensor of size n^d, so working with moments naively is assumed to be prohibitively expensive for $d > 2$ and larger values of n. In this work, we develop theory and numerical methods for *implicit computations* with moment tensors of GMMs. We derive concise analytic expressions for the moments in terms of symmetrized tensor products, relying on the correspondence between symmetric tensors and homogeneous polynomials, and combinatorial identities involving Bell polynomials. The primary application of this theory is to estimating GMM parameters (means and covariances) from a set of observations, when formulated as a moment-matching optimization problem. If there is a known and common covariance matrix, we also show it is possible to debias the data observations, in which case the problem of estimating the unknown means reduces to symmetric CP tensor decomposition. Numerical results validate and illustrate the numerical efficiency of our approaches. This work potentially opens the door to the competitiveness of the method of moments as compared to expectation maximization methods for parameter estimation of GMMs.

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

Joao M. Pereira
IMPA - Instituto de Matemática Pura e Aplicada
Rio de Janeiro, Brazil
jpereira@impa.br

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

MS134

Tensor Perron Eigenvectors for Classification and Ranking

Matrices are a very useful tool to represent and analyse networks represented as graphs. When dealing with higher-order network models (e.g. based on multiple layers or non-dyadic interactions), tensors provide the natural analogue of matrices. Unfortunately, unlike matrices, computing eigen and singular values of higher-order tensors is generally an NP-hard problem, which requires solving systems of homogeneous polynomial equations. However, when the tensor at hand has nonnegative entries, recent advances in nonlinear Perron-Frobenius theory based on multihomogeneous mappings provide us simple conditions that guarantee existence, uniqueness and computability of so-called tensor Perron eigen and singular values and vectors. In this talk I will review the main results in the resulting Perron-Frobenius theory for nonnegative tensors and I will discuss how suitable Perron vectors can be used to diffuse initially labeled points and thus classify the nodes as well as to define spectral measures of centrality for higher-order networks.

Francesco Tudisco
Gran Sasso Science Institute
School of Mathematics
francesco.tudisco@gssi.it

MS135

Generalization Bounds for Sparse Random Feature Expansions

Random feature methods have been successful in various machine learning tasks, are easy to compute, and come with theoretical accuracy bounds. They serve as an alternative approach to standard neural networks since they can represent similar function spaces without a costly training phase. However, for accuracy, random feature methods require more measurements than trainable parameters, limiting their use for data-scarce applications or problems in scientific machine learning. This work introduces the sparse random feature expansion to obtain parsimonious random feature models. Specifically, we leverage ideas from compressive sensing to generate random feature expansions with theoretical guarantees even in the data-scarce setting. In particular, we provide generalization bounds for functions in a certain class (that is dense in a reproducing kernel Hilbert space) depending on the number of samples and the distribution of features. The generalization bounds improve with additional structural conditions, such as coordinate sparsity, compact clusters of the spectrum, or rapid spectral decay. In particular, by introducing sparse features, i.e., features with random sparse weights, we provide improved bounds for low order functions. Finally, we show that the sparse random feature expansions outperforms shallow networks in several scientific machine learning tasks.

Abolfazl Hashemi
Purdue University
abolfazl@purdue.edu

MS135

Diffrax: JAX-Based Numerical Differential Equation Solvers

We introduce Diffrax, a JAX-based suite of ordinary/stochastic/controlled differential equation solvers. Highlights include: - High performance; 200 times faster than torchdiffeq and of similar speed to DifferentialEquations.j. - Numerous features: high-order solvers, implicit solvers, dense solutions, multiple adjoints methods, etc. - Integrates directly with JAX: jit/grad/vmap/etc. all work as normal. - Easily extensible with custom ops (solvers etc.); includes the ability to handle the step-

ping yourself if writing a differentiable simulator yourself. - The main technical novelty: ODEs and SDEs are all solved in a unified way, by lowering them to CDEs. Code: https://github.com/patrick-kidger/diffrax Documentation: https://docs.kidger.site/diffrax/

Patrick Kidger
University of Oxford
kidger@maths.ox.ac.uk

MS135

Theoretical Guarantees for Distribution Approximation and Sampling via Neural ODEs

Neural ordinary differential equations (neural ODEs) provide an invertible neural network architecture that can be used for density estimation, Bayesian inference, and generative modeling. In this work, we establish theoretical guarantees for distribution approximation using neural ODEs. In particular, we show that one can efficiently approximate a wide class of distributions by choosing the ODE velocity field from a class of multi-layer feedforward neural networks whose size can be explicitly bounded, with the accuracy of the distribution approximation measured by a variety of probability metrics. We also study the statistical consistency of neural ODEs as generative models, and obtain finite sample guarantees.

Robert Ren, Youssef M. Marzouk
Massachusetts Institute of Technology
zren@mit.edu, ymarz@mit.edu

Jakob Zech
Heidelberg University
jakob.zech@uni-heidelberg.de

Sven Wang
Massachusetts Institute of Technology
svenwang@mit.edu

MS135

Score-Based Generative Modeling with Stochastic and Ordinary Differential Equations

Generating data with complex patterns, such as images, audio, and molecular structures, requires fitting very flexible statistical models to the data distribution. Even in the age of deep neural networks, building such models is difficult because they typically require an intractable normalization procedure to represent a probability distribution. To address this challenge, we propose to model the vector field of gradients of the data distribution (known as the score function), which does not require normalization and therefore can take full advantage of the flexibility of deep neural networks. We will show how to (1) estimate the score function from data with flexible deep neural networks and efficient statistical methods, (2) generate new data by forming stochastic differential equations with the estimated score functions, and even (3) evaluate probability values accurately using ordinary differential equations built on score functions. The resulting method, called score-based generative modeling, achieves record-breaking performance in applications including image synthesis, text-to-speech generation, time series prediction, and point cloud generation, challenging the long-time dominance of generative adversarial networks (GANs) on many of these tasks.

Yang Song
Stanford University
thusongyang@gmail.com

MS136

Tensor Structured Sketching for Constrained Least Squares Problems

Constrained least squares problems arise in many applications. The memory and com- putation cost are usually expensive with high-dimensional input data. We employ the so-called sketching strategy to project the least squares problem into a space of a lower sketching dimension" via a random sketching matrix. The key idea of sketching is to reduce the dimension of the problem as much as possible while maintaining the approximation accuracy. In this talk, we will focusing on least square problems with tensor data matrix. Such structure is often present in linearized inverse PDE problems and tensor decomposition optimizations. To match with the tensor structures of the problem, we utilize a general class of row-wise tensorized sub-Gaussian matrices as sketching matrices. We provide theoretical guarantees on the sketching dimension in terms of error criterion and probability failure rate. For unconstrained linear regressions, we obtain an optimal estimate for the sketching dimension. For optimization problems with general constraint sets, we show that the sketching dimension depends on a statistical complexity that characterizes the geometry of the underlying problems. Our theories are demonstrated and validated in a few concrete examples, including unconstrained linear regression and sparse recovery problems.

Ke Chen
University of Texas Austin
kechen@math.utexas.edu

Ruhui Jin
University of Wisconsin
rjin@math.wisc.edu

MS136

Randomized Algorithms for Efficient Preconditioning and UQ in Inverse Transport Problems

We discuss different applications of randomized algorithms in inverse control problems. We are going to consider optimization problems governed by dynamical systems, in particular, transport equations. We will discuss how randomized algorithms allow us to design efficient strategies for preconditioning and uncertainty quantification.

Andreas Mang
University of Houston
Department of Mathematics
amang@Central.UH.EDU

Jae Youn Kim
Department of Mathematics
University of Houston
jaeyoun@math.uh.edu

MS137

Galerkin Neural Network Approximation of Multiscale Problems

We consider the neural network approximation of systems of partial differential equations exhibiting multiscale features. Specifically, we consider the approximation of the

Reissner-Mindlin plate model which poses significant challenges due to the presence of boundary layers and numerical phenomena such as locking. This work builds on the basic Galerkin Neural Network approach established in previous work for symmetric, positive-definite problems. The key contributions of this work are (1) the analysis and comparison of several new least squares-type variational formulations for the Reissner-Mindlin plate, and (2) their numerical approximation using the Galerkin Neural Network approach. Numerical examples are presented which demonstrate the ability of the approach to resolve multiscale phenomena such for the Reissner-Mindlin plate model for which we develop a new family of benchmark solutions which exhibit boundary layers.

Justin Dong
Brown University
justin_dong@brown.edu

MS138

Short Weather Forecast Ensembles Inform Long-Term Climatology of Sudden Stratospheric Warming

In some physical systems such as the Earth's atmosphere, the most extreme and rare events matter a great deal, both for their human impact and for their dynamical consequences. A prototypical example is sudden stratospheric warming (SSW), a rapid breakdown of the winter stratospheric polar vortex, which causes extreme mid-latitude cold spells and alters surface weather for months. The historical scarcity of observations, and an unusually SSW-rich 2000's decade, lead to uncertain SSW climatology: when do they occur, how often, and how predictably? Long, expensive model runs could answer these statistical questions, but with a tradeoff between cost and bias. We instead utilize weather forecast ensembles that are high-resolution, but short (subseasonal) in duration. A simple analogue forecasting procedure chains them together to estimate key climate statistics, such as annual frequencies and timing distributions of SSW events, as formulated in Transition Path Theory. Using forecast ensembles initialized between 1996 and 2018, we find that the SSW statistics match well with 20th-century reanalysis. Our method extrapolates the climatology well beyond what is possible with the short observational dataset that initialized the forecasts, yielding accurate estimates of once in a century events. This suggests exciting new uses for ensemble forecasts in rare event analysis.

Justin M. Finkel
University of Chicago
jfinkel@uchicago.edu

Edwin Gerber
NYU
epg2@nyu.edu

Dorian S. Abbot
Department of Geophsyical Sciences
University of Chicago
abbot@uchicago.edu

Jonathan Weare
New York University
Courant Institute of Mathematical Sciences
jqw200@nyu.edu

MS138

Monte Carlo and Markov Chain Monte Carlo in High-Dimensional, Nonlinear Problems in Earth Science

Earth science nearly always requires estimating models, or model parameters, from data. This could mean to infer the state of the southern ocean from ARGO floats, to compute the state of our atmosphere based on atmospheric observations of the past six hours, or to construct a resistivity model of the Earths subsurface from electromagnetic data. All these problems have in common that the number of unknowns is large (millions to hundreds of millions) and that the underlying processes are nonlinear. Due to the nonlinearity, Markov chain Monte Carlo (MCMC) is a good candidate for the numerical solution of such inverse problems. But MCMC is known to be slow when the number of unknowns is large. In this talk, I will point to two approaches that enable the solution of high-dimensional inverse problems in Earth science and both have in common that they rely on efficient, but biased estimation algorithms. The main message is that the unbiased solution of nonlinear, high-dimensional problems is and remains infeasible, but one can construct efficient and accurate biased estimators that are feasible to apply to high-dimensional problems.

Matthias Morzfeld
University of California, San Diego
Scripps Institution of Oceanography
mmorzfeld@ucsd.edu

MS138

The Stein Geometry in Machine Learning: Gradient Flows, Optimal Transport and Large Deviations

Sampling or approximating high-dimensional probability distributions is a key challenge in computational statistics and machine learning. This talk will present connections to gradient flow PDEs, optimal transport and interacting particle systems, focusing on the recently introduced Stein variational gradient descent methodology and some variations. The construction induces a novel geometrical structure on the set of probability distributions related to a positive definite kernel function. We discuss the corresponding geodesic equations, infinitesimal optimal transport maps, as well as large deviation functionals. This is joint work with A. Duncan (Imperial College London), L. Szpruch (University of Edinburgh) and M. Renger (Weierstrass Institute Berlin).

Nikolas Nüsken
Universität Potsdam
nuesken@uni-potsdam.de

MS138

Approximate Sampling Using Gaussian Projections

We present a novel method for sampling and optimization tasks based on a stochastic interacting particle system. We explain how this method can be used for the following two goals: (i) generating approximate samples from a given target distribution; (ii) optimizing a given objective function. The approach is derivative-free and affine invariant, and is therefore well-suited for solving inverse problems defined by complex forward models: (i) allows generation of samples

from the Bayesian posterior and (ii) allows determination of the maximum a posteriori estimator. We investigate the properties of the proposed method in terms of various parameter choices, both analytically and by means of numerical simulations. The analysis and numerical simulation establish that the method has potential for general purpose optimization tasks over Euclidean space; contraction properties of the algorithm are established under suitable conditions, and computational experiments demonstrate wide basins of attraction for various specific problems. The analysis and experiments also demonstrate the potential for the sampling methodology in regimes in which the target distribution is unimodal and close to Gaussian; indeed we prove that the method recovers a Laplace approximation to the measure in certain parametric regimes and provide numerical evidence that this Laplace approximation attracts a large set of initial conditions in a number of examples.

Carrillo José A
University of Oxford
carrillo@maths.ox.ac.uk

Franca Hoffmann
University of Bonn
franca.hoffmann@hcm.uni-bonn.de

Andrew Stuart
Computing + Mathematical Sciences
California Institute of Technology
astuart@caltech.edu

Urbain Vaes
Ecole des Points and Inria
urbain.vaes@inria.fr

MS139

Statistical Benefits of Convolutional Models: A Kernel Perspective

Many supervised learning problems involve high-dimensional data such as images, text, or graphs. In order to make efficient use of data, it is often useful to leverage priors in the problem at hand, such as invariance to certain transformations or dependence on local information. Empirically, deep convolutional networks have been very successful on such data, raising the question of how they are able to capture the structure of these problems for efficient learning. This work studies this question from a theoretical perspective using kernel methods, in particular convolutional kernels. These are constructed following similar architectural principles as convolutional networks, are closely related to their infinite-width limits in certain regimes, and provide good empirical performance on standard computer vision benchmarks such as Cifar10. I will present two contributions that highlight the statistical benefits of (deep) convolutional architectures for learning target functions with certain structures, in particular invariant/stable function, as well as additive interaction models with symmetries.

Alberto Bietti
Inria Grenoble, France
alberto.bietti@gmail.com

MS139

Deep Networks and the Multiple Manifold Problem

Data with low-dimensional nonlinear structure are ubiquitous in engineering and scientific problems. We study a model problem with such structure: a binary classification task that uses a deep fully-connected neural network to classify data drawn from two disjoint smooth curves on the unit sphere. Aside from mild regularity conditions, we place no restrictions on the configuration of the curves. We prove that when (i) the network depth is large relative to certain geometric properties that set the difficulty of the problem and (ii) the network width and number of samples is polynomial in the depth, randomly-initialized gradient descent quickly learns to correctly classify all points on the two curves with high probability. To our knowledge, this is the first generalization guarantee for deep networks with nonlinear data that depends only on intrinsic data properties. Our analysis draws on ideas from harmonic analysis and martingale concentration for handling statistical dependencies in the initial (random) network. We sketch applications to invariant vision, where leveraging low-dimensional structure leads to novel resource-efficient neural network architectures for detecting deformations of visual motifs.

Sam Buchanan
Toyota Technical Institute at Chicago
sambuchanan6@gmail.com

MS139

The Role of Data Geometry in High-Dimensional Learning

The intrinsic low-dimensionality of natural high-dimensional data plays a very important role in the success of modern data science, but mathematical understanding still lags behind practice. This talk will be an overview of this area of research and will serve as an introduction to the minisymposium. Topics will include the following: (1) Classical models for low dimensional structure in high dimensions as well as their limitations in practice; (2) Current understanding and open questions about the role data structure play in how high dimensional learning algorithms find estimators that both interpolate the data and generalize well.

Eliza W. O'Reilly
Caltech
eoreilly@caltech.edu

MS139

Training Data Size Induced Double Descent For Denoising Neural Networks and the Role of Training Noise Level

Training a network to denoise noisy inputs is the most widely used technique for pre-training deep neural networks. Hence one important question is the effect of scaling the number of training data points. We formalize the question of how many data points should be used by looking at the generalization error for denoising noisy test data. Prior work on computing the generalization error focus on adding noise to target outputs. However, adding noise to the input is more in line with current pre-training practices. In the linear (in the inputs) regime, we provide an asymptotically exact formula for the generalization error for rank 1 data and an approximation for the generalization error for rank r data. We show using our formulas, that the generalization error versus number of data points follows a double descent curve. From this, we derive a formula for the amount of noise that needs to be added to the training data to minimize the denoising error and see that

this follows a double descent curve as well.

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

MS140

Projected Variational Methods for High-Dimensional Bayesian Inference

In this talk, I will present a class of transport-based projected variational methods to tackle the computational challenges of the curse of dimensionality and unaffordable evaluation cost for high-dimensional Bayesian inverse problems governed by complex models. We project the high-dimensional parameters to intrinsically low-dimensional data-informed subspaces, and employ transport-based variational methods to push samples drawn from the prior to a projected posterior. Moreover, we employ fast surrogate models to approximate the parameter-to-observable map. I will present error bounds for the projected posterior distribution measured in Kullback–Leibler divergence. Numerical experiments will be presented to demonstrate the properties of our methods, including improved accuracy, fast convergence with complexity independent of the parameter dimension and the number of samples, strong parallel scalability in processor cores, and weak data scalability in data dimension.

Peng Chen
The University of Texas at Austin
peng@oden.utexas.edu

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

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

MS140

A Model-Constrained Bayesian Neural Networks for Bayesian Inverse Problems

Bayesian inverse problem for practical applications are notoriously difficult due to: 1) high dimensional parameter space, 2) high dimensional state space, and 3) (possibly) high dimensional data space. This talk presents a triple reduction approach to address these three challenges simultaneously. Several theoretical and numerical results will be presented to demonstrate the approach.

Russell Philley
University of Texas at Austin, U.S.
rsphilley@gmail.com

Tan Bui-Thanh, Hai Nguyen
The University of Texas at Austin
tanbui@ices.utexas.edu, hainguyen@utexas.edu

MS140

An Exponential Speedup in the Rigorous Operator Learning of Elliptic PDEs

The so-called "operator learning" of solution operators of partial differential equations (PDEs) from solution pairs has attracted considerable attention. In the case of elliptic PDEs, existing methods with rigorous convergence rates require $\mathrm{poly}(1/\epsilon)$ solution pairs to achieve an ϵ-accurate approximation of the solution operator. In the present work, we achieve an exponential improvement over these methods by proposing an algorithm that can recover the discretized solution operators of general elliptic PDEs on a d-dimensional domain to accuracy ϵ from only $\mathcal{O}\left(\log\left(N\right)\log^{d}\left(N/\epsilon\right)\right)$ solution pairs selected a-priori. Here, N is the number of degrees of freedom of the discrete function space. Our method has computational cost $\mathcal{O}\left(N\log^{2}\left(N\right)\log^{2d}\left(N/\epsilon\right)\right)$ and returns a sparse Cholesky factor with $\mathcal{O}\left(N\log\left(N\right)\log^{d}\left(N/\epsilon\right)\right)$ nonzero entries. This Cholesky factor can be interpreted as a transport map that maps a standard Gaussian vector to the Gaussian process associated with the PDE. Prior work on the conditional independence properties of these Gaussian processes allows us to rigorously prove the error-vs-complexity bounds mentioned above. We further provide numerical experiments confirming the practicality of our algorithm.

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

Houman Owhadi
Applied Mathematics
Caltech
owhadi@caltech.edu

MS140

Mcmc Procedures for Dimension Reduction with Theoretical Guarantees

The likelihood-informed subspace (LIS) method offers a viable route to reducing the dimensionality of high-dimensional probability distributions arising in Bayesian inference. Intuitively, the accuracy of the LIS approximation, and hence the performance of the inference algorithms, are influenced by three factors—the dimension truncation error in identifying the subspace, Monte Carlo error in estimating the Gram matrices, and Monte Carlo error in constructing marginalizations. This talk discusses how to analyze each of these three factors and their interplay. We also discuss how to extend these results to problems with heavy tailed priors using transformation method.

Xin Tong
National University of Singapore
xin.t.tong@nus.edu.sg

Tiangang Cui
Monash University
tiangang.cui@monash.edu

Olivier Zahm
Univ. Grenoble Alpes, Inria
olivier.zahm@inria.fr

MS141

Hamilton-Jacobi Equations on Graphs with Applications to Semi-Supervised Learning and Data Depth

Shortest path graph distances are widely used in data science and machine learning, since they can approximate the

underlying geodesic distance on the data manifold. However, the shortest path distance is highly sensitive to the addition of corrupted edges in the graph, either through noise or an adversarial perturbation. In this talk we present a family of Hamilton-Jacobi equations on graphs that we call the p-eikonal equation. We show that the p-eikonal equation with p=1 is a provably robust distance-type function on a graph, and the limiting case in which p goes to infinity recovers shortest path distances. While the p-eikonal equation does not correspond to a shortest-path graph distance, we nonetheless show that the continuum limit of the p-eikonal equation on a random geometric graph recovers a geodesic density weighted distance in the continuum. At the end we present application of the p-eikonal equation to data depth and semi-supervised learning, and use the continuum limit to show asymptotic consistency results for both applications. This is a joint work with Jeff Calder from University of Minnesota.

Mahmood Ettehad
University of Minnesota, Twin Cities
etteh001@umn.edu

MS141

Ensuring Exploration and Exploitation in Graph-Based Active Learning

Uncertainty sampling has traditionally been the de facto, simplest acquisition function for active learning in semi-supervised learning. Comparatively cheap to compute and straightforward to interpret, uncertainty sampling has been known to suffer from myopic sampling bias that fails to properly explore the extent of geometric structure of the dataset prior to exploiting learning decision boundaries. As such, most work in active learning for graph-based learning has focused on the design of more intricate acquisition functions that are explorative in nature, though are almost always more costly to compute. We show that exploration and exploitation in graph-based semi-supervised learning can be achieved by ensuring that the underlying graph-based model reflects uncertainty at unlabeled nodes that in turn captures the underlying clustering structure of the dataset. We present theoretical results and empirical evidence to demonstrate that reweighted Laplacian learning, which has a well-defined continuum limit model, allows us to ensure exploration and exploitation by simply applying uncertainty sampling.

Kevin Miller
University of California, Los Angeles
ksmill327@gmail.com

MS141

A Variational Approach to Multi-Dimensional Scaling

Dimension reduction is a crucial pre-processing step in many learning algorithms, and can be seen as a particular form of metric or graph embedding problem. Despite the widespread utilization of these algorithms, many of the non-linear versions of these algorithms lack rigorous theory. This talk will discuss recent work, joint with Adam Pickarski, on multi-dimensional scaling, one of the most basic forms of non-linear dimension reduction. In particular, we utilize a variational approach to derive regularity properties of the embeddings selected a simple version of multi-dimensional scaling, and to provide new asymptotic consistency results for the algorithm.

Ryan Murray
North Carolina State University
rwmurray@ncsu.edu

MS141

Unlabeled Data Help in Graph-Based Bayesian Semi-Supervised Learning

In this talk we consider the graph-based approach to semi-supervised learning in a Bayesian perspective. Under a manifold assumption, we shall demonstrate that the unlabeled data help in constructing a suitable prior that leads to improved posterior performance. In particular, we will see that optimal posterior contraction rates can be achieved if the number of unlabeleddata scales polynomiallywith respect to the labeled ones.

Daniel Sanz-Alonso, Ruiyi Yang
University of Chicago
sanzalonso@uchicago.edu, yry@uchicago.edu

MS142

Reduced Basis Methods Applied to Many-Body Physics Problems

We present the reduced basis method (RBM) as a tool for developing emulators for equations with tunable parameters within the context of the nuclear many-body problem. The RBM uses a basis expansion informed by a set of solutions for a few values of the model parameters and then projects the equations over a well-chosen low-dimensional subspace. We connect some of the results in the eigenvector continuation literature (a set of dimensionality-reduction techniques independently discovered in the nuclear theory community) to the formalism of RBMs and show how RBMs can be applied to a broad set of problems. As we illustrate, the possible success of the RBM on such problems can be diagnosed beforehand by a principal component analysis. We apply the RBM to the one-dimensional Gross-Pitaevskii equation with a harmonic trapping potential and to nuclear density functional theory for ^{48}Ca. The outstanding performance of the approach, together with its straightforward implementation, show promise for its application to the emulation of computationally demanding calculations, including uncertainty quantification.

Edgard Bonilla
Stanford University
edgard@stanford.edu

MS142

Wassmap: Wasserstein Isometric Mapping for Image Manifold Learning

We propose Wasserstein Isometric Mapping (Wassmap), a parameter-free nonlinear dimensionality reduction technique that provides solutions to some drawbacks in existing global nonlinear dimensionality reduction algorithms in imaging applications. Wassmap represents images via probability measures in Wasserstein space, then uses pairwise quadratic Wasserstein distances between the associated measures to produce a low-dimensional, approximately isometric embedding. We show that the algorithm is able to exactly recover parameters of some image manifolds including those generated by translations or dilations of a fixed generating measure. Additionally, we show that a discrete version of the algorithm retrieves parameters

from manifolds generated from discrete measures by providing a theoretical bridge to transfer recovery results from functional data to discrete data. Testing of the proposed algorithms on various image data manifolds show that Wassmap yields good embeddings compared with other global techniques.

Nick Henscheid
University of Arizona
nhenscheid@math.arizona.edu

Keaton Hamm, Shujie Kang
University of Texas at Arlington
keaton.hamm@uta.edu, shujie.kang@uta.edu

MS142

The Multimarginal Optimal Transport Formulation of Adversarial Multiclass Classification

In this talk, I will introduce a family of adversarial multiclass classification problems and provide equivalent reformulations in terms of: 1) a new generalized version of the Wasserstein barycenter problem, and 2) a family of multimarginal optimal transport problems where the number of marginals is equal to the number of classes in the original classification problem. The direct computational implication of these results is that by solving either the barycenter problem and its dual, or the MOT problem and its dual, we can recover the solution to the robust classification problem and the optimal adversarial action. Our results generalize recent results in the literature establishing connections between adversarial learning and optimal transport in the binary case. Several examples with synthetic and real data will illustrate the results. Joint work with Nicolas Garcia-Trillos and Jakwang Kim.

Matt Jacobs
Purdue University
jacob225@purdue.edu

MS142

Manifold Learning with Non-Euclidean Norms

Manifold learning methods play a prominent role in nonlinear dimensionality reduction and other tasks involving high-dimensional data sets with low intrinsic dimensionality. Many of these methods are graph-based: they associate a vertex with each data point and a weighted edge with each pair. Prior theory shows that the Laplacian matrix of the graph converges to the LaplaceBeltrami operator of the data manifold, under the assumption that the pairwise affinities are based on the Euclidean norm. In this talk, we determine the limiting differential operator for graph Laplacians constructed using any norm. Our proof involves an interplay between the second fundamental form of the manifold and the convex geometry of the given norms unit ball. To demonstrate the potential benefits of non-Euclidean norms in manifold learning, I will consider the task of mapping the motion of large molecules with continuous variability. In a numerical simulation we show that a modified Laplacian eigenmaps algorithm, based on the Earthmovers distance, outperforms the classic Euclidean Laplacian eigenmaps, both in terms of computational cost and the sample size needed to recover the intrinsic geometry. This talk is based on joint work with Amit Moscovich, Nathan Zelesko and Amit Singer.

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

MS143

Computational Inference of Population Models Based on Genomic Data

Sequencing the genomes of many individuals in one or multiple populations is now standard. In contrast, analysis of these large data sets is still problematic. Today, two approaches are used: site-frequency-spectra summary statistic as the data and full probabilistic inference using sequence data. The site frequency spectra use expectations derived from the coalescent assuming all data is only two-allelic. In contrast, the probabilistic inference separates the coalescent (inference of relationship among individuals) from the mutation model (change of particular sites on the genome due to random mutation over evolutionary time). We compare results of the different approaches using different amounts of data and different complexities of the model (population structure + mutation model), suggesting that even so full probabilistic models are slow compared to the summary statistic methods, they have the potential to be more accurate.

Peter Beerli
Florida State University
School of Computational Science
beerli@fsu.edu

MS143

Visualization of Tree Space and Finding Best Trees

Phylogenetic trees are fundamental for understanding the evolutionary history of a set of species. Understanding the local neighborhoods of a phylogenetic tree is essential, but since trees are high-dimensional objects, discussing these neighborhoods is difficult. Based on the geodesic distance between pairs of trees, we describe a method to generate intermediate trees on the shortest path between two arbitrary trees, called pathtrees. These pathtrees allow characterizing the Bilera-Holmes-Vogtman tree space and can also be used to find high likelihood trees independently of traditional heuristic search mechanisms. We implemented our algorithm in the Python package PATHTREES which enables the construction of the continuous tree landscape interior of the convex hull of starting trees, low-dimensional visualization of the generated tree landscape using multidimensional scaling (MDS), interpolating the log-likelihood values between trees, identifying clusters of trees with the same topologies, and searching for the best tree.

Marzieh Khodaei
Florida State Unversity
Scientific Computing
mk16e@fsu.edu

MS143

Fractional Coalescent

The fractional coalescent is a generalization of Kingman's n-coalescent. It facilitates the development of the theory of population genetic processes that deviate from Poisson-distributed waiting times. It also marks the first use of methods developed in fractional calculus, in population genetics. The fractional coalescent is an extension of the Canning's model, where the variance of the number of offspring per parent is a random variable.

Somayeh Mashayekhi

Kennesaw State University
smashay1@kennesaw.edu

MS143

Advances for Massive Viral Trees in Bayesian Phylogenetics

Comparative biologists are often interested in inferring covariation between multiple biological traits sampled across numerous related taxa. Existing control techniques for shared evolutionary history almost universally scale poorly as the number of taxa increases. An additional challenge arises as obtaining a full suite of measurements becomes increasingly difficult with increasing taxa. This typically necessitates data imputation or integration that further exacerbates scalability. We propose an inference technique that integrates out missing measurements analytically and scales linearly with the number of taxa by using a post-order traversal algorithm under a multivariate Brownian diffusion (MBD) model to characterize trait evolution. We further exploit this technique to extend the MBD model to account for sampling error or non-heritable residual variance. We test these methods to examine mammalian life history traits, prokaryotic genomic and phenotypic traits, and HIV infection traits. We find computational efficiency increases that top two orders-of-magnitude over current best practices.

Marc A. Suchard
David Geffen School of Medicine at UCLA
msuchard@ucla.edu

MS144

Bi-Fidelity Training of Neural Networks and Neural Operators

Neural networks have recently been at the forefront of scientific machine learning (SciML) research because of their capability of accurately representing a functional relationship between input and a quantity of interest. However, these networks require a large training dataset for accurate prediction. For engineering systems, we often have multiple models to describe the system. Some models, known as high-fidelity, describe the behavior with higher levels of accuracy but, in general, are expensive to simulate. On the other hand, inexpensive models, known as low-fidelity, often produce inaccurate predictions. In this talk, we will introduce training methods based on transfer learning for uncertainty quantification of engineering systems, which ameliorate the cost of generating a large training dataset from accurate high-fidelity models by using a combination of datasets generated from low- and high-fidelity models. We will also discuss the effect of regularization on the training of neural networks using a dataset from a low-fidelity model. In a recent development, neural network operators, such as the Deep Operator Network (DeepONet), are used to approximate solution operators associated with governing equations of physical systems independent of the finite element mesh used to generate the training dataset. We will also overview training methods for these operator networks in the presence of bi-fidelity datasets.

Subhayan De, Alireza Doostan
Department of Aerospace Engineering Sciences
University of Colorado, Boulder
subhayan.de@colorado.edu,
Alireza.Doostan@Colorado.EDU

MS144

Transfer Learning of Closure Terms and in Reduced Order Models of Chemically Reactive Flows

One of the primary challenges facing numerical modeling of turbulent reactive flows is associated with the high dimensionality of thermo-chemical scalars and the associated chemical stiffness. Principal component analysis (PCA) has recently been proposed as a data-driven reduced order modeling strategy to reduce the oftentimes high-dimensional composition space of the combustion systems to a low-dimensional principal component (PC) space. The transport equations for the PCs are solved in reformulated PCA-based workflows, in which the closure source terms and diffusion coefficients for the PCs are tabulated using artificial neural networks (ANNs). The training of such ANNs require training data obtained from a priori (offline) forward simulations in the original high-dimensional setting, with an associated prohibitive computational expense. To alleviate the aforementioned computational demands in the training of ANNs, we propose the use of probabilistic transfer learning whereby pre-trained neural network models from other similar tasks (operational regimes) can be leveraged for a task of interest. We will present results demonstrating the advantages of transfer learning across varying operating conditions for homogeneous hydrogen/air mixture combustion simulations.

Kisung Jung
Sandia National Labs
kjung@sandia.gov

Tarek Echekki
North Carolina State University
techekk@ncsu.edu

Jacqueline Chen, Mohammad Khalil
Sandia National Laboratories
jhchen@sandia.gov, mkhalil@sandia.gov

MS144

Multi-Fidelity Bayesian Neural Networks for Inverse Problems with Noisy Data

We propose a new class of Bayesian neural networks that can be trained using noisy data of variable fidelity, and we apply them to learn function approximations as well as to solve inverse problems based on PDES. These multi-fidelity BNNs consist of three neural networks: The first is a fully connected neural network, which is trained following the maximum a posteriori probability (MAP) method to fit the low-fidelity data; the second is a Bayesian neural network employed to capture the cross-correlation with uncertainty quantification between the low- and high-fidelity data; and the last one is the physics-informed neural network, which encodes the physical laws described by PDEs. We demonstrate the accuracy of the present method using synthetic data as well as real measurements. Specifically, we first approximate a one- and four-dimensional function, and then infer the reaction rates in one- and two-dimensional diffusion-reaction systems. Moreover, we infer the sea surface temperature (SST) in the Massachusetts and Cape Cod Bays using satellite images and in-situ measurements. Taken together, our results demonstrate that the present method can capture both linear and nonlinear correlation between the low- and high-fidelity data adaptively, identify unknown parameters in PDEs, and quantify

uncertainties in predictions, given a few scattered noisy high-fidelity data.

Xuhui Meng, George E. Karniadakis
Brown University
xuhui_meng@brown.edu, george karniadakis@brown.edu

MS144

Transfer Learning of Gaussian Processes to Capture Unmodeled Physics: Application to Control of Nonlinear Dynamical Systems

Gaussian process regression is a machine learning methodology gaining popularity because of its ability to consolidate prior knowledge and observed data in a Bayesian manner. This capability allows Gaussian process regression to estimate unmodeled phenomena in physics-based predictive modeling and simulation. However, much like most machine learning algorithms, Gaussian process regression can become prohibitive in its requirement of training data volume and computational resources. On the one hand, sparse data sets lack representational richness to train models with sufficient predictive accuracy. Conversely, training on large data sets is sometimes not feasible due to the poor scalability of computational resources as a function of the data set volume. We, therefore, propose to leverage a novel probabilistic transfer learning strategy whereby knowledge gained through Gaussian process modeling on similar source tasks is transferred to a novel task of interest. This learning scheme will assess the similarity of the source and target tasks, both for the priors and posteriors, and determine the optimal amount of knowledge transfer. We will apply the methodology to a vehicle control problem, whereby Gaussian process regression is employed to learn unmodeled nonlinear effects and transfer learning is utilized to alleviate the data sparsity and computational complexity challenges across vehicles of similar types of unmodeled physics.

Erin Swansen, Aditya Gahlawat, Naira Hovakimyan
University of Illinois at Urbana-Champaign
swansen2@illinois.edu, gahlawat@illinois.edu,
nhovakim@illinois.edu

Mohammad Khalil
Sandia National Laboratories
mkhalil@sandia.gov

MS145

Modeling and Learning Paradigms for Learning to Optimize

Optimization is a ubiquitous modeling tool and is often deployed in settings which repeatedly solve similar instances of the same problem. Amortized optimization methods use learning to predict the solutions to problems in these settings. This leverages the shared structure between similar problem instances. In this talk, we will discuss the key design choices behind amortized optimization, roughly categorizing 1) models into fully-amortized and semi-amortized approaches, and 2) learning methods into regression-based and objective-based. We then view existing applications through these foundations to draw connections between them, including for manifold optimization, variational inference, sparse coding, meta-learning, control, reinforcement learning, convex optimization, and deep equilibrium networks. This framing enables us easily see, for example, that the amortized inference in variational autoencoders is conceptually identical to value gradients in control and reinforcement learning as they both use fully-amortized models with an objective-based loss. This talk will cover topics from the tutorial: https://arxiv.org/abs/2202.00665

Brandon Amos
Facebook AI Research
brandon.amos.cs@gmail.com

MS145

Shine: Sharing the Inverse Estimate from the Forward Pass for Bi-Level Optimization and Implicit Models

Implicit deep learning has emerged as a method to increase the effective depth of deep neural networks. While their training is memory-efficient, they are still slower to train than their explicit counterparts. In Deep Equilibrium Models (DEQs), the training is performed as a bi-level problem, and its computational complexity is partially driven by the iterative inversion of a huge Jacobian matrix. In this paper, we propose a novel strategy to tackle this computational bottleneck from which many bi-level problems suffer. The main idea is to use the quasi-Newton matrices from the forward pass to efficiently approximate the inverse Jacobian matrix in the direction needed for the gradient computation. We provide a theorem that motivates using our method with the original forward algorithms. In addition, by modifying these forward algorithms, we further provide theoretical guarantees that our method asymptotically estimates the true implicit gradient. We empirically study this approach and the recent Jacobian-Free method in different settings, ranging from hyperparameter optimization to large Multiscale DEQs (MDEQs) applied to CIFAR and ImageNet. Both methods reduce significantly the computational cost of the backward pass. While SHINE has a clear advantage on hyperparameter optimization problems, both methods attain similar computational performances for larger scale problems such as MDEQs at the cost of a limited performance drop compared to the original models.

Zaccharie Ramzi
CNRS
ENS Ulm
zaccharie.ramzi@gmail.com

MS145

Deep Equilibrium Architectures for Inverse Problems in Imaging

Recent efforts on solving inverse problems in imaging via deep neural networks use architectures inspired by a fixed number of iterations of an optimization method. The number of iterations is typically quite small due to difficulties in training networks corresponding to more iterations; the resulting solvers cannot be run for more iterations at test time without incurring significant errors. I will describe an alternative approach corresponding to an infinite number of iterations, yielding up to a 4dB PSNR improvement in reconstruction accuracy above state-of-the-art alternatives and where the computational budget can be selected at test time to optimize context-dependent trade-offs between accuracy and computation. The proposed approach leverages ideas from Deep Equilibrium Models, where the fixed-point iteration is constructed to incorporate a known forward model and insights from classical optimization-based reconstruction methods. This is joint work with Davis Gilton

and Greg Ongie.

Rebecca Willett
University of Chicago
willett@uchicago.edu

MS145

Efficient Training of Implicit Networks via Jacobian-Free Backpropagation

A promising trend in deep learning replaces fixed depth models by approximations of the limit as network depth approaches infinity. This approach uses a portion of network weights to prescribe behavior by defining a limit condition. This makes network depth implicit, varying based on the provided data and an error tolerance. Moreover, existing implicit models can be implemented and trained with fixed memory costs in exchange for additional computational costs. In particular, backpropagation through implicit depth models requires solving a Jacobian-based equation arising from the implicit function theorem. We propose a new Jacobian-free backpropagation (JFB) scheme that circumvents the need to solve Jacobian-based equations while maintaining fixed memory costs. This makes implicit depth models much cheaper to train and easy to implement. Numerical experiments on classification and CT reconstructions are provided.

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

MS146

Learning and Control of Hamiltonian Dynamics on the SE(3) Manifold

Accurate models of robot dynamics are critical for safe and stable control and generalization to novel operational conditions. Hand-designed models, however, may be insufficiently accurate, even after careful parameter tuning. This motivates the use of machine learning techniques to approximate the robot dynamics over a training set of state-control trajectories. The dynamics of many robots, including ground, aerial, and underwater vehicles, are described in terms of their SE(3) pose and generalized velocity, and satisfy conservation of energy principles. In this talk, we present a Hamiltonian formulation over the SE(3) manifold of the structure of a neural ordinary differential equation (ODE) network to approximate the dynamics of a rigid body. In contrast to a black-box ODE network, our formulation guarantees total energy conservation by construction. We develop energy shaping and damping injection control for the learned, potentially underactuated SE(3) Hamiltonian dynamics to enable a unified approach for stabilization and trajectory tracking with various platforms, including pendulum, rigid-body, and quadrotor systems.

Thai Duong, Nikolay Atanasov
University of California San Diego
tduong@eng.ucsd.edu, natanasov@eng.ucsd.edu

MS146

Symplectic Model Reduction of Hamiltonian Systems on Nonlinear Manifolds

Classical model reduction techniques project the governing equations onto linear subspaces of the high-dimensional state-space. However, for problems with slowly decaying Kolmogorov-n-widths such as certain transport-dominated problems, classical linear-subspace reduced order models (ROMs) of low dimension might yield inaccurate results. Thus, the reduced space needs to be extended to more general nonlinear manifolds. Moreover, as we are dealing with Hamiltonian systems, it is crucial that the underlying symplectic structure is preserved in the reduced model. To the best of our knowledge, existing literatures addresses either model reduction on manifolds or symplectic model reduction for Hamiltonian systems, but not their combination. In this talk, we bridge these two approaches by providing a novel projection technique called symplectic manifold Galerkin, which projects the Hamiltonian system onto a nonlinear symplectic trial manifold such that the ROM is again a Hamiltonian system. We provide numerical results which demonstrate the ability of the method to outperform linear-subspace ROMs.

Silke M. Glas
University of Twente
s.m.glas@utwente.nl

MS146

Hamiltonian Operator Inference: Physics-Preserving Learning of Reduced-Order Models for Canonical Hamiltonian Systems

We present a nonintrusive physics-preserving method to learn reduced-order models (ROMs) of Hamiltonian systems. The proposed Hamiltonian operator inference approach embeds the physics into the operator inference framework to develop a data-driven model reduction method that preserves the underlying symplectic structure. Our method exploits knowledge of the Hamiltonian functional to define and parametrize a Hamiltonian ROM form which can then be learned from data projected via symplectic projectors. The proposed method is 'gray-box' in that it utilizes knowledge of the Hamiltonian structure at the partial differential equation level, as well as knowledge of spatially local components in the system. However, it does not require access to computer code, only data to learn the models. Our numerical results demonstrate Hamiltonian operator inference on a linear wave equation, the cubic nonlinear Schroedinger equation, and a nonpolynomial sine-Gordon equation. Accurate long-time predictions far outside the training time interval for nonlinear examples illustrate the generalizability of our learned models.

Boris Kramer, Harsh Sharma
University of California San Diego
bmkramer@ucsd.edu, hasharma@ucsd.edu

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

MS146

Leveraging Hamiltonian Structure to Learn Rigid Body Dynamics from Images

In learning Hamiltonian dynamics, we can exploit our knowledge of physics to improve the efficiency and accuracy of our learned models. We explore use of physics-informed datasets and loss functions to learn 3D rigid-body rotational dynamics from different forms of data (states, images, etc.). Here we will focus on learning dynamics from images, as it is a topic with wide applicability (e.g. situational awareness, visual-based planning, and control). We introduce a framework developed for learning 3D ro-

tational dynamics of rigid bodies from images. We encode images into a physically interpretable latent space and parametrize the systems Hamiltonian using a neural network. The learned equations of motion predict sequential future latent states and images. This interpretable latent space allows for applications to control. Our model is trained on a dataset we developed that consists of trajectories for a variety of 3D objects that rotate according to rigid-body dynamics. We will focus on introducing the structure and mechanics of our approach, our models performance on the synthetic dataset, as well as future applications to real-world data and tasks.

Justice Mason
Princeton University
jjmason@princeton.edu

MS147

Learning the Constraints in Adaptive Domain Decomposition Methods

n/a

Axel Klawonn
Universität zu Köln
axel.klawonn@uni-koeln.de

Martin Lanser
Universitaet zu Koeln
Mathematisches Institut
martin.lanser@uni-koeln.de

Janine Weber
Universität zu Köln
Department of Mathematics and Computer Science
janine.weber@uni-koeln.de

MS148

Efficient Nonlinear Manifold Reduced Order Models with Shallow Sparse Neural Network

Traditional linear subspace reduced order models (LS-ROMs) are able to accelerate physical simulations, in which the intrinsic solution space falls into a subspace with a small dimension, i.e., the solution space has a small Kolmogorov n-width. However, for physical phenomena not of this type, e.g., any advection-dominated flow phenomena, such as in traffic flow, atmospheric flows, and air flow over vehicles, a low-dimensional linear subspace poorly approximates the solution. To address cases such as these, we have developed a fast and accurate physics-informed neural network ROM, namely nonlinear manifold ROM (NM-ROM), which can better approximate high-fidelity model solutions with a smaller latent space dimension than the LS-ROMs. Our method takes advantage of the existing numerical methods that are used to solve the corresponding full order models. The efficiency is achieved by developing a hyper-reduction technique in the context of the NM-ROM. Numerical results show that neural networks can learn a more efficient latent space representation on advection-dominated data from 1D and 2D Burgers' equations. A speedup of up to 2.6 for 1D Burgers' and a speedup of 11.7 for 2D Burgers' equations are achieved with an appropriate treatment of the nonlinear terms through a hyper-reduction technique. Finally, a posteriori error bounds for the NM-ROMs are derived that take account of the hyper-reduced operators.

Youngsoo Choi
Lawrence Livermore National Laboratory
choi15@llnl.gov

MS148

Bayesian Reinforcement Learning for Optimal Sensor Relocation in Convection-Diffusion Fields

We present a Bayesian reinforcement learning approach to optimally place and relocate sensors. We formulate this sequential design problem as a finite-horizon partially observable Markov decision process (POMDP) in a Bayesian setting and with information-theoretic utilities. In particular, we seek to handle continuous random variables, general non-Gaussian posteriors, and expensive nonlinear forward models. The optimal design policy then incorporates elements of both feedback and lookahead, generalizing the suboptimal batch and greedy design strategies. We solve for the optimal policy numerically via policy gradient (PG) techniques, and derive and prove the PG expression. Adopting an actor-critic approach, we parameterize the policy and value functions using deep neural networks and improve them using gradient estimates produced from simulated episodes of designs and observations. The overall method is then applied to a sensor relocation problem for contaminant source inversion in a convection-diffusion field, illustrating its advantages over batch and greedy designs.

Wanggang Shen, Xun Huan
University of Michigan
wgshen@umich.edu, xhuan@umich.edu

MS148

DPM: A Novel Training Method for Physics-Informed Neural Networks in Extrapolation

We present a method for learning dynamics of complex physical processes described by time-dependent nonlinear partial differential equations (PDEs). Our particular interest lies in extrapolating solutions in time beyond the range of temporal domain used in training. Our choice for a baseline method is physics-informed neural network (PINN)[Raissi et al., J. Comput. Phys., 378: 686707, 2019] because the method parameterizes not only the solutions, but also the equations that describe the dynamics of physical processes. We demonstrate that PINN performs poorly on extrapolation tasks in many benchmark problems. To address this, we propose a novel method for better training PINN and demonstrate that our newly enhanced PINNs can accurately extrapolate solutions in time. Our method shows up to 72% smaller errors than existing methods in terms of the standard L2-norm metric.

Kookjin Lee
Arizona State University
klee263@asu.edu

MS148

A Learning Based Projection Method for Model Order Reduction

The Kolmogorovn-width of the solution manifolds of transport-dominated problems can decay slowly. As a result, it can be challenging to design efficient and accurate reduced order models (ROMs) for such problems. To address this issue, we propose a new learning-based projection method to construct nonlinear adaptive ROMs for transport problems. The construction follows the offline-

online decomposition. In the offline stage, we train a neural network to construct adaptive reduced basis dependent on time and model parameters. In the online stage, we project the solution to the learned reduced manifold. Inheriting the merits from both deep learning and the projection method, the proposed method is more efficient than the conventional linear projection-based methods, and may reduce the generalization error of a solely learning-based ROM.

Zhichao Peng
Michigan State University
pengzhic@msu.edu

Min Wang
Duke University
wangminhelen@gmail.com

Fengyan Li
Rensselaer Polytechnic Institute
lif@rpi.edu

MS149

BE: Lightning Talks

Participants in the Broader Engagement program will have a chance to share their research and prepare for the poster session during this Lightning Talk session. All participants presenting a poster or talk at the SIAM MDS22 conference will have a chance to share their work in a lightning round session. Participants will field questions from the audience and will be provided with constructive feedback on their presentation.

Sreeranjani Ramprakash
Argonne National Laboratory, U.S.
ramprakash@anl.gov

MS150

Do Input Gradients Highlight Discriminative Features?

Post-hoc gradient-based attribution methods are regularly used to identify the key discriminative features for a model. But, due to lack of ground truth, a thorough evaluation of even the most basic input gradient attribution method is still missing in literature. Our main contribution in this work is to overcome this challenge through experiments and theory on real and designed datasets. Our results demonstrate that (i) input gradient attribution does NOT highlight correct features on standard models (i.e., trained on original data) but surprisingly, it does highlight correct features on adversarially trained models (i.e., trained using adversarial training) and (ii) "feature leakage", which refers to the phenomenon wherein, given an instance, its input gradients highlight the location of discriminative features in the given instance as well as in other instances that are present in the dataset, is the reason behind why input gradient attribution fails for standard models. Our work raises more questions than it answers, so we will end with interesting directions for future work.

Praneeth Netrapalli
Google Research India
pnetrapalli@google.com

MS150

From Shapley back to Pearson: Hypothesis Testing via the Shapley Value

While neural networks have achieved remarkable performance in many settings, their complex nature raises concerns on their reliability, trustworthiness, and fairness in real-world scenarios. As a result, several a-posteriori explanation methods have been proposed to highlight the features that influence a model's prediction. Notably, the Shapley value–a game theoretic quantity that satisfies several desirable properties–has gained popularity in the machine learning explainability literature. More traditionally, however, feature importance in statistical learning has been formalized by conditional independence, and a standard way to test for it is via Conditional Randomization Tests (CRTs). So far, these two perspectives on interpretability and feature importance have been considered distinct and separate. In this work, we show that Shapley-based explanation methods and conditional independence testing for feature importance are closely related. More precisely, we prove that evaluating a Shapley coefficient amounts to performing a specific set of conditional independence tests, as implemented by a procedure similar to the CRT but for a different null hypothesis. Furthermore, the obtained game-theoretic values upper bound the p-values of such tests. As a result, we grant large Shapley coefficients with a precise statistical sense of importance with controlled type I error.

Jacopo Teneggi, Beepul Bharti
Johns Hopkins University
jtenegg1@jhu.edu, bbharti1@jhu.edu

Yaniv Romano
EE Department, Technion
Israel
yromano@technion.ac.il

Jeremias Sulam
Johns Hopkins University
jsulam1@jhu.edu

MS150

Interpretability That Matters: Causal Hypothesis Testing with Black-Box Predictive Models in Cancer

Biomarker hypotheses form the basis for preclinical investigation, drug development, and clinical trials in cancer. The success of a biomarker and subsequent trials hinges on a causal connection between the biomarker and patient outcomes—specifically, that intervening on the marker will improve outcomes. Increasingly, biomarker hypotheses are generated by interpreting machine learning models trained on large observational datasets. In this talk, I'll present methods for going beyond heuristic feature selection and marginal association tests when using black box predictive models. In the first part of the talk, I'll present our recent work building scalable hypothesis testing methods that use black box models to test for conditional independence between each feature and an outcome of interest. In the second part, I'll describe work that goes beyond conditional independence to discover causal features in the presence of multiple outcomes with unknown causal structure among them. I will motivate the methods by a recent biomarker analysis of metastatic potential in a large dataset of 22K cancer patients treated at Memorial Sloan Kettering Cancer Center.

Wesley Tansey
Memorial Sloan Kettering Cancer Center

tanseyw@mskcc.org

MS150

What Does This Data Do? A Data-Centric Perspective on Interpretable ML

Many machine learning systems are developed using heterogeneous data from different sources. How do different types of data contribute to the behavior and predictions of the model? I will discuss recent advances in mathematically characterizing the impact of specific data in the context of machine learning, providing a data-centric interpretation of model predictions. We will discuss applications of this approach for data curation and model editing with examples from healthcare and medicine.

James Zou
Stanford
jamesz@stanford.edu

MS151

Multi-Perspective Data Visualization

We propose two algorithms for visualizing a dataset/graph or a network by embedding it in a 3-dimensional Euclidean space based on various given distances between the same pairs of objects. The goal is to find an Embedding that preserves either local or global distance information simultaneously. In order to preserve global distances, we propose an algorithm: Multi-Perspective Simultaneous Embedding (MPSE) that generalizes the Multi-Dimensional Scaling (MDS). In order to preserve local distances, we propose an algorithm: Embedding Neighborhoods Simultaneously t-SNE (ENS-t-SNE) which generalized the t-Stochastic Neighborhood Embedding approach (t-SNE). We illustrate the utility of both algorithms on some synthetic and real-world datasets.

Vahan Huroyan
University of Arizona
vahanhuroyan@math.arizona.edu

MS151

The Geometry of Graph Projection, Interpolation, and Sketching

We study algorithms for sketching and interpolation between graphs in a geometric way. In particular, we use geodesics based on the Coordinated Optimal Transport (COPT) distance to interpolate between graphs. We first generalize this distance to allow for nonuniform distributions over graphs. This allows us to tackle various tasks, such as graph averaging and graph summarization.

Tyler Maunu
Brandeis University
maunu@brandeis.edu

MS151

Representation Power and Theoretical Foundations of Modern Node Embeddings

I will overview our recent work studying theoretical foundations of modern node embedding methods, with the goal of helping to explain their strong empirical performance. I will first discuss work showing that the popular DeepWalk method can be interpreted as a variant of a classic Laplacian eigenvector embedding, with an added non-linearity. This connection to a large prior literature on spectral embedding and spectral graph theory, opens up a pathway for an improved understanding of DeepWalk and related methods. I will also discuss results exploring the power of low-dimensional embeddings in general, in representing graphs and distributions over graphs, and I will overview open questions in the area. The presented work is joint with Sudhanshu Chanpuriya, Konstantinos Sotiropoulos, and Charalampos Tsourakakis.

Cameron Musco
University of Massachusetts Amherst
cmusco@cs.umass.edu

MS151

Dynamic Graph Embedding with Uncertainty Quantification

Dynamic graph embedding has gained great attention due to its capability of learning low-dimensional graph embeddings for complex temporal graphs with high accuracy. However, recent advances mostly focus on learning node embeddings as deterministic vectors for static graphs, hence disregarding the key graph temporal dynamics and the evolving uncertainty associated with node embedding in the latent space. We propose an efficient stochastic dynamic graph embedding method (DynG2G) that applies an inductive feed-forward encoder trained with node triplet energy-based ranking loss. Every node per timestamp is encoded as a time-dependent probabilistic multivariate Gaussian distribution in the latent space. We adopted eight benchmarks of different sizes and evolving dynamics (from slowly changing dynamics to rapidly varying multi-rate dynamics). Our experiments indicate that DynG2G achieves new state-of-the-art performance in capturing the temporal node embeddings and simultaneously predicting the evolving node embedding uncertainty, which plays a crucial role in quantifying the intrinsic dimensionality of the dynamical system over time. We also obtain a universal relation of the optimal embedding dimension L versus the effective dimensionality of uncertainty (D). The L - D correlation provides a clear path for selecting the optimum embedding size adaptively per timestamp by L = Du.

Mengjia Xu
Brown University
mengjia_xu@brown.edu

MS152

Discovering Differential Equations from Partial Measurements with Deep Delay Autoencoders

A central challenge in data-driven model discovery is the presence of hidden, or latent, variables that are not directly measured but are dynamically important. Takens' theorem provides conditions for when it is possible to augment partial measurements with time delayed information, resulting in an attractor that is diffeomorphic to that of the original full-state system. This diffeomorphism is typically unknown, and learning the dynamics in the embedding space has remained an open challenge for decades. Here, we design a deep autoencoder network to learn a coordinate transformation from the delay embedded space into a new space where it is possible to represent the dynamics in a sparse, closed form. We demonstrate this approach on the Lorenz, Rössler, and Lotka-Volterra systems, as well as a Lorenz analogue from a video of a chaotic waterwheel experiment. This framework combines deep learning and

the sparse identification of nonlinear dynamics (SINDy) methods to uncover interpretable models within effective coordinates.

Joseph Bakarji
University of Washington
jbakarji@uw.edu

Jared L. Callaham
University ofWashington
jc244@uw.edu

Steven Brunton
University of Washington
sbrunton@uw.edu

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

MS152

Weight-Parameterization in Neural ODEs for Surrogate Modeling and Sampling

Neural ODEs hinge on time-dependent weight functions as continuous counterparts of weight matrices in residual neural networks. Parameterization of weight functions generates a tradeoff between model complexity and performance. In this talk, we examine and compare different strategies to parameterize the weights of neural ODEs in time. We demonstrate the impact of these parametrizations from viewpoints of training, generalization and efficiency. More specifically, we focus on supervised machine learning tasks such as surrogate model construction for computationally expensive physical models. In the context of various types of neural ODEs and common training algorithms, our results underscore the importance of choosing the basis functions for adequate weight parameterizations. We demonstrate that weight-parametrized neural ODEs can yield accurate surrogates that generalize well and are efficient to store and evaluate.

Haley Rosso
Emory University
haley.rosso@emory.edu

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

Khachik Sargsyan
Sandia National Laboratories
ksargsy@sandia.gov

MS152

Multigrid-Augmented Deep Learning Preconditioners for the Helmholtz Equation

In this work, we present a data-driven approach to iteratively solve the discrete heterogeneous Helmholtz equation at high wavenumbers. In our approach, we combine classical iterative solvers with convolutional neural networks (CNNs) to form a preconditioner which is applied within a Krylov solver. For the preconditioner, we use a CNN of type U-Net that operates in conjunction with multigrid ingredients. Two types of preconditioners are proposed 1) U-Net as a coarse grid solver, and 2) U-Net as a deflation operator with shifted Laplacian V-cycles. Following our training scheme and data-augmentation, our CNN preconditioner can generalize over residuals and a relatively general set of wave slowness models. On top of that, we also offer an encoder-solver framework where an "encoder" network generalizes over the medium and sends context vectors to another "solver" network, which generalizes over the right-hand-sides. We show that this option is more robust and efficient than the stand-alone variant. Lastly, we also offer a mini-retraining procedure, to improve the solver after the model is known. This option is beneficial when solving multiple right-hand-sides, like in inverse problems. We demonstrate the efficiency and generalization abilities of our approach on a variety of 2D problems.

Yael Azulay
Ben-Gurion University of the Negev
yaelaz@post.bgu.ac.il

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

MS154

Addressing Generalization Issues in Subgrid-Scale Modeling of Turbulence and Ocean Flows: Approaches Using Transfer Learning and Graph Neural Networks

The past few years have seen machine learning used to improve the modeling and prediction of nonlinear, multiscale, high dimensional systems. While machine learning methods continue to improve, fully addressing the limited accuracy when models predict outside of their training set remains an open issue. Performance drop can arise when the system parameters or forcing vary starkly, but also as long-term models need to maintain accuracy, for example, in a changing climate. One method of addressing this problem in neural networks is transfer learning. This uses minimal data from a target system to retrain only a subset of network layers, extending the utility of networks across applications and making more effective use of data from data-scarce regimes. Working with 2D turbulence, this work demonstrates the effectiveness of transfer learning, fully recovering the gap in performance in response to changing the forcing and increasing the Reynolds number by two orders of magnitude and presents a framework for understanding and optimizing layer selection in transfer learning for subgrid parameterization models. Another source of generalization error can arise when the discretion of domain varies in either resolution, geometry, or with different irregular boundaries between the training and test setups. To address these issues, we can leverage graph neural networks' ability to represent arbitrary domains and show some preliminary results using graph-based networks to predict subgrid forcing from ocean model data.

Adam Subel
NYU
as15415@nyu.edu

Yifei Guan, Ashesh K. Chattopadhyay
Rice University
yg62@rice.edu, akc6@rice.edu

Pedram Hassanzadeh
Mechanical Engineering and Earth Science
Rice University

pedram@rice.edu

Laure Zanna
Courant Institute of Mathematical Sciences
laure.zanna@nyu.edu

MS155

Ensemble Kalman Methods: 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. It has primarily been developed by practitioners in the geophysical sciences, with notable impact on the fields of oceanography, oil reservoir simulation and weather forecasting. Despite its wide adoption in fields of application, firm theoretical foundations are only now starting to emerge. In this talk we consider the problem from the perspectives of both control theory and probability, and provide a unifying approach to algorithms that rests on transport of measures and mean field stochastic dynamical systems. The ensemble Kalman methods as implemented in practice are found as particle approximations to these mean field models; we explain this Monte Carlo perspective.

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

Sebastian Reich
Universität Potsdam, Germany
sereich@uni-potsdam.de

Andrew Stuart
Computing + Mathematical Sciences
California Institute of Technology
astuart@caltech.edu

MS155

Statistical Inference of RNA Spatial Diffusion Patterns for Mechanistic Insight

Recent advances in imaging have produced data of the spatial positions of RNA molecules within the nucleus of live cells. Several biological questions about RNA dynamics remain to be understood and, in this talk, I will discuss our exploration of whether these spatial positions reveal any insight. To do so, we have constructed a PDE model for the particle positions, and inference on this differs significantly from classical inverse problems due to observations occurring via discrete particles. In our Bayesian approach, the expense of solving the PDE model requires efficient sampling via Monte Carlo techniques. We find that the inference problem has many interesting identifiability features and lends itself well to abstraction toward other questions in the realm of inference of spatial biological processes. This work is in collaboration with the Ding lab at UCI.

Christopher Miles
University of California, Irvine
cemiles@uci.edu

MS155

Sampling Molecular Systems Using Multigrid

Molecular systems are hard to simulate since all the interesting events happen rarely. For example, in simulations of proteins, resolving the fastest vibrations may require a time step as small as a femtosecond, but conformational changes like folding may occur only once in a microsecond. Many techniques have been devised to efficiently calculate the steady-state distributions of such systems, for example parallel tempering and umbrella sampling. We consider a class of efficient simulation methods that use a coarse model to accelerate convergence to the steady-state. Nonequilibrium umbrella sampling, exact milestoning, injection measures, and some versions of weighted ensemble fall into this class. We draw parallels between these methods and algebraic multigrid for the numerical solution of partial differential equations. We analyze the efficiency of these methods for a simple class of test problems modeling molecular simulation.

Brian Van Koten
University of Massachusetts Amherst
vankoten@math.umass.edu

MS155

Transport Map Unadjusted Langevin Algorithm: Guarantees and Connections

Langevin dynamics (LD) are widely used to sample high-dimensional, non-Gaussian distributions whose densities are known up to a normalizing constant. There has been recent interest in the unadjusted Langevin algorithm (ULA) in which a single realization of LD is used to estimate expectations with respect to the target distribution. When the target distribution is not strongly log-concave, the method is known to exhibit slow convergence, which affects the efficiency of the resulting estimator. Meanwhile transport maps provide a way to couple complex non-Gaussian target distributions with simple reference ones, and are a way to generate cheap samples of the target distribution. However, transport maps between random variables can only be approximated which results in biased samples. We present a method that combines transport maps with the unadjusted Langevin algorithm. Given a target distribution and a transport map, we define a stochastic process that is the pullback of a Langevin process on the pushforward of the target distribution through the map. We prove that when the pushforward distribution is strongly log-concave, the output process exhibits geometric convergence in the 2Wasserstein distance even when the target distribution does not satisfy the typical condition conditions. Moreover, we also show that in continuous time, when a transport map is applied to LD, the result is a Riemannian manifold Langevin dynamics with a metric that is defined by the transport map.

Benjamin J. Zhang
Massachusetts Institute of Technology
bjz@mit.edu

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

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

MS156

Function Approximation and Statistical Estimation

Using Neural Networks on Structured Data

Deep neural networks have demonstrated remarkable generalization performance in high dimensional problems, e.g., image classification, where each image contains a large number of pixels. Such appealing performance contradicts a fundamental theoretical challenge curse of data dimensionality. To explain this huge gap, we take the data intrinsic geometric structures into consideration, and model high-dimensional data as samples on a low-dimensional manifold. We show that neural networks can efficiently approximate functions supported on a low-dimensional manifold. The network size scales exponentially in the approximation error, with an exponent depending on the data intrinsic dimension. As an application of our function approximation theory in statistics, we show that deep neural networks can circumvent the curse of data ambient dimensionality by capturing unknown data intrinsic structures and attain fast statistical convergence in regression and distribution estimation.

Minshuo Chen
Georgia Institute of Technology
minshuochen@gmail.com

MS156

On the Convex Dimension of Data

With the ever-growing prevalence of high-dimensional data sources, it is often important to characterize the intrinsic low-dimensionality of such datasets. Understanding such low-dimensionality has broad applications in downstream tasks such as inverse problems, where the low-dimensionality can be learned and incorporated in the inversion process via regularization functionals. We propose and analyze a new notion of complexity for natural data, which we call the convex dimension. This notion characterizes a dataset's dimensionality as the minimal gaussian width a convex body can prescribe to the datapoints on average. In convex geometric terms, learning a regularizer from data can also be phrased in terms of this complexity measure: how does one learn a convex body with minimal gaussian width such that the given datapoints lie on its low-dimensional faces? We study various properties of this quantity and how it relates to the performance of regularizers learned in matrix factorization problems.

Oscar F. Leong
California Institute of Technology
oleong@caltech.edu

MS156

Rethinking Double Descent: Learned Model Complexity Through a Piecewise Affine Representation

The ability of a model to generalize from the training data is characterized by the bias-variance tradeoff with respect to model complexity. Yet the intriguing double descent phenomenon shows that overparameterization can lead to better generalization. However, increasing the capacity of the hypothesis class may not necessarily increase the complexity of the learned model. The key observation is that a (hypothesis) parameter has dual roles in controlling complexity and inducing regularization. The trajectory of added parameters interplays with the data distribution and algorithmic choices to yield various generalization behaviors. We demonstrate this phenomenon in both parametric and nonparametric settings, and discuss how to reconcile the seemingly paradoxical double descent phenomenon by understanding the role of parameters and through effective complexity measures for learned models based on their piecewise affine representations.

Ronan Perry
Max Planck Institute for Intelligent Systems
rflperry@gmail.com

Ningyuan Huang
Johns Hopkins University
nhuang19@jhu.edu

Joshua Vogelstein
Department of Biomedical Engineering
Johns Hopkins University
jovo@jhu.edu

MS156

Exact High-Dimensional Asymptotics for Statistical Estimation via Approximate Message Passing

Approximate message passing (AMP) is a class of computationally-efficient algorithms derived from loopy belief propagation on dense factor graphs. They have two features that make them particularly attractive: they can easily be tailored to take advantage of prior information on the structure of the signal, such as sparsity or other constraints, and under suitable assumptions on a design or data matrix, AMP theory provides precise asymptotic guarantees for statistical procedures in the high-dimensional regime where the ratio of the number of observations n to dimensions p converges to a constant. This talk will present a general program for systematically deriving exact expressions for the asymptotic risk of estimators that are solutions to a broad class of convex optimization problems using AMP. This talk is based on (https://arxiv.org/abs/2105.02180) written in collaboration with Oliver Feng, Ramji Venkataramanan, and Richard Samworth.

Cynthia Rush
Columbia University
cgr2130@columbia.edu

MS157

Low-Rank Conditional Structure in Bayesian Inverse Problems with Transport Maps

Approximate sampling via probability transport maps has recently gained popularity as an alternative and a complement to standard sampling strategies such as MCMC. The computational effort needed to train a transport map depends on the total number of map parameters, making large-scale Bayesian inverse problems challenging. We present a new method for encoding sparsity in transport maps by exploiting low-rank conditional (LRC) structure in the posterior distribution. By enforcing that components of the map only depend on low-dimensional summaries of the input variables, we can greatly reduce the number of map parameters needed for high-dimensional problems. We show how to uncover LRC structure in the posterior by solving a particular eigenvalue problem and provide intuition connecting LRC structure to hierarchical matrix approximations of the Hessian of the log-likelihood. We demonstrate a significant reduction in the number of map parameters needed for several PDE-constrained inverse problems and show improvements in the training be-

havior of these maps compared to unstructured maps.

Michael C. Brennan
MIT, CCSE
mcbrenn@mit.edu

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

MS157

Transport Optimization Methods in Bayesian Sampling Problems

We present a class of information metric optimization methods for Bayesian inverse problems. Two examples of information metrics are studied, including the Fisher-Rao metric and Wasserstein-2 metric. Focusing on the Wasserstein-2 metric, we introduce accelerated gradient and Newton flows. We also formulate their practical discrete-time algorithms in particles. Facing a curse of dimensionality problem, we further introduce a projected Wasserstein gradient method to handle it. Numerical experiments, including PDE-constrained Bayesian inference and parameter estimation in COVID-19 modeling, demonstrate the effectiveness of the proposed method. This is based on joint works with Yifei Wang and Peng Chen.

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

MS157

Constrained Variational Inference with Gradient Flow

Approximate inference and learning techniques, especially Bayesian inference, provide an essential approach to handling uncertainty in machine learning (ML). As these techniques are increasingly used in daily life, it becomes essential to safeguard the ML systems with various trustworthy-related constraints, such as fairness, safety, interpretability. Enforcing these constraints in probabilistic inference can be challenging because it requires to trade off the approximate inference of target distributions and the nonlinear constraints. This talk will introduce a framework for solving the constrained inference problem by leveraging the variational view of approximate inference as taken in recent particle-based methods such as Langevin dynamics and Stein variational gradient descent (SVGD): We formulate the constrained inference problem as a constrained optimization of KL divergence objective in the space of probability measures, and solve it by deriving constrained gradient flow methods that generalizes Langevin dynamics and SVGD. These methods provide a simple approach to handling general nonlinear constraints in probabilistic inference in a wide range of applications.

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

MS157

Data-Driven Summary Functions for Scalable Amortized Variational Inference

Bayesian inference for inverse problems is challenged by the computational costs associated with sampling the posterior distribution. Amortized variational inference reduces these costs by pretraining a neural network that approximates the posterior distribution for previously unseen observations. Given observed data during inference, this network provides samples from the posterior distribution virtually for free, thereby accelerating Bayesian inference. For multi-source inverse problems, as in seismic imaging, training such a network becomes computationally challenging due to the high dimensionality of the observed data. To address this challenge, we propose using learned summary functions to reduce the dimensionality of the observed data while maintaining the most informative data statistics. This has the advantage of controlling the network size, leading to reduction in the computational cost of amortized variational inference.

Rafael Orozco, Ali Siahkoohi
Georgia Institute of Technology
rorozco@gatech.edu, alisk@gatech.edu

Gabrio Rizzuti
Utrecht University
g.rizzuti@uu.nl

Tristan Van Leeuwen
Centrum Wiskunde & Informatica (CWI)
t.van.leeuwen@cwi.nl

Felix Herrmann
Georgia Institute of Technology
felix.herrmann@gatech.edu

MS158

Graph Learning for Surface Water and Sediment Detection in Multispectral Images

We develop a method based on graph learning and active learning to detect surface water pixels in satellite images. For this task, current cutting-edge products are mostly based on convolutional neural networks and huge training datasets. Taking the advantage of graph learning approaches, our method is trained on a much smaller dataset and reaches a higher accuracy. Furthermore, we provide a pipeline that combines graph learning and active learning techniques to detect surface water from scratch. Specifically, in order to have a fairly good classification, a user only needs to label around 0.1% pixels according to the guidance of our method. This is joint work with Andrea L. Bertozzi, Kevin Miller, and Jon Schwenk.

Bohan Chen
University of California, Los Angeles
bhchenyz@g.ucla.edu

MS158

Graph-Based Approach for Federated Multi-Task Learning in a Privacy-Preserving Manner

In multi-task learning, multiple learningtasks are performed jointly rather than separately to leverage theirsimilarities and improve performance.We focus on the federated learning setting, whereeach machine possessesits own data for individual tasks and sharing the full local data between machines is prohibited due to privacy, etc. This introduces a low-sample regime for each of the learning task, further motivating the need for inter-machine collaboration, i.e. multi-task learning.Motivated by graph regularization,we propose a novel fusion framework that only requires a one-shotcommunication of local estimates.Our method linearly

combinesthe local estimates to produce an improved estimate for each task,and we show that the ideal mixing weight for fusion is a function oftask similarity and task difficulty. A practical algorithm is developedand shown to significantly reduce mean squared error (MSE) on synthetic data, as well as improve performance on an income predictiontask where the real-world data is disaggregated by race. On-going work on hyperspectral image analysis and active learning will also be presented.

Harlin Lee
UCLA
harlin@math.ucla.edu

MS158

Active Learning via Learned Local Exemplars

We propose a unified optimization framework for nonlinear dimensionality reduction that combines the strengths of manifold learning and sparse dictionary learning. In this framework, a data point is represented as a sparse convex combination of local exemplars. The sparse coefficients and exemplars are optimized efficiently using alternating minimization. We endow the set of points and the learned exemplars with a graph structure where the sparse coefficients become edge weights. We discuss ways in which labels are to be propagated from any few labeled examples. Using the obtained graph, we also give some characterization for the problem of identifying impactful examples for label query.

Abiy F. Tasissa
Tufts University
Abiy.Tasissa@tufts.edu

MS159

Networks that Adapt to Intrinsic Dimensionality Beyond the Domain

This talk discusses the approximation of two-layer compositions $f(x) = g(\phi(x))$ via deep networks with ReLU activation, where ϕ is a geometrically intuitive, dimensionality reducing feature map. We'll focus on two intuitive and practically relevant choices for ϕ: the projection onto a low-dimensional embedded submanifold and a distance to a collection of low-dimensional sets. This will cover the capacity of neural networks for both regression and classification models. Since ϕ encapsulates all nonlinear features that are material to the function f, this suggests that deep nets are faithful to an intrinsic dimension governed by f rather than the complexity of the domain / data on which f is defined. In particular, the prevalent model of approximating functions on low-dimensional manifolds can be relaxed to include significant off-manifold noise by using functions of this type, with ϕ representing an orthogonal projection onto the same manifold. We'll also discuss connections of this work to manifold autoencoders and data generation.

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

Timo Klock
Deeptech Consulting
tmklock@googlemail.com

Rongjie Lai
Rensselaer Polytechnic Institute
lair@rpi.edu

Stefan C. Schonsheck
Univeristy of California Davis
sschonsh@gmail.com

Scott Mahan
University of California, San Diego, U.S.
scmahan@ucsd.edu

MS159

Supervised Learning of Sheared Distributions Using Linearized Optimal Transport

Detecting differences and building classifiers between distributions, given only finite samples, are important tasks in a number of scientific fields. Optimal transport (OT) has evolved as the most natural concept to measure the distance between distributions and has gained significant importance in machine learning in recent years. However, OT often fails to exploit reduced complexity in case the family of distributions is generated by simple group actions. In this talk, we discuss how optimal transport embeddings can be used to deal with this issue, both on a theoretical and a computational level. In particular, we embed the space of distributions into an L^2-space by mapping a distribution to its OT map with respect to a fixed reference distribution. We further give an exact characterization of distributions for which this embedding is an isometry. In the embedding space, we use regular machine learning techniques to achieve linear separability when the classes of distributions are generated by a family of shearings, describing conditions under which two classes of sheared distributions can be linearly separated. We also give necessary bounds on these shearing transformations to achieve a pre-specified separation level. Furthermore, embedding into multiple L^2 spaces allows for not only larger families of transformations but also a greater level of linear separation. Finally, our theoretical results are verified empirically on image classification tasks.

Varun Khurana
University of California, San Diego
vkhurana@ucsd.edu

MS159

Wasserstein Embeddings in the Deep Learning Era

Computational optimal transport has found many applications in machine learning and, more specifically, deep learning as a fundamental tool to manipulate and compare probability distributions. The Wasserstein distances arising from the optimal transport problem have been of particular interest in recent years. However, a consistent roadblock against the more prevalent use of transport-based methods has been their computational cost. Besides the more well-known ideas for faster computational approaches, including entropy regularization, several fundamental concepts have emerged that enable the integration of transport-based methods as part of the computational graph of a deep neural network. Sliced-Wasserstein distances and the Linear Optimal Transport (LOT) framework are among fundamental concepts well suited for integration into today's deep neural networks. In this talk, we will present the idea of Linear Optimal Transport (otherwise known as the Wasserstein Embedding) and its extension to Sliced-Wasserstein Embeddings and demonstrate their various applications in deep learning with a particular

interest in learning from graphs and set-structured data.

Soheil Kolouri
Vanderbilt University
soheil.kolouri@vanderbilt.edu

MS159

From the Simplex to the Sphere: Faster Constrained Optimization using the Hadamard Parametrization

Minimizing a smooth function over the standard probability simplex is an important sub-routine in many machine learning tasks. Although standard algorithms (e.g., Projected Gradient Descent) perform adequately, there is a need for new algorithms which are amenable to parallelization, and possess convergence guarantees for non-convex functions. We propose to rephrase this problem, via the Hadamard parametrization, as a minimization problem over the unit sphere. This allows us to leverage new theory on Riemannian optimization to provide an algorithm satisfying the aforementioned desiderata.

Daniel McKenzie
University of California, Los Angeles, U.S.
mckenzie@math.ucla.edu

MS160

Effective Matrix Designs for Covid-19 Group Testing

For large-scale testing, a particularly effective strategy is using *pooling methods*. Instead of testing individual samples one-by-one, pooling tests mix a group of samples into one pool and then use one test on the pool. If the test result is negative consequently one can conclude that all samples in the pool had to be negative (potentially saving scarce resources). When the pools are tested in parallel, matrices defined by elements of 0's and 1's encode all essential information of the tests. The specific matrix designs enable identifying up to a guaranteed number of positives among all samples. Recent methods have been based on prime number pool sizes or in other ways not fully exploiting all possible properties. This talk presents methods for constructing pooling matrices with pool sizes that are powers of prime numbers, and that extend previous designs. Algorithms are proposed that may be used to test a large number of samples for a range of disease prevalences.

David Brust
forschungszentrum Jülich
david.brust93@gmail.com

Johannes J. Brust
University of California, San Diego
Mathematics Department
jjbrust@ucsd.edu

MS160

Network-Based Functional Regions and Social Segregation in Mexico

Mexico City is one of the largest metropolitan areas in the world and the main social and economic hub in the country. COVID-19 pandemic changed how people interact and move within the city in many ways, yielding complex consequences of social relevance. Starting from a large dataset of geolocation data, we built daily origin-destination networks for Mexico City during 2020 and 2021. We utilize community detection algorithms in these large networks to delineate functional regions in the city and we studied the structure and evolution of these regions during the COVID-19 pandemic. Krackhardt E/I Ratio of network communities allows us to study patterns of mobility and social segregation within the city and how these patterns evolved with mobility restrictions and lockdowns. Our findings show that, even though gross mobility levels have reached their pre-pandemic levels, the way in which people move within the city is not the same as before. We discuss social implications of these results.

Oscar Fontanelli
Facultad de Ciencias, Universidad Nacional Autónoma de Méxic
oscarfontanelli@gmail.com

Maribel Hernández-Rosales
Centro de Estudios e Investigaciones Avanzadas del IPN
maribel.hr@cinvestav.mx

MS160

Probability and Measure Theory for Optimal SARS-CoV-2 Testing

Estimating *prevalence*, the fraction of a population that has been previously infected by a disease such as SARS-CoV-2, is among the most important tasks in epidemiology. When estimating prevalence, it is often assumed that diagnostic classification of samples is necessary. Using a perspective motivated by metrology, we demonstrate that this longstanding assumption is false. Conditional probability models of diagnostic measurement outcomes for positive and negative samples are used to construct a new, unbiased, and converging estimator of prevalence. We further demonstrate that the uncertainty in this estimate can be minimized using a *bathtub principle* from measure theory. This result establishes that optimal prevalence estimation and optimal diagnostic classification have fundamentally distinct solutions. Using examples from real-world SARS-CoV-2 antibody, we compare our technique with traditional prevalence estimates that fail to converge and illustrate the differences between prevalence estimation and classification.

Anthony Kearsly
National Institute of Standards and Technology
Applied and Computational Mathematics Division
anthony.kearsley@nist.gov

Paul Patrone
NIST Gaithersburg, Information Technology Lab
Applied and Computational Mathematics Division
paul.patrone@nist.gov

MS160

Dynamical Clustering of U.S. States Reveals Four Distinct Infection Patterns that Predict Sars-CoV-2 Pandemic Behavior

The SARS-CoV-2 pandemic has so far unfolded diversely across the fifty United States of America, reflected both in different time progressions of infection "waves" and in magnitudes of local infection rates. Despite a marked diversity of presentations, most U.S. states experienced their single greatest surge in daily new cases during the transition from Fall 2020 to Winter 2021. Popular media also cite additional similarities between states – often despite dis-

parities in governmental policies, reported mask-wearing compliance rates, and vaccination percentages. Here, we identify a set of robust, low-dimensional clusters that 1) summarize the timings and relative heights of four historical COVID-19 "wave opportunities" accessible to all 50 U.S. states, 2) correlate with geographical and intervention patterns associated with those groups of states they encompass, and 3) predict aspects of the "fifth wave" of new infections in the late Summer of 2021. In particular, we argue that clustering elucidates a negative relationship between vaccination rates and subsequent case-load variabilities within state groups. We advance the hypothesis that vaccination acts as a "seat belt," in effect constraining the likely range of new-case upticks, even in the context of the Summer 2021, variant-driven surge.

Joseph Natale
Halicioglu Data Science Institute
University of California, San Diego, La Jolla, CA 90293
jonatale@ucsd.edu

Johannes J. Brust
University of California, San Diego
Mathematics Department
jjbrust@ucsd.edu

Varun Viswanath
Department of Electrical and Computer Engineering
University of California, San Diego, La Jolla, CA 90293
varunv9@eng.ucsd.edu

Oscar Trujillo Acevedo
Division de Ingenieria Bioquimica
Instituto Tecnologico Superior of Atlixco, Mexico
oscar.trujillo@itsatlixco.edu.mx

Sophia Pérez Giottonini
Student
Colegio de Bachilleres del Estado de Sonora, Sonora, Mexico
sp19010189@bachilleresdesonora.edu.mx

Sandy Ihuiyan Romero Hernández
Student
Instituto Tecnologico Superior de Atlixco, Mexico
ihuiyan99@gmail.com

Diana Cruz Millán
Department of Health Sciences
Universidad de Sonora, campus Cajeme, Mexico
dgcm98@gmail.com

Jay Park
Department of Emergency Medicine
University of California, San Diego, La Jolla, CA 92037
jp006@ucsd.edu

Benjamin Smarr
University of California, San Diego
bsmarr@eng.ucsd.edu

MS161

Training and Predictive Uncertainty for Multifidelity Convolutional Neural Networks

The limited availability of high-fidelity predictions often prevents the training of data-driven models. In this contribution, we present our recent progress towards lowering the high-fidelity data requirements by training convolutional neural networks (CNNs) from ensembles of high- and low-fidelity data. CNNs, which are obtained by assembling convolutional encoders, decoders and skip connections, have the advantage of applying the same parameters over a large space, limiting the number of parameters required to learn patterns in high-dimensional data as compared to a fully-connected network. We discuss and explore strategies for the all-at-once training of these networks and we also consider the estimation of the uncertainty in their predictions via Monte Carlo dropBlocks. We demonstrate the predictive capabilities of these networks on several verification problems and data generated from partial differential equations.

Lauren Partin
University of Notre Dame
lhensley@nd.edu

Gianluca Geraci
Sandia National Laboratories, NM
ggeraci@sandia.gov

Ahmad A. Rushdi
Stanford University
rushdi@stanford.edu

Michael S. Eldred
Sandia National Laboratories
Optimization and Uncertainty Quantification Dept.
mseldre@sandia.gov

Daniele E. Schiavazzi
University of Notre Dame
dschiavazzi@nd.edu

MS161

Bayesian Approaches to Multi-Fidelity Uncertainty Quantifications of Unstructured Ensembles of Models

This talk discusses Bayesian inference methodologies for MFNets – a paradigm for multifidelity information fusion via directed acyclic graphs. MFNets provide a flexible approach to modeling the relationships between unstructured ensembles of models and information sources by linking the outputs of each information source though a network of models. As a result, data on a high-fidelity information source informs a full cascade of it's ancestral information sources. In this talk we discuss Bayesian learning of the parameters of this network. We discuss both sampling-based methodologies targeting the full posterior, as well as variational approaches for its approximation. We demonstrate the complexity of the shape of the posterior and its resulting challenges for sampling methods. We also demonstrate the performance of variational inference methodologies and discuss their challenges for representing highly complex multi-modal behavior. Examples from both synthetic and physical models are provided.

Alex Gorodetsky, Trung Pham
University of Michigan
goroda@umich.edu, trungp@umich.edu

John D. Jakeman
Sandia National Labs

jdjakem@sandia.gov

MS161

Enhancing Polynomial Chaos Expansion Surrogates Through Probabilistic Transfer Learning

In the context of surrogate modeling, polynomial chaos expansions (PCE) allow practitioners to construct cheap yet sufficiently-accurate surrogates to be used in place of the expensive forward model simulations. For black box simulations, non-intrusive PCE allows the construction of said surrogates using a set of simulation response evaluations. In that context, the PCE coefficients can be obtained using linear regression approach, also known as point collocation or stochastic response surfaces, which exhibits better scalability and can handle noisy function evaluations in contrast to other non-intrusive approaches. However, since over-sampling is generally advisable for the linear regression approach, the simulation requirements become prohibitive for expensive forward models. We propose to leverage transfer learning whereby knowledge gained through similar PCE surrogate construction tasks (source domains) is transferred to a new surrogate-construction task (target domain) having limited number of forward model simulations (training data). The proposed transfer learning strategy will determine when it is worth applying transfer learning and how much knowledge is to be transferred using new techniques inspired by Bayesian modeling and sequential data assimilation.?

Mohammad Khalil, Uma Balakrishnan, Reese Jones, Cosmin Safta
Sandia National Laboratories
mkhalil@sandia.gov, ubalakr@sandia.gov, rjones@sandia.gov, csafta@sandia.gov

Jiefu Chen, Xuqing Wu
University of Houston
jchen82@central.uh.edu, xwu8@central.uh.edu

Yueqin Huang
Cyentech Consulting LLC
yueqinhuang@cyentech.com

MS162

Multigrid Proximal Gradient Method for Non-Smooth Convex Optimization

We present a recent result on accelerating a 1st-order method for solving convex (possibly non-smooth) composite optimization problem of the form f(x) + g(x). We propose a multigrid (MG) based convergence acceleration method for the proximal gradient method. Coming from the domain of PDEs and scientific computing, the idea of multigrid assumes that the optimization problem has a hierarchical structure that can be exploited. By utilizing such hierarchy, acceleration can be achieved by a multilevel process. We provide several theoretical results for the proposed method. We show a fixed-point property of the sequence generated by the method, and we provide a simple convergence analysis, based on a recent result on the Polyak-Lojasiewicz inequality, to show that the proposed method achieves a linear convergence rate on strongly convex problems. Finally, we illustrate that the proposed MG-accelerated proximal gradient outperforms the proximal gradient method with Nesterov's acceleration, especially for large-sized problems in certain problem classes, such as a class of PDEs with a free boundary condition known as the elastic obstacle problem.

Man Shun Ang
University of Waterloo
Canada
ms3ang@uwaterloo.ca

Hans De Sterck
University of Waterloo, Canada
hdesterck@uwaterloo.ca

Stephen A. Vavasis
University of Waterloo
Dept of Combinatorics and Optimization
vavasis@uwaterloo.ca

MS162

Linear Asymptotic Convergence Analysis of Anderson Acceleration as a Krylov Method

We consider Anderson acceleration (AA) with a moving window of size m, and investigate the linear asymptotic convergence behaviour of AA(m) applied to linear and nonlinear fixed-point methods. Anderson acceleration has been shown empirically to be very effective for accelerating optimization methods such as ADMM applied to machine learning problems, but there is no theory to explain and quantify the asymptotic convergence improvement that is observed in practice. We first observe numerically that the root-linear convergence factor of sequences generated by AA(m) strongly depends on the initial condition, and that the acceleration coefficients oscillate while the approximation converges to the fixed point. To shed light on this behaviour, we write AA(m) itself as an augmented fixed-point method and investigate the continuity and differentiability of the fixed-point iteration function and the acceleration coefficients. To further investigate AA(m) convergence, we consider the case of accelerating linear fixed-point methods and write AA(m) as a Krylov space method. We obtain polynomial residual update formulas for AA(m) and derive an (m+2)-term recurrence relation for the AA(m) polynomials. This allows us to demonstrate several properties of AA(m). The main question to quantify the asymptotic convergence acceleration provided by AA(m) remains, however, an open problem.

Hans De Sterck
University of Waterloo, Canada
hdesterck@uwaterloo.ca

Yunhui He
The University of British Columbia
yunhui.he@ubc.ca

MS162

Multiscale Training Algorithms for Deep Neural Networks

Deep networks (DNNs) are used in a wide range of application areas and scientific fields. Since the representation capacity of DNNs is tightly coupled to their width and depth, networks have grown considerably over the last years. As this growing trend is expected to continue, the development of efficient and scalable training algorithms becomes an important task. In this talk, we discuss how to enhance the training of large-scale DNNs using two multilevel training methods. Our first approach utilizes the MG-OPT framework. Here, we exploit the finite-sum structure of the loss function and propose to create the multilevel hierarchy and

transfer operators by reducing the number of samples. This gives rise to a novel class of multilevel solution strategies, which also includes stochastic variance reduction methods. Secondly, we consider continuous-in-depth neural network architectures and employ a hierarchical basis for the parametrization of the controls/weights. Exploiting the properties of a hierarchical basis, we then propose a cascadic training algorithm, which incorporates adaptive refinement in the network depth. The convergence properties and scaling behavior of the proposed training methods will be demonstrated using several state-of-the-art benchmark problems. Moreover, a comparison with the widely-used stochastic gradient optimizer will be presented, showing a significant reduction in the number of iterations and the execution time.

Alena Kopanicakova
Università della Svizzera italiana
Switzerland
alena.kopanicakova13@gmail.com

Rolf Krause
Università della Svizzera italiana, Switzerland
rolf.krause@usi.ch

MS162

Layer-Parallel Multi-Level Optimization for Deep Residual Neural Networks

Deep neural networks (DNNs) exhibit excellent performance for many machine learning tasks, e.g., image classification, natural language processing, and game playing. However, training DNNs remains challenging and computationally expensive, with much room for improvement, both in terms of new sources of parallelism and algorithmic speedup. Recently, the parallel-in-time method, multigrid-reduction-in-time (MGRIT), has been applied to neural ODEs and certain recurrent neural networks to provide new parallelism to the layer dimension (layer-parallelism) with resulting parallel speedups. In this talk, we consider combining layer-parallelism with recent advances applying multilevel optimization strategies (called MGOPT) to training DNNs. These MGOPT approaches choose search directions in a hierarchical fashion and can potentially provide an algorithmic speedup (as opposed to new parallelism). MGOPT uses a similar hierarchy to layer-parallel, thus making the combination natural. We demonstrate our approach on several benchmark machine learning problems, explore the applicability in the context of physics-informed neural networks, and suggest possible improvements to the MGOPT framework for stochastic problems.

Jacob B. Schroder
Department of Mathematics and Statistics
University of New Mexico
jbschroder@unm.edu

Eric C. Cyr
Computational Mathematics Department
Sandia National Laboratotories
eccyr@sandia.gov

MS163

Visualizing the Shape of Gene Expression Across Plant Evolution

Topological Data Analysis (TDA) is a growing field that provides methods to characterize and summarize the shape of data. Mapper is one such method which allows us to reduce the often complex topological structure of the data to a one dimensional graph. It is simple and scalable, and the flexibility it provides makes mapper an ideal exploratory visual analysis tool. In this study, we apply mapper to visualize the shape of gene expression across numerous species and families of flowering plants, also known as angiosperms. Angiosperms diversified relatively recently, between 130 and 90 million years ago, and comprise an astounding variety of plants. In this meta-analysis, we collect publicly available datasets of angiosperms that include gene expression along with the information about the tissue type and stresses the samples have been exposed to. After identifying groups of homologous genes between species, or orthogroups, we create a combined dataset of gene expression profiles of 2671 samples, across 16 plant families, 10 tissue types and 8 stresses. The process of photosynthesis is conserved across all plants. Therefore, we design a lens function that captures the photosynthetic vs non-photosynthetic separation between various tissue types. We then construct a mapper graph to visualize the shape of the data as seen through this lens. This work was done in collaboration with 2021 MSU HRT841/CSS844 course students.

Sourabh Palande
Michigan State University
palandes@msu.edu

MS163

Using Mapper to Reveal Morphological Relationships in Passiflora Leaves

As collections of data grow in size, it is increasingly important to have efficient means of analyzing large data sets. In this project, we use Mapper [Singh et al, 2007], a tool that summarizes data into a graph, to discover an underlying structure relating the shapes of more than 3,300 *Passiflora* leaves from 40 different species [Chitwood et al, 2017]. We choose to study leaves of the *Passiflora* genus in particular due to their extraordinary diversity of shape. As the Mapper graph has a structure, or "shape' of its own, we think of it as a "shape of shapes' that provides information on the interplay between the developmental processes determining leaf shape within a single plant and the evolutionary processes between species. In particular, we examine the interactions between leaf species and both leaf age and leaf area by constructing a Mapper graph for each measure. For each node in the resulting graphs, we then compute the average leaf shape to obtain a graph structure that reveals how morphometric differences between species relate to the developmental changes that must occur for those shapes to be realized. We additionally present a new Mapper visualization in Javascript which incorporates both tooltips that display the average leaf shape per node as well as supporting pie chart nodes, useful for visualization of categorical data such as *Passiflora* species.

Sarah Percival, Erik J. Amezquita, Sourabh Palande
Michigan State University
perciva9@msu.edu, amezqui3@msu.edu,
palandes@msu.edu

Aman Husbands
Department of Biology
University of Pennsylvania
ayh@upenn.edu

Arjun Krishnan
Computational Mathematics, Science and Engineering
Michigan State University

arjun@msu.edu

Beronda Montgomery
Department of Biochemistry and Molecular Biology
Michigan State University
montg133@msu.edu

Elizabeth Munch, Daniel Chitwood
Michigan State University
muncheli@msu.edu, chitwoo9@msu.edu

MS163

A Topological Data Analysis Perspective on Coronavirus Evolution

COVID-19 is a disease caused by a coronavirus named SARS-CoV-2. As the virus has been spreading around the world, millions of viral genomes have been sequenced. Genomes provide an accurate record of variation and evolution and can inform how these viruses emerge and evolve. In this talk I will discuss how mutations and recombinations shape SARS-CoV-2 evolution, and how topological data analysis techniques can help to understand the role of these processes in the emergence of this and other potential future pandemics.

Raul Rabadan
Columbia University
rr2579@cumc.columbia.edu

MS163

Temporal Mapper: a Mapper-Inspired Approach to Analyzing Nonlinear Brain Dynamics

The human brain exhibits complex network dynamics. Modeling brain dynamics can facilitate the understanding of human cognition and psychiatric disorders. Data-driven methods seek to identify stable brain states and transitions in time series. Intuitively, the states and transitions resonate with the concepts of attractors and phase transitions in mechanistic (dynamical systems) modeling of the brain. Yet, a direct link between data-driven and mechanistic modeling approaches remains missing. Building on our previous work using Mapper to characterize human brain dynamics, we developed a variant named the Temporal Mapper, bridging data-driven and mechanistic modeling. Temporal Mapper summarizes neural time series data as a directed graph, with nodes mapping to attractors and edges mapping to phase transitions in dynamical systems. For theoretical validation, we simulated neural dynamics using a biophysical network model of the human brain, where phase transitions were induced in a controlled manner. Temporal Mapper reasonably reconstructs the attractor transition networks compared to the theoretical ground truth. For empirical validation, we apply the method to an existing human fMRI dataset during a continuous multitask experiment. Features of the high-degree nodes and cycles of the directed group were significantly associated with human behavioral performance. This work is an initial step towards unifying data-driven and mechanistic modeling of brain dynamics.

Mengsen Zhang
Stanford University
Department of Psychiatry and Behavioral Sciences
mengsen_zhang@med.unc.edu

MS164

Iterative Algorithms for Partitioned Neural Network Approximation

Recently, with the success of deep learning methodologies in many applications, neural networks have been also applied to approximate functions or solutions of differential equations. They showed promising results and are expected to serve as an alternative to classical approximation methods with better flexibility. On the other hand, the parameter optimization in the neural network approximation needs enormous computing cost as the size of parameters gets often very large in many application problems. To address this issue, in the many recent works partitioned neural networks are formed and their parameters are optimized in parallel. In those methods, the limitation is that the parallel computing needs a lot of communication cost. In this work, with the aim of reducing the communication cost, iterative algorithms for the partitioned neural network approximation will be proposed by adapting domain decomposition methods and concrete convergence analysis on the iterative algorithms will be presented. Numerical results will be included to show the performance of the proposed methods.

Hyea Hyun Kim
Kyung Hee University
1732, Deogyeong-daero, Giheung-gu, Yongin-si,
Gyeonggi-do 17
hhkim@khu.ac.kr

Hee Jun Yang
Kyung Hee University
yhjj109@khu.ac.kr

MS164

Two-Level Group Convolution

Group convolution has been widely used in order to reduce the computation time of convolution, which takes most of the training time of convolutional neural networks. However, it is well known that a large number of groups significantly reduce the performance of group convolution. In this talk, we propose a new convolution methodology called "two-level" group convolution that is robust with respect to the increase of the number of groups and suitable for multi-GPU parallel computation. We first observe that the group convolution can be interpreted as a one-level block Jacobi approximation of the standard convolution, which is a popular notion in the field of numerical analysis. In numerical analysis, there have been numerous studies on the two-level method that introduces an intergroup structure that resolves the performance degradation issue without disturbing parallel computation. Motivated by these, we introduce a coarse-level structure which promotes intergroup communication without being a bottleneck in the group convolution. We show that all the additional work induced by the coarse-level structure can be efficiently processed in a distributed memory system. We compare the proposed method to various approaches for group convolution in order to highlight the superiority of the proposed method in terms of execution time, memory efficiency, and performance.

Chang-Ock Lee
Department of Mathematical Sciences
KAIST

colee@kaist.edu

Youngkyu Lee, Jongho Park
KAIST
lyk92@kaist.ac.kr, jongho.park@kaist.ac.kr

MS165
Machine Learning X Finite Element Methods

Finite Element Methods (FEM) are arguably one of the most widely used numerical methods to approve PDEs in science and engineering disciplines. The key elements include using discrete basis functions on a mesh to approximate the target function. Recently, Neural Networks, facilitated by the user friendly interface of the auto-differentiation in various machine learning packages, are used to approximate the target function in a mesh free way. However, using NN changes a well-conditioned problem into a nonconvex ill-conditioned problem. The worse part is that, to achieve the same accuracy, NN and optimization-based PDE solver costs hundreds of thousand more Floating Point Operations (FLOPs) than FEM. In this talk, we shall introduce a new auto-differentiation based FEM implemented natively in PyTorch to handle both linear and nonlinear PDEs, and its application in solving inverse problems.

Shuhao Cao
Washington University in St. Louis
s.cao@wustl.edu

Ruchi Guo
University of California, Irvine
Department of Mathematics
ruchig@uci.edu

Long Chen
University of California at Irvine
chenlong@math.uci.edu

MS165
Accelerating Learning Neural ODEs via Proximal Algorithms

Learning neural ODEs often requires solving very stiff ODE systems, primarily using explicit adaptive ODE solvers. These solvers are computationally expensive, requiring the use of tiny step sizes for numerical stability and accuracy guarantees. We considers learning neural ODEs using implicit ODE solvers of different orders leveraging proximal operators. The proximal implicit solver consists of inner-outer iterations: the inner iterations approximate each implicit update step using a fast optimization algorithm, and the outer iterations solve the ODE system over time. The proximal implicit ODE solver guarantees superiority over explicit solvers in numerical stability and computational efficiency. We validate the advantages of proximal implicit solvers over existing popular neural ODE solvers on various challenging benchmark tasks, including learning continuous-depth graph neural networks and continuous normalizing flows

Long Chen
University of California at Irvine
chenlong@math.uci.edu

MS165
A Continuous Data Assimilation for Two-Phase Flow

In this work, we discuss, propose and analyze a novel continuous data assimilation two-phase flow algorithm for reservoir simulation. We show that the solutions of the algorithm, constructed using coarse mesh observations, converge at an exponential rate in time to the corresponding exact reference solution of the two-phase model. More precisely, we obtain a stability estimate which illustrates an exponential decay of the residual error between the reference and approximate solution, until the error hits a threshold depending on the order of data resolution. Numerical computations illustrate the effectiveness of this approach, as well as variants with data on sub-domains. In particular, we demonstrate numerically that synchronization is achieved for data collected from a small fraction of the domain.

Yat Tin Chow
UC Riverside
yattinc@ucr.edu

Wing Tat Leung
Department of Mathematics
University of California, Irvine
wtleung@uci.edu

Ali Pakzad
Indiana University-Bloomington
apakzad@iu.edu

MS165
Learning Stochastic Closures Using Sparsity-Promoting Ensemble Kalman Inversion

Closure models are widely used in simulating complex systems such as turbulence and Earths climate, for which direct numerical simulation is too expensive. Although it is almost impossible to perfectly reproduce the true system with closure models, it is often sufficient to correctly reproduce time-averaged statistics. Here we present a sparsity-promoting, derivative-free optimization method to estimate model error from time-averaged statistics. Specifically, we show how sparsity can be imposed as a constraint in ensemble Kalman inversion (EKI), resulting in an iterative quadratic programming problem. We illustrate how this approach can be used to quantify model error in the closures of dynamical systems. In addition, we demonstrate the merit of introducing stochastic processes to quantify model error for certain systems. We also present the potential of replacing existing closures with purely data-driven closures using the proposed methodology. The results show that the proposed methodology provides a systematic approach to estimate model error in closures of dynamical systems.

Jinlong Wu, Tapio Schneider
California Institute of Technology
jinlong@caltech.edu, tapio@caltech.edu

Andrew Stuart
Computing + Mathematical Sciences
California Institute of Technology
astuart@caltech.edu

MS166
Spectral Convergence of Kernel Integral Operators

The theory of kernel integral operators lies at the intersec-

tion of diverse fields such as operator theory, data-mining and Geometric analysis. Several connections have been established between kernels depending on an underlying Riemannian metric, and geometric properties such as the curvature, the Laplace and advection operators. We consider a class of kernels called local kernels, equipped with a bandwidth parameter ϵ controlling the decay of tails. We provide a more complete theory on understanding the behavior of these kernels as $\epsilon \to 0^+$, using the language of tensors. We demonstrate that the collection of limiting operators is much broader than Laplacians. In addition, we prove spectral convergence of the kernel integral operator under certain conditions. We also establish pointwise and spectral convergence properties of the data-driven versions of these kernels and operators.

Suddhasattwa Das
Department of Mathematical Sciences, George Mason University
dass@cims.nyu.edu

Ryan Vaughan
Department of Mathematics
Dartmouth College
rv2152@nyu.edu

Tyrus Berry
Department of Mathematical Sciences
George Mason University
tberry@gmu.edu

Dimitrios Giannakis
Dartmouth College
Dimitrios.Giannakis@dartmouth.edu

MS166

Deep Geometric and Topological Representation Learning

Diffusion condensation is a dynamic process that yields a sequence of multiscale data representations that aim to encode meaningful abstractions. It is effective for manifold learning, denoising, clustering, and visualization of high-dimensional data. Diffusion condensation is constructed as a time-inhomogeneous process where each step first computes and then applies a diffusion operator to the data. We theoretically analyze the convergence and evolution of this process from geometric, spectral, and topological perspectives. From a geometric perspective, we obtain convergence bounds based on the smallest transition probability and the radius of the data, whereas from a spectral perspective, our bounds are based on the eigenspectrum of the diffusion kernel. Our spectral results are of particular interest since most of the literature on data diffusion is focused on homogeneous processes. From a topological perspective, we show diffusion condensation generalizes centroid-based hierarchical clustering. To understand the evolution of the data geometry beyond convergence, we use topological data analysis. We show that the condensation process itself defines an intrinsic diffusion homology. We use this intrinsic topology as well as an ambient topology to study how the data changes over diffusion time. Our work gives theoretical insights into the convergence of diffusion condensation, and shows that it provides a link between topological and geometric data analysis.

Smita Krishnaswamy
Yale University
smita.krishnaswamy@yale.edu

MS166

Understanding Dataset Characteristics via Diffusion on Graph

Classical graph signal processing provides powerful techniques for understanding and modifying graph signals from the spectral domain, but they come with high computational costs. More recently, diffusion on graphs has been sought as an alternative approach to modifying graph signals; it is much more computationally efficient and is easy to interpret from the spatial perspective. Here, we present two different studies utilizing diffusion wavelets on a graph to filter graph signals for downstream analysis. In the first study, we aim to understand how and what is being utilized by Graph Neural Networks to achieve graph-related tasks. We do so by observing the performance difference between using the filtered graph and the original graph. We demonstrate that some image datasets, such as CIFAR and MNIST, rely on low-frequency signals; on the contrary, heterophilic datasets, such as WebKB, rely more heavily on high-frequency signals. In the second study on computational biology using gene interaction networks and gene expression data, we observe similar results where different frequency bands perform differently in a task-specific manner. In summary, our studies demonstrate the practical usage of graph diffusion to modify graph signals, leading to improved downstream prediction performance and a better understanding of the graph datasets' characteristics.

Renming Liu
Michigan State University, U.S.
liurenmi@msu.edu

MS166

Low Distortion Local Eigenmaps

In this talk I will present Low Distortion Local Eigenmaps (LDLE), a "bottom-up" manifold learning technique which constructs a set of low distortion local views of a dataset in lower dimension and registers them to obtain a global embedding. The local views are constructed using subsets of the global eigenvectors of the graph Laplacian and are registered using Procrustes analysis. The choice of these eigenvectors may vary across the regions. In contrast to existing techniques, LDLE is more geometric and can embed manifolds without boundary as well as non-orientable manifolds into their intrinsic dimension.

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

MS167

Beyond Conformal Prediction: Distribution-Free Uncertainty Quantification for Modern Machine Learning

As we begin deploying machine learning models in conse-

quential settings like medical diagnostics or self-driving vehicles, we need ways of knowing when the model may make a consequential error (for example, that the car doesn't hit a human). I'll be discussing how to generate rigorous, finite-sample confidence intervals for any prediction task, any model, and any dataset, for free. This will be a chalk talk. I will primarily discuss a flexible method called Learn then Test that works for a large class of prediction problems including those with high-dimensional, structured outputs (e.g. instance segmentation, multiclass or hierarchical classification, protein folding, and so on).

Anastasios N. Angelopoulos
University of California, Berkeley
angelopoulos@berkeley.edu

MS167

Chances and Limitations of Explainability of Deep Neural Networks

In this talk, we provide a theoretical framework for interpreting neural network decisions by formalizing the problem in an information theoretic framework. The solver of the associated optimization, coined Rate-Distortion Explanation (RDE), is then accessible to a mathematical analysis. We will discuss theoretical results as well as present numerical experiments showing that our algorithmic approach outperforms established methods, in particular, for sparse explanations of neural network decisions and is flexible enough to also be applied to challenging modalities such as from telecomunication. We further demonstrate that in combination with a wavelet decomposition, the extended approach, called CartoonX, can detect the reason for adversarial examples. We will finish with a general word of caution about neural networks trained on classical digital hardware, and present fundamental limitations, also for related interpretations.

Gitta Kutyniok
Ludwig-Maximilians-Universität München
kutyniok@math.lmu.de

MS167

Conformalized Online Learning

Modern machine learning algorithms have achieved remarkable performance in a myriad of applications, and are increasingly used to make impactful decisions in the hiring process, criminal sentencing, and healthcare diagnostics. The use of data-driven algorithms in high-stakes applications is exciting yet alarming: these methods are extremely complex, often brittle, and notoriously hard to analyze or interpret. Naturally, concerns have been raised about the reliability of the output of such machines. This talk focuses on making reliable predictions in an online setting, in which the underlying data distribution can drastically—and even adversarially—shift over time. We will introduce statistical tools that can be wrapped around any online "black-box' algorithm to provide valid and informative uncertainty estimates.

Yaniv Romano
EE Department, Technion
Israel
yromano@technion.ac.il

MS167

Individualized Conditional Independence Testing under Model-X with Heterogeneous Samples and Interactions

Model-X knockoffs and the conditional randomization test are methods that search for conditional associations in large data sets, controlling the type-I errors if the joint distribution of the predictors is known. However, they cannot test for interactions nor find whether an association is only significant within a latent subset of a heterogeneous population. We address this limitation by developing an extension of the knockoff filter that tests conditional associations within automatically detected subsets of individuals, provably controlling the false discovery rate for the selected hypotheses. Then, under the additional assumption of a partially linear model with a binary predictor, we extend the conditional randomization test as to make inferences about quantiles of individual effects that are robust to sample heterogeneity and interactions. The performances of these methods are investigated through simulations and with the analysis of data from a randomized blood donation experiment with several treatments.

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

Tianshu Sun
University of Southern California
tianshus@marshall.usc.edu

MS168

Algebraic Methods for Tensor Data

We develop algebraic methods for computations with tensor data. We give 3 applications: extracting features that are invariant under the orthogonal symmetries in each of the modes, approximation of the tensor spectral norm, and amplification of low-rank tensor structure. We introduce colored Brauer diagrams, which are used for algebraic computations and in analyzing their computational complexity. We present numerical experiments whose results show that the performance of the alternating least square algorithm for the low-rank approximation of tensors can be improved using tensor amplification.

Harm Derksen
Northeastern University
Department of Mathematics
ha.derksen@northeastern.edu

MS168

Tensor Decompositions: Algorithms and Uniqueness

Tensors are natural generalizations of matrices to higher-way arrays. Decompositions of tensors into sums of product(rank-one) tensors are useful in many areas for compressing and interpreting the information stored in a tensor. The tensor rank, which naturally generalizes the matrix rank, is the smallest number of product tensors that can decompose a given tensor. In contrast to matrices, tensor rank decompositions are often unique (up to trivialities). Uniqueness is useful in applications, as it corresponds to a unique interpretation of the information stored in a tensor. It is thus important to develop efficient algorithms to decompose tensors and to verify that the decomposition is unique. In this talk, we will review recent work on both fronts. This talk will not assume prior

knowledge of tensor decompositions.

Benjamin Lovitz
Northeastern University
benjamin.lovitz@gmail.com

MS168

Tensor-Based Frameworks in Cancer Genomics

The tumor microenvironment (TME) is a complex milieu around the tumor, whereby cancer cells interact with stromal, immune, vascular, and extracellular components. The TME is being increasingly recognized as a key determinant of tumor growth, disease progression, and response to therapies. We build a generalizable and robust tensor-based framework capable of integrating dissociated single-cell and spatially resolved RNA-seq data for a comprehensive analysis of the TME. Tensors are a generalization of matrices to higher dimensions. Tensor methods are known to be able to successfully incorporate data from multiple sources and perform a joint analysis of heterogeneous high-dimensional data sets. The methodologies developed as part of this effort will advance our understanding of the TME in multiple directions. These include cellular heterogeneity within the TME, crosstalks between cells, and tumor-intrinsic pathways stimulating tumor growth and immune evasion.

Neriman Tokcan
Broad Institute of MIT and Harvard
ntokcan@broadinstitute.org

MS169

Broader Engagement (BE): Tutorial on Best Practices and Tools For Secure Scientific Software Development

This tutorial will include components for evaluating design practices for creating secure software, appropriate software processes for managing secure software, threat modeling, and quality assurance testing using both static and dynamic analysis tools. The tutorial will be conducted with stakeholders from both industry and academia. Moreover, the training modules of the tutorial will be made available to community at large via open-source collaboration platforms. The research and the output of this tutorial participation will enable community members to analyze the security of scientific software using threat models such as, DREAD and STRIDE. The threat models will aid in identifying risks that often costs an organization thousands of dollars in patching vulnerabilities stemming from unsecured code. This tutorial will not only introduce the students and the software developers from various domains with the best practices for developing secure scientific software but also will facilitate the hands-on experience with penetration testing tools to mitigate threats such as, losing business and sensitive information due to a variety of potential vulnerabilities, thus will enable development of trustworthy and secure scientific software.

Nitin Sukhija
Slippery Rock University of Pennsylvania
nitin.sukhija@sru.edu

MS170

Broader Engagement (BE): Tutorial Hands On HPC Crash Course

There are many situations in nature that can only be thoroughly or efficiently studied through large massively parallel computer simulations enabled by High Performance Computers (HPC). For example, HPC simulations are the only way to fully explore the mechanisms that cause supernova, the catastrophic explosions that end massive stars lives and make the elements that form us. This symposium is designed to introduce science and computer science students to core concepts in High Performance Computing. It will consist of a mixture of lectures and self guided computing challenge exercises. The first part of the course will take place in Zoom, before the conference on September 22 from 1:00pm-4:00 pm EDT and is designed to give participants foundational skills in Linux, Vim and C. Please ¡a href="https://www.olcf.ornl.gov/siam-mds22-hands-on-hpc-crash-course/"¿ register here¡/a¿ for the virtual day by September 15th. The second part will take place at the conference where students will use the Summit Supercomputer to explore a series of guided challenge exercises with C and Python codes designed to introduce methods, such as the Message Passing Interface (MPI), and OpenMP, for parallel and accelerated computing. The target audience for this event are those new to HPC or those who need a refresher on basic HPC skills.

Leah Huk
Oak Ridge National Laboratory, U.S.
hukln@ornl.gov

MS170

Broader Engagement (BE): Tutorial Hands On HPC Crash Course

There are many situations in nature that can only be thoroughly or efficiently studied through large massively parallel computer simulations enabled by High Performance Computers (HPC). For example, HPC simulations are the only way to fully explore the mechanisms that cause supernova, the catastrophic explosions that end massive stars lives and make the elements that form us. This symposium is designed to introduce science and computer science students to core concepts in High Performance Computing. It will consist of a mixture of lectures and self guided computing challenge exercises. The first part of the course will take place in Zoom, before the conference on September 22 from 1:00pm-4:00 pm EDT and is designed to give participants foundational skills in Linux, Vim and C. Please ¡a href="https://www.olcf.ornl.gov/siam-mds22-hands-on-hpc-crash-course/"¿ register here¡/a¿ for the virtual day by September 15th. The second part will take place at the conference where students will use the Summit Supercomputer to explore a series of guided challenge exercises with C and Python codes designed to introduce methods, such as the Message Passing Interface (MPI), and OpenMP, for parallel and accelerated computing. The target audience for this event are those new to HPC or those who need a refresher on basic HPC skills.

Suzanne Parete-Koon
Oak Ridge National Laboratory, U.S.
paretekoonst@ornl.gov

MS171

Broader Engagement (BE): Tutorial Hands On HPC Crash Course

There are many situations in nature that can only be thoroughly or efficiently studied through large massively parallel computer simulations enabled by High Performance Computers (HPC). For example, HPC simulations are the only way to fully explore the mechanisms that cause super-

nova, the catastrophic explosions that end massive stars lives and make the elements that form us. This symposium is designed to introduce science and computer science students to core concepts in High Performance Computing. It will consist of a mixture of lectures and self guided computing challenge exercises. The first part of the course will take place in Zoom, before the conference on September 22 from 1:00pm-4:00 pm EDT and is designed to give participants foundational skills in Linux, Vim and C. Please ¡a href="https://www.olcf.ornl.gov/siam-mds22-hands-on-hpc-crash-course/"¿ register here¡/a¿ for the virtual day by September 15th. The second part will take place at the conference where students will use the Summit Supercomputer to explore a series of guided challenge exercises with C and Python codes designed to introduce methods, such as the Message Passing Interface (MPI), and OpenMP, for parallel and accelerated computing. The target audience for this event are those new to HPC or those who need a refresher on basic HPC skills.

Leah Huk
Oak Ridge National Laboratory, U.S.
hukln@ornl.gov

MS171

Broader Engagement (BE): Tutorial Hands On HPC Crash Course

There are many situations in nature that can only be thoroughly or efficiently studied through large massively parallel computer simulations enabled by High Performance Computers (HPC). For example, HPC simulations are the only way to fully explore the mechanisms that cause supernova, the catastrophic explosions that end massive stars lives and make the elements that form us. This symposium is designed to introduce science and computer science students to core concepts in High Performance Computing. It will consist of a mixture of lectures and self guided computing challenge exercises. The first part of the course will take place in Zoom, before the conference on September 22 from 1:00pm-4:00 pm EDT and is designed to give participants foundational skills in Linux, Vim and C. Please ¡a href="https://www.olcf.ornl.gov/siam-mds22-hands-on-hpc-crash-course/"¿ register here¡/a¿ for the virtual day by September 15th. The second part will take place at the conference where students will use the Summit Supercomputer to explore a series of guided challenge exercises with C and Python codes designed to introduce methods, such as the Message Passing Interface (MPI), and OpenMP, for parallel and accelerated computing. The target audience for this event are those new to HPC or those who need a refresher on basic HPC skills.

Suzanne Parete-Koon
Oak Ridge National Laboratory, U.S.
paretekoonst@ornl.gov

MS172

Broader Engagement (be): Guided Affinity Group Presentations and Wrap Up 2

The Broader Engagement program seeks to promote diversity and inclusion at the SIAM MDS22 conference and in the Mathematics of Data Science community. The program provides travel funding and specialized programming for CSE students and professionals from underrepresented and underprivileged backgrounds, and it invites members from the broader community to engage in activities and discussion that promote a diverse and welcoming environment. A key component to BE@MDS22 are the Guided Affinity Groups (GAGs) which are designed to help students get more out of SIAM MDS conference sessions. Led by volunteer community members, affinity groups explore conference topics from an entry-level perspective by meeting prior to conference sessions, attending conference sessions together, and meeting afterwards. BE attendees meet with affinity group leads virtually prior to the conference and then in-person before each day of the conference begins. In this final session, each group provides a 10-15 minute presentation on what was discovered and learned from the experience. In the presentation each group shares: What was their affinity group? What were the pedagogical goals of the team? Who was involved? (Leader, team, others?) What did they learn? What were the most effective ways to learn? What particular talks/researchers did they feel helped them? Whats next? All conference participants are invited to attend the final session.

Malena Espanol
Arizona State University
malena.espanol@asu.edu

MS172

Broader Engagement (BE): Guided Affinity Group Presentations and Wrap Up

The Broader Engagement program seeks to promote diversity and inclusion at the SIAM MDS22 conference and in the Mathematics of Data Science community. The program provides travel funding and specialized programming for CSE students and professionals from underrepresented and underprivileged backgrounds, and it invites members from the broader community to engage in activities and discussion that promote a diverse and welcoming environment. A key component to BE@MDS22 are the Guided Affinity Groups (GAGs) which are designed to help students get more out of SIAM MDS conference sessions. Led by volunteer community members, affinity groups explore conference topics from an entry-level perspective by meeting prior to conference sessions, attending conference sessions together, and meeting afterwards. BE attendees meet with affinity group leads virtually prior to the conference and then in-person before each day of the conference begins. In this final session, each group provides a 10-15 minute presentation on what was discovered and learned from the experience. In the presentation each group shares: What was their affinity group? What were the pedagogical goals of the team? Who was involved? (Leader, team, others?) What did they learn? What were the most effective ways to learn? What particular talks/researchers did they feel helped them? Whats next? All conference participants are invited to attend the final session.

Mary Ann E. Leung
Sustainable Horizons Institute
mleung@shinstitute.org

MS172

Broader Engagement (BE): Guided Affinity Group Presentations and Wrap Up

The Broader Engagement program seeks to promote diversity and inclusion at the SIAM MDS22 conference and in the Mathematics of Data Science community. The program provides travel funding and specialized programming for CSE students and professionals from underrepresented and underprivileged backgrounds, and it invites members from the broader community to engage in activities and discus-

sion that promote a diverse and welcoming environment. A key component to BE@MDS22 are the Guided Affinity Groups (GAGs) which are designed to help students get more out of SIAM MDS conference sessions. Led by volunteer community members, affinity groups explore conference topics from an entry-level perspective by meeting prior to conference sessions, attending conference sessions together, and meeting afterwards. BE attendees meet with affinity group leads virtually prior to the conference and then in-person before each day of the conference begins. In this final session, each group provides a 10-15 minute presentation on what was discovered and learned from the experience. In the presentation each group shares: What was their affinity group? What were the pedagogical goals of the team? Who was involved? (Leader, team, others?) What did they learn? What were the most effective ways to learn? What particular talks/researchers did they feel helped them? Whats next? All conference participants are invited to attend the final session

Aimee Maurais
Massachusetts Institute of Technology
maurais@mit.edu

MT1

Optimization for Data Analysis

Optimization has proved to be a rich source of techniques for formulating and solving computational problems that arise in data analysis and machine learning. This tutorial surveys problems in kernel learning, regression, graph analysis, neural networks, low-rank matrix analysis, and other areas that can be formulated as optimization problems over spaces of real vectors or matrices. We touch on the role of regularization in promoting useful solution structures. Finally, we describe the primary algorithmic techniques, focusing gradient and stochastic gradient methods.

Stephen Wright
University of Wisconsin-Madison
Dept. of Computer Sciences
swright@cs.wisc.edu

MT2

Broader Engagement (be): Introduction to Graph Neural Networks

Graph Neural Networks (GNNs) are considered a subset of deep learning methods designed to make predictions on graph representations. Most practical applications come from the areas of physics simulations, object detection and recommendation systems. Given the extended application areas, GNNs are one of fastest growing and most active research topic that attracts increasing attention not only from the machine learning and data science community, but from the larger scientific community. The materials for this tutorial will be selected for researchers with no prior knowledge of GNNs. Further reading, applications and most popular software packages and frameworks will be discussed.

Xiangyang Ju
Lawrence Berkeley National Laboratory, U.S.
XJu@lbl.gov

MT2

Broader Engagement (be): Introduction to Graph Neural Networks

Graph Neural Networks (GNNs) are considered a subset of deep learning methods designed to extract important information and make useful predictions on graph representations. Researchers have been working to adapt neural networks to operate on graph data for more than a decade. Most practical applications come from the areas of physics simulations, object detection and recommendation systems. Given the extended application areas, GNNs are one of fastest growing and most active research topic, that attracts increasing attention not only from the machine learning and data science community, but from the larger scientific community as well. The materials for this tutorial will be selected and organized for researchers with no prior knowledge of GNNs. Further reading, applications and most popular software packages and frameworks will also be discussed.

Alina Lazar
Youngstown University
alazar@ysu.edu

MT3

Broader Engagement (BE): Julia for Python Programmers

Julia is an open-source programming language targeting scientific computing and data science with many abstractions built into the language and ecosystem. It has been gaining popularity since its first stable release in 2018 with the promise of being an approachable language that compiles to efficient native code via LLVM, while providing rich capabilities for interactive data analysis and visualization making Julia an excellent language for technical and scalable computing. The objective of this tutorial is to provide the audience with enough hands-on experience in Julia to attend the Data Science with Julia session, and start developing their own applications.

Johannes P. Blaschke
Lawrence Berkeley National Laboratory
jpblaschke@lbl.gov

MT4

Broader Engagement (be): Data Science with Julia

Julia is an open-source programming language targeting scientific computing and data science. It has been gaining popularity since its first stable release in 2018 with the promise of being an approachable language that compiles to efficient native code via LLVM, while providing rich capabilities for interactive data analysis and visualization making Julia an excellent language to explore data analysis, statistics, and machine learning algorithms. The objective of this tutorial is to provide the audience with enough hands-on experience (covering several popular data science tools) in Julia to start applying it to their own projects.

Johannes P. Blaschke
Lawrence Berkeley National Laboratory
jpblaschke@lbl.gov

MT5

Geometric Deep Learning

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.

Michael Bronstein
University of Oxford, U.K.
michael.bronstein@cs.ox.ac.uk

MT6

Inverse Problems: Integrating Data with PDE-based Models under Uncertainty

Recent years have seen tremendous growth in the volumes of observational and experimental data. One question is: How do we extract knowledge from this data? When the data correspond to observations of physical systems (represented by mathematical models), this knowledge-from-data problem is fundamentally an inverse problem. This mini-tutorial aims to introduce the mathematical and computational aspects of inverse problems governed by partial differential equations (PDEs), particularly modern developments that emphasize the quantification of uncertainty in the inverse solution within the framework of Bayesian inference. The concepts introduced in this mini-tutorial will be put into practice in hands-on sessions using hIPPYlib.

Tucker Hartland
University of California, Merced
thartland@ucmerced.edu

MT6

Inverse Problems: Integrating Data with PDE-based Models under Uncertainty

Recent years have seen tremendous growth in the volumes of observational and experimental data. One question is: How do we extract knowledge from this data? When the data correspond to observations of physical systems (represented by mathematical models), this knowledge-from-data problem is fundamentally an inverse problem. This mini-tutorial aims to introduce the mathematical and computational aspects of inverse problems governed by partial differential equations (PDEs), particularly modern developments that emphasize the quantification of uncertainty in the inverse solution within the framework of Bayesian inference. The concepts introduced in this mini-tutorial will be put into practice in hands-on sessions using hIPPYlib.

Noemi Petra
University of California, Merced
npetra@ucmerced.edu

MT7

Computational Optimal Transport

I will propose in this mini tutorial an overview of novel approaches to compute optimal transport between measures, using a blend of recent techniques. After starting with the basics, describing the Monge and Kantorovich problems, I will show how convex optimization (using for instance regularization), automatic differentiation and neural networks can all prove useful to approximate optimal transport at scale, for various cost structures, and in high dimensional regimes.

Marco Cuturi
Apple, Inc., U.S.
marcocuturicameto@gmail.com

MT8

Tensor Decompositions: A Quick Tour of Illustrative Applications

Tensor decompositions are ubiquitous for analysis and dimensionality reduction of data. They have found application in areas such as neuroscience, market segmentation, hyperspectral image processing, network science, financial portfolio allocation, deep learning, quantum information theory, reinforcement learning, computer vision, drug design, energy demand forecasting, reduced-order models, etc. We will explain the fundamentals of tensor decomposition, focusing on the Tucker, Canonical Polyadic (CP), and Generalized CP (GCP) decompositions, with an emphasis on their application as a tool for unsupervised learning. We will walk through examples of applying tensor decomposition to several illustrative real-world datasets, giving attendees the opportunity to run existing software and analyze results themselves. This tutorial is appropriate for both novices and experts that are interested in better understanding applications of tensor decomposition. The examples will be available for attendees future use in their own classroom projects. The materials for this mini-tutorial will be made available via the following GITLAB repository to ¡a href=https://tinyurl.com/29mdk99x¿ download the materials in advance of the tutorial session, if possible. ¡/a¿

Grey Ballard
Wake Forest University
ballard@wfu.edu

MT8

Tensor Decompositions: A Quick Tour of Illustrative Applications

Tensor decompositions are ubiquitous for analysis and dimensionality reduction of data. They have found application in areas such as neuroscience, market segmentation, hyperspectral image processing, network science, financial portfolio allocation, deep learning, quantum information theory, reinforcement learning, computer vision, drug design, energy demand forecasting, reduced-order models, etc. We will explain the fundamentals of tensor decomposition, focusing on the Tucker, Canonical Polyadic (CP), and Generalized CP (GCP) decompositions, with an emphasis on their application as a tool for unsupervised learning. We will walk through examples of applying tensor decomposition to several illustrative real-world datasets, giving attendees the opportunity to run existing software and analyze results themselves. This tutorial is appropriate for both novices and experts that are interested in better understanding applications of tensor decomposition. The examples will be available for attendees future use in their own classroom projects. The materials for this mini-tutorial will be made available via the following GITLAB repository to ¡a href=https://tinyurl.com/29mdk99x¿ download the materials in advance of the tutorial session, if possible. ¡/a¿

Daniel M. Dunlavy
Sandia National Laboratories
Computer Science and Informatics

dmdunla@sandia.gov

MT8

Tensor Decompositions: A Quick Tour of Illustrative Applications

Tensor decompositions are ubiquitous for analysis and dimensionality reduction of data. They have found application in areas such as neuroscience, market segmentation, hyperspectral image processing, network science, financial portfolio allocation, deep learning, quantum information theory, reinforcement learning, computer vision, drug design, energy demand forecasting, reduced-order models, etc. We will explain the fundamentals of tensor decomposition, focusing on the Tucker, Canonical Polyadic (CP), and Generalized CP (GCP) decompositions, with an emphasis on their application as a tool for unsupervised learning. We will walk through examples of applying tensor decomposition to several illustrative real-world datasets, giving attendees the opportunity to run existing software and analyze results themselves. This tutorial is appropriate for both novices and experts that are interested in better understanding applications of tensor decomposition. The examples will be available for attendees future use in their own classroom projects. The materials for this mini-tutorial will be made available via the following GITLAB repository to ¡a href=https://tinyurl.com/29mdk99x¿ download the materials in advance of the tutorial session, if possible. ¡/a¿

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

PP1

Frequency-Domain Based Learning of Dynamical Systems from Purely Time-Domain Data

Constructing reduced-order models (ROMs) for large scale dynamical systems typically uses information about internal system matrices. In some cases, explicit access to these quantities are not available. Data-driven approaches provide a remedy in which the reduced-model is directly learned from input/output data. Frequency-based data-driven methods, which require access to values (and in some cases to derivatives) of the transfer function, have been very successful in constructing high-fidelity ROMs from data. However, at times this frequency domain data can be difficult to obtain or one might have only access to time-domain data. Recently, Burohman et.al. in 2020 introduced a framework to approximate transfer function values using only time-domain data. We first discuss improvements to this method to allow a more efficient and more robust numerical implementation. Then, we develop an algorithm that performs optimal-$\mathcal{H}_2$ approximation using purely time-domain data; thus significantly extending the applicability of $\mathcal{H}_2$-optimal approximation without a need for frequency domain samples. We also investigate how well other established frequency-based ROM techniques (such as Loewner framework and Vector Fitting) perform on this identified data.

Michael S. Ackermann
Virginia Tech
amike98@vt.edu

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

PP1

Conservation Aware Compact Model Extraction from Semiconductor PDE Data

Semiconductor physics modeling with Technology Computer Aided Design (TCAD) simulation tools are often utilized during design of individual electronic components. However, TCAD models are typically based on variations of finite element or finite volume numerical partial differential equation (PDE) discretization schemes, and so simulation of any more than a handful of coupled devices presents a computationally intractable challenge. For large-scale circuit modeling, it is necessary to utilize reduced order models, or compact models, counterpart to those TCAD PDE-based models which can faithfully reproduce the current-voltage responses of those devices. In this presentation, we will outline a method to produce machine-learned compact models trained from TCAD device simulation data. The tools presented include a data-driven discrete exterior calculus, which serves the purpose of coarse-graining the PDE model while ensuring expected conservation laws are satisfied.

Jonas Actor, Nathaniel Trask, Ravi Patel
Sandia National Laboratories
jaactor@sandia.gov, natrask@sandia.gov,
rgpatel@sandia.gov

Xiaozhe Hu
Tufts University
xiaozhe.hu@tufts.edu

Xujiao Gao
Sandia Natl. Labs
xngao@sandia.gov

Andy Huang
Sandia National Laboratories
ahuang@sandia.gov

PP1

Context-Aware Importance Sampling for Multi-Fidelity Bayesian Inference

Multi-fidelity methods leverage low-cost surrogate models to speed up computations and make occasional recourse to expensive high-fidelity models to establish accuracy guarantees. Because surrogate and high-fidelity models are used together, poor predictions by the surrogate models can be compensated with frequent recourse to high-fidelity models. Thus, there is a trade-off between investing computational resources to improve surrogate models and the frequency of making recourse to expensive high-fidelity models; however, this trade-off is ignored by traditional modeling methods that construct surrogate models that are meant to replace high-fidelity models rather than being used together with high-fidelity models. We present context-aware importance sampling which trades off increasing the fidelity of surrogate models for constructing more accurate biasing densities and the numbers of samples that are required from the high-fidelity models to compensate for poor biasing densities. Numerical examples demonstrate that such context-aware surrogate models for multi-fidelity importance sampling have lower fidelity than what typically is set as tolerance in traditional model reduction, leading to runtime speedups of up to one order of

magnitude in the presented examples.

Terrence Alsup
New York University
Courant Institute of Mathematical Sciences
alsup@cims.nyu.edu

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

PP1

Tracking Groundwater Contamination Plume Movement with Time Using Data Visualization Techniques

Recent research on nodal point positioning has been shown to have good success in modeling potential problems such as contaminant transport in surface or groundwater and also thermal environments, among other situations. These prior applications used a Nodal Positioning Algorithm (NPA) to aid in developing a distribution of nodal locations that reduces computational error. In this paper we show the advantages gained in extending the examined NPA by locating nodes using a probabilistic distribution coupled to nodal point coordinate development. We show that using a probabilistic coordinate generator allows for model computational nodes to be placed at random locations throughout the study region whereas use of more geometrically established nodes may be limited by the discretization spacing.

Michael Altenburg
United States Military Academy
michael.altenburg@westpoint.edu

Bryce Wilkins
Carnegie Mellon University
bwilkins@andrew.cmu.edu

Theodore Hromadka II
United States Military Academy
theodore.hromadka@westpoint.edu

PP1

Neural Network Kalman Filtering for 3D Object Tracking from Linear Array Ultrasound Data

Many interventional surgical procedures rely on medical imaging to visualise and track instruments. Such imaging methods not only need to be real-time capable, but also provide accurate and robust positional information. In ultrasound applications, typically only two-dimensional data from a linear array are available, and as such obtaining accurate positional estimation in three dimensions is non-trivial. In this work, we first train a neural network, using realistic synthetic training data, to estimate the out-of-plane offset of an object with the associated axial aberration in the reconstructed ultrasound image. The obtained estimate is then combined with a Kalman filtering approach that utilises positioning estimates obtained in previous time-frames to improve localisation robustness and reduce the impact of measurement noise. The accuracy of the proposed method is evaluated using simulations, and its practical applicability is demonstrated on experimental data obtained using a novel optical ultrasound imaging setup. Accurate and robust positional information is provided in real-time. Axial and lateral coordinates for out-of-plane objects are estimated with a mean error of 0.1mm for simulated data and a mean error of 0.2mm for experimental data. Three-dimensional localisation is most accurate for elevational distances larger than 1mm, with a maximum distance of 6mm considered for a 25mm aperture.

Arttu Arjas
University of Oulu
arttu.arjas@oulu.fi

Erwin J. Alles, Efthymios Maneas
University College London
e.alles@ucl.ac.uk, efthymios.maneas@ucl.ac.uk

Simon Arridge
Department of Computer Science
University College London
S.Arridge@cs.ucl.ac.uk

Adrien Desjardins
University College London
a.desjardins@ucl.ac.uk

Mikko J. Sillanpää
University of Oulu
mikko.sillanpaa@oulu.fi

Andreas Hauptmann
University of Oulu
University College London
andreas.hauptmann@oulu.fi

PP1

Linear and Nonlinear Models for Vitality Prediction of Potato Plants

Vitality quantifies a plants emergence and growth during the first several weeks influencing the plant success overall. While it is generally believed that vitality depends on the genotype and environment, it is not clear what exactly determines the vitality, whether it can be predicted, and which targeted treatments should be applied. Isolating the measurable qualities of a seed having the most significant influence on vitality is a gateway to environment-tailored seeds, increased yield and reduced agricultural waste. In collaboration with potato seed companies and academic plant biologists we analyze a broad spectrum of chemical and biological markers of potato seed tubers (FTIR, XRF, and HSI spectra, microbiome, metabolome) and compare their predictive power with respect to the plant vitality (Atza, Budko, Progr. Ind. Math. ECMI 2021). We utilize the data from the field trials of six genotypes, each characterized by 30 seed batches of unique origin grown in different regions of Europe over three years. While scrutinizing the industry-standard linear model we faced the following question: At what point does one conclude that an overparameterized linear model should be replaced by a nonlinear one? We shall present our considerations on this subject and discuss the performance of non-linear kernel models based on application-informed distance functions.

Elisa Atza
TU Delft
e.atza@tudelft.nl

Neil Budko
Delft University of Technology

n.v.budko@tudelft.nl

PP1

Using Fractal Geometry in Studying Biomedical Signals with Scale-Free Dynamics

Sometimes when we study signal regularity, we may see some phenomena that do not have characteristic scale. These scale free signals have been observed in biomedical signal processing, geophysics, finance, and internet traffic. To clarify more, when data is translation invariant, we may need to estimate autocorrelation or power spectral density (PSD) means that signal statistics like mean and variance do not change over time. In another hand, those signals that do not have characteristic scale, called scale-invariant signals that means the signal statistics do not change if we stretch or shrink the time axis. When we study translation invariant or scale invariant signals, or the signals with different scaling behaviour, we are not able to use the classical signal processing and we need to perform fractal analysis. When we study real world signals in biology, finance and so on, depending on scale and higher order moments, we may confront with signals that display nonlinear power-law behaviours. For these type signals, we need to apply multifractal analysis. In multifractal analysis we discover whether some type of power-law scaling exists for various statistical moments at different scales. In this study, we will review these type processes and will provide some examples to compare the multifractal spectra of biomedical databases.

Tahmineh Azizi
Kansas State University
tazizi@ksu.edu

PP1

Extremal Event Graphs: A (Stable) Tool for Analyzing Noisy Time Series

Local maxima and minima, or extremal events, in experimental time series can be used as a coarse summary to characterize data. However, the discrete sampling in recording experimental measurements suggests uncertainty on the true timing of extrema during the experiment. This in turn gives uncertainty in the timing order of extrema within the time series. Motivated by applications in genomic time series and biological network analysis, we construct a weighted directed acyclic graph (DAG) called an extremal event DAG using techniques from persistent homology that is robust to measurement noise. Furthermore, we define a distance between extremal event DAGs based on the edit distance between strings. We prove several properties including local stability for the extremal event DAG distance with respect to pairwise L_∞ distances between functions in the time series data. Lastly, we provide algorithms, publicly free software, and implementations on extremal event DAG construction and comparison.

Robin Belton
Smith College, U.S.
rbelton@smith.edu

PP1

Using Reinforcement Learning to Optimize Cryptocurrency Portfolio

Given a cryptocurrency, this paper aims to find the best approach in modeling the price in order to optimize the portfolio according to the price changes of its current inventory over time. There exists a pattern within the data which will allow us to predict which action (buy/sell/hold) to take regarding each asset in the portfolio in order to optimize returns through minimizing risk and maximize profit. Data will be modeled using sequential modeling, probabilistic sequence modeling, and advanced neural networks to create a decision making process with reinforcement learning. The machine learning algorithm will utilize the different models efficiently to ultimately decide the best course of action that can be taken, given an asset with respect to time in order to optimize the rewards function.

Neila Bennamane
Florida International University, U.S.
nbenn017@fiu.edu

PP1

Semi-Supervised Learning for Aggregated Multilayer Graphs Using Diffuse Interface Methods and Fast Matrix Vector Products

We generalize a graph-based multiclass semi-supervised classification technique based on diffuse interface methods to multilayer graphs. Besides the treatment of various applications with an inherent multilayer structure, we present a very flexible approach that interprets high-dimensional data in a low-dimensional multilayer graph representation. Highly efficient numerical methods involving the spectral decomposition of the corresponding differential graph operators as well as fast matrix-vector products based on the nonequispaced fast Fourier transform enable the rapid treatment of large and high-dimensional data sets. We perform various numerical tests putting a special focus on image segmentation. In particular, we test the performance of our method on data sets with up to 10 million nodes per layer as well as up to 104 dimensions, resulting in graphs with up to 52 layers. While all presented numerical experiments can be run on an average laptop computer, the linear dependence per iteration step of the runtime on the network size in all stages of our algorithm makes it scalable to even larger and higher-dimensional problems.

Kai Bergermann, Martin Stoll, Toni Volkmer
Department of Mathematics,
Technische Universitaet Chemnitz
kai.bergermann@math.tu-chemnitz.de,
martin.stoll@math.tu-chemnitz.de,
toni.volkmer@math.tu-chemnitz.de

PP1

Automatic Computing of Second-Order Edgeworth Expansions and Related Quantities of Any Function of the Mean of An Iid Sample of An Absolutely Continuous Distribution

We designed a completely automated Maple (15) worksheet for deriving Edgeworth and Cornish-Fisher expansions as well as the acceleration constant of the bootstrap bias-corrected and accelerated technique. It is valid for non-parametric or parametric bootstrap, of any (studentized) statistics that is a -regular enough- function of the mean of an iid sample of an absolutely continuous distribution. In addition to verifying known complex second-order Edgeworth and Cornish-Fisher expansions of the t-distribution, Theorem 11.4.2 by Lehman and Romano in [Lehman and Romano, 2005, p. 460], we successfully applied the worksheet to a complex maximum likelihood esti-

mator as a first step to deriving more accurate confidence intervals in order to enhance quality controls. The worksheet also features the export of Maple results into R code. In addition, we provide R code to plot these expansions as well as their increasing rearrangements. All these supplemental materials are available upon request.

Frédéric Bertrand, Myriam Maumy
Troyes Technology University
frederic.bertrand@utt.fr, myriam.maumy@utt.fr

PP1

Scalable Spike-and-Slab

Spike-and-slab priors are commonly used for Bayesian variable selection, due to their interpretability and favorable statistical properties. However, existing samplers for spike-and-slab posteriors incur prohibitive computational costs when the number of variables is large. In this article, we propose *Scalable Spike-and-Slab* (S^3), a scalable Gibbs sampling implementation for high-dimensional Bayesian regression with the continuous spike-and-slab prior of George and Mcculloch (1993). For a dataset with n observations and p covariates, S^3 has $\mathcal{O}(\max\{n^2 p_t, np\})$ computational cost at iteration t where p_t does not exceed the number of covariates switching spike-and-slab states between iterations t and $t-1$ of the Markov chain. This improves upon the $\Omega(n^2 p)$ per-iteration cost of state-of-the-art implementations as, typically, δ_t is substantially smaller than p. We apply S^3 on synthetic and real-world datasets, demonstrating orders of magnitude speed-ups over existing exact samplers and significant gains in inferential quality over approximate, asymptotically biased samplers with comparable cost.

Niloy Biswas, Xiao-Li Meng
Harvard University
niloy_biswas@g.harvard.edu, meng@stat.harvard.edu

Lester Mackey
Microsoft Research New England
lmackey@stanford.edu

PP1

Geometric Anomalies and Mass Spectral Libraries

Mass spectrometry is an important area of analytical chemistry that studies the measurement of mass-to-charge ratios of molecules present in a sample of interest. These measurements are assembled into an ion signal as a function of the mass-to-charge ratio to yield a mass spectrum. The resulting mass spectra are then used to identify molecules present in a measured sample by comparison to a library of mass spectra. Comparisons between sample and library spectra are made using a search function that approximates the similarity/dissimilarity between two mass spectra. There are a variety of search functions in use, but a search function based on the cosine between spectra is commonly employed. Mass spectral libraries are important reference data for researchers employing mass spectrometry-based methods for analysis, and the need for high quality libraries is ubiquitous in industries like pharmaceutical development, forensic applications and chemometrics. In this poster we investigate the use of geometric structures in these libraries to detect anomalies. In this straight-forward approach we investigate decision thresholds and how they affect the detection of outliers in these large libraries of compound mass spectra. We consider examples from various disciplines including forensics and food science.

Danielle Brager
National Institute of Standards and Technology
danielle.brager@nist.gov

Anthony J. Kearsley
Applied and Computational Mathematics Division
National Institute for Standards and Technology
anthony.kearsley@nist.gov

Edward Erisman
National Institute of Standards and Technology
edward.erisman@nist.gov

PP1

Using Data Visualization Techniques to Compare Four Different Water Flooding Estimates

In this presentation we demonstrate the use of data visualization techniques to depict the similarities and differences between four different methods of estimating floodwater elevations computed to model a severe flooding condition. Because the computational methods employed to estimate floodwater elevations are data intensive and develop considerable outcome results, visualization of these different data outcomes provide valuable help in better understanding the similarities and differences between modeling estimates depending on the computational method used. Exhibits are provided that depict the water surface elevations within the severe storm floodplain.

Morgan Brown
United States Military Academy
morgan.brown@westpoint.edu

Bryce Wilkins
Carnegie Mellon University
bwilkins@andrew.cmu.edu

Theodore Hromadka II
United States Military Academy
theodore.hromadka@westpoint.edu

PP1

Spatial Effect Removal by Manifold Transformation

A good estimate of the expectation and variance of a phenotype parameter requires observation of a statistically significant number of plants. Since plants cannot grow too close to each other, a typical field trial has a spatial scale where the spatial uniformity of soil properties and other factors can not be maintained. Therefore, one cannot consider the mean of, say, canopy size, for plants growing at different locations in the test field to be a good estimate of the expected canopy. It is recognized that the simple row-column random-effect model does not capture the actual spatial heterogeneity and a more sophisticated mixed spline-based model has been proposed (Rodrguez-lvarez et al, Spatial Statistics, 23, 5271, 2018). Here we reformulate the problem by considering the manifold transformation of the probability density function of a phenotype parameter between different spatial locations. Limiting the class of transformations, we derive statistical models applicable to the common split-block design. The model requires spatial regularization and we consider the gradient minimizing regularizer implemented with the finite-element method and

a kernel-based technique. Using both synthetic and real data we show that the optimal regularization parameters can be determined by cross validation leading to successful removal of field effects with complex spatial patterns.

Neil Budko
Delft University of Technology
n.v.budko@tudelft.nl

Elisa Atza
TU Delft
e.atza@tudelft.nl

PP1

Towards a Mathematical Model for Human Reliance and Trust in Automation

Currently, most of the autonomous vehicles are designed to be used in conjunction with human operators. Since the driving performance of humans can degrade with factors like stress and workload, the decision to assign some tasks to an automation can lead to better driving performance and safety. We trained a machine learning model, with data from a previous study on simulated driving with an automated assistant, to predict whether a human operator will use the driving assistance at a given time. We obtained the most important features driving human reliance on automation and use them to make associations between reliance and trust. Additionally, we analyze the dynamics of reliance on automation under different conditions and propose a framework to model human reliance and trust in automation.

Carlos Enrique Bustamante Orellana
Arizona State University, U.S.
cbustam3@asu.edu

PP1

Approximate Matrix Multiplication and Laplacian Sparsifiers

A ubiquitous operation in data science and scientific computing is matrix multiplication. However, it presents a major computational bottleneck when the matrix dimension is high, as can occur for large data size or feature dimension. A common approach in approximating the product, is to subsample row vectors from the two matrices, and sum the rank-1 outer products of the sampled pairs. We propose a sampling distribution based on the leverage scores of the two matrices. We give a characterization of our approximation in terms of the Euclidean norm, analogous to that of a ℓ_2-subspace embedding. We then show connections between our algorithm; CR-multiplication, with Laplacian spectral sparsifiers, which also have numerous applications in data science, and how approximate matrix multiplication can be used to devise sparsifiers. We also review some applications where these approaches may be useful.

Neophytos Charalambides
University of Michigan, Ann Arbor
Department of Electrical Engineering and Computer Science
neochara@umich.edu

Mert Pilanci
Department of Electrical Engineering
Stanford University
pilanci@stanford.edu

Alfred O. Hero
The University of Michigan
Dept of Elect Eng/Comp Science
hero@eecs.umich.edu

PP1

Classix: Fast and Explainable Clustering Based on Sorting

We introduce a fast and explainable clustering method called CLASSIX, which shares features of both distance and density based clustering methods. It consists of two phases, namely a greedy aggregation phase of the sorted data into groups of nearby data points, followed by the merging of groups into clusters. The algorithm is controlled by two main parameters, namely a distance parameter for the aggregation and another parameter controlling the minimal cluster size. The algorithm has linear space complexity and achieves near linear time complexity on a wide range of problems. Its inherent simplicity allows for the generation of intuitive explanations of the computed clusters. The software and experimental data are publicly available at https://github.com/nla-group/classix.

Xinye Chen, Stefan Güttel
The University of Manchester
xinye.chen@postgrad.manchester.ac.uk, stefan.guettel@manchester.ac.uk

PP1

Classification of Acute Lymphoblastic and Acute Myeloid Leukemia: Exploration of Binary Classification Methods for Small-Unbalanced Dataset

Accurate classification of different types of leukemia is necessary since the effectiveness of treatment methods varies on different leukemia types. Inspired by T. R. Golubs paper published on Science by October 1999 that developed a prediction procedure for Acute Lymphoblastic Leukemia(ALL) and Acute Myeloid Leukemia(AML): selecting influential genes by neighborhood analysis and use these genes for weighted calculation, this study aimed at finding another concise cancer type prediction procedure by exploring common feature selection methods and binary classification algorithms. This study used the same DNA microarray results dataset of Golubs paper collected from 47 ALL and 25 AML patients. Two rounds of feature selection were performed. We first removed genes classified as absent according to DNA microarray method for over 50% samples in both types, and then used ANOVA-Linear Support Vector Classification Pipeline to find top 50 discriminative genes. We tried Biased Random Forest classifier (BRAF), Synthetic Minority Oversampling Technique (SMOTE), and Multi-layer neural network on selected genes. After evaluation and comparison of the three methods, Neural network comes up to have the best behavior of prediction accuracy around 90% and might be competent for a reference method in similar small-unbalanced data binary classification problems.

Xuanhe Chen
Columbia University, U.S.
xc2610@cumc.columbia.edu

PP1

Implicitly Regularized Empirical Risk Minimiza-

tion

Intuitive yet effective, empirical risk minimization plays a crucial role in not only statistical learning theory but also in practice. The realization of this principles takes different form depending on the specific model and algorithm used. Implicit regularization, a phenomenon that is observed more and more frequently, suggests that there is a built-in regularization in gradient descent and empirical risk minimization. However, the connections remain vague and lack theoretical understanding. I will present a precise description and theoretical results in a general setting, which can be used in various fields. I will do so by analyze the implicitly regularized empirical risk minimization problem (IRERM).

Hung-Hsu Chou
RWTH Aachen University
chou@mathc.rwth-aachen.de

Johannes Maly, Dominik Stöger
Catholic University of Eichstaett-Ingolstadt, Germany
johannes.maly@ku.de, dominik.stoeger@ku.de

PP1

Multi-Resolution Dynamic Mode Decomposition for Damage Detection in Wind Turbine Gearboxes

Classical methods for damage detection in gearboxes have proven to be successful in scenarios with steady operating conditions. Unfortunately, when considering wind turbine gearboxes, standard data analysis techniques are strongly challenged by the presence of wind fluctuations. In this talk, we will discuss how to cope with the stochastic nature of the arising sensor signals due to wind turbulence by introducing an approach for damage detection in gearboxes based on the equation-free linear characterization of the sensor datas stochastic temporal evolution with the multi-resolution dynamic mode decomposition (mrDMD). We analyze data stemming from a simulated vibration response of a simple nonlinear gearbox model in a healthy and damaged scenario and under different wind conditions. Our approach exploits the mrDMD capability to find multiscale linear characterizations of nonlinear dynamical systems by describing the temporal evolution of the systems state as a weighted sum of terms, each associated with a known temporal coefficient that evolves with time. With mrDMD, we can extract components in the analyzed vibration signals that capture their geometrical characteristics in the time-domain at different time scales, highlighting features related to damage and enabling its identification. Our results show how the proposed approach can overcome the issues experienced by classical damage detection techniques, such as Fourier analysis and Empirical Mode Decomposition.

Paolo Climaco
Institut für Numerische Simulation, Universität Bonn
climaco@ins.uni-bonn.de

Jochen Garcke
Institut für Numerische Simulation, Universität Bonn
Fraunhofer SCAI
garcke@ins.uni-bonn.de

Rodrigo Iza-Teran
Fraunhofer SCAI
victor.rodrigo.iza.teran@scai.fraunhofer.de

PP1

Depression and Its Socio-Economic Factors among Racial/ethnic Group

During the course of this semester, we were fortunate to perform research on depression among racial/ethnic groups. While brainstorming on what exactly to focus on, like our topic, research question, hypothesis, and so on, we wanted to focus on factors that we could all relate to. We were trying to decide on specific factors related to depression and the different racial groups and we came up with this question, What is the prevalence of depression among people of different socioeconomic and racial/ethnic groups? What are the different factors in racial/ethnic groups that cause depression?

Oluwapelumi Marvelous Dada
Huston-Tillotson University, U.S.
omdada@htu.edu

PP1

Breast Cancer Prediction Using Machine Learning Algorithms - A Deep Learning Approach

Breast Cancer is the deadliest and commonly diagnosed cancer in women globally. Early diagnosis and treatment of breast cancer increases the chance of a five-year survival rate by 99%. Recent technological and computational advancement have led to the discovery of machine learning algorithms for the analysis of complex data. Machine learning algorithms have been widely applied for the analysis of breast cancer data. In this paper, we propose to implement machine learning algorithms, using a deep learning approach, for automatic detection and prediction of breast cancer using mammogram images. To achieve this, we implement transfer learning on a deep learning algorithm called Convolutional Neural Network (CNN). Two datasets of breast cancer images are analyzed using three CNN models in existing deep learning frameworks. The models perform a binary and multiclass classification task on the images. Experimental results showed that CNN models can accurately identify and predict breast cancer when provided with a large and balanced dataset.

Darlington S. David
University of Liberia, College of Engineering
davidds@ul.edu.lr

Darnette M. Herron, TITUS G. Gooding
University of Liberia
darnetteherron@gmail.com, tgoodingsr@gmail.com

PP1

Detailed Description of Phase Transitions of Non-Magnetic Body-Centered Cubic (bcc) Iron at High Pressure Through a Marginalized Graph Kernel

Gaussian process regression on a molecular graph kernel (Tang, 2019) was performed to construct a relation between the par-atom forces and displacements of a non-spin-polarized BCC iron (Fe) ab initio simulation in the direction of their first, second and third nearest neighbors. When training on 400 randomly selected molecular dynamics steps, the standard deviation of errors in the testing predictions of the total system's potential was of 8meV, while the mean of this errors was 5meV. The ma-

chine learning potential of a model resulting from training is used to approximate the forces acting upon each atom in the displacement axes of interest through numerical differentiation according to the central difference approximation and through derivation of a polynomial fit to the potential functional given by the model. The resulting forces are consistent within approximation techniques and show stability in the second nearest-neighbors' axes, but not for the first nearest neighbors. BCC Iron is unstable without magnetism (Heine, 2019), but these observations suggest that the effect of magnetism in the stability of the system of interest is not noticeably affected in the second nearest-neighbor interactions. By analyzing the system-wide tendencies of the resulting force curves in the studied axes for the output of a DFT simulation of high temperature, high pressure BCC Fe lattice, a phase transition into an hexagonal closed packed lattice can be inferred.

Adrian De La Rocha Galan
The University of Texas at El Paso, U.S.
adelaroch@miners.utep.edu

PP1

Extracting Thermodynamic and Fluorescent Properties of Intercalating Dyes from Temperature-Programmed PCR Measurements with Modeling, Optimization, and Uncertainty Quantification

Fluorescent dyes that intercalate with DNA are commonly used in real-time Polymerase Chain Reaction (PCR) instruments. In this poster, we present a mathematical model which provides a quantitative relationship between the measured fluorescence signal and underlying biochemical phenomena. The model accounts for the partitioning of dye between solution and DNA, as well as the fluorescence of intercalated dye. A fixed number of DNA strands, subject to a series of step decreases in temperature, is considered. The model predicts a measured signal to be linear in total dye concentration. The experimental measurements possess trends that deviate from linearity, due to noise both within the same 96-well plate and between different plates. Numerical optimization allows for ascertainment of the linear trends in the experimental measurements, accounting for error both in signal and total dye concentration. Uncertainty quantification and propagation of optimal solutions associated with different DNA concentrations allows for calculation of dye partitioning coefficients and adsorbed molar fluorescences. It also sheds new light on the role of noise in total dye concentration in this important measurement technology. The temperature dependence of the partitioning coefficients, in turn, allows for calculation of thermodynamics of intercalation. All properties calculated separately for single-stranded and double-stranded DNA are consistent with biochemical intuition.

Robert Dejaco, Jacob Majikes
National Institute of Standards and Technology
robert.dejaco@nist.gov, jacob.majikes@nist.gov

Paul Patrone
NIST Gaithersburg, Information Technology Lab
Applied and Computational Mathematics Division
paul.patrone@nist.gov

James Liddle
National Institute of Standards and Technology
james.liddle@nist.gov

Anthony J. Kearsley
Applied and Computational Mathematics Division
National Institute for Standards and Technology
anthony.kearsley@nist.gov

PP1

Distributional Reinforcement Learning for Supply Chain Optimization

Reinforcement learning through policy optimization is an effective method to solve stochastic sequential decision-making problems. It so happens that supply chains can be modelled as these types of systems. In this work, we propose a reinforcement learning methodology specifically tailored for supply chain problems which efficiently optimizes complex supply chains. The proposed algorithm can balance the exploration-exploitation paradigm in the search process, while being efficient and avoiding most local optima. Additionally, we construct a novel distributional reinforcement learning framework to learn a risk-sensitive policy by optimizing the conditional value-at-risk. The capabilities of our proposed algorithm are tested on a multi-echelon supply chain problem. The results show that the policy identified by our method outperforms policy gradient methods and its robust risk-aversion policy can offer more protection from unlikely, but catastrophic scenarios. Finally, we conducted a sensitivity analysis of the algorithms parameters within the distributional reinforcement learning framework.

Ehecatl Antonio Del Rio Chanona
Imperial College London
a.del-rio-chanona@imperial.ac.uk

Sophie Sitter
Sargent Centre for Process Systems engineering
sophie.sitter20@imperial.ac.uk

PP1

Two-Sample Neighborhood Hypothesis Tests for High Dimensional Data

Because data is increasingly high- or infinite-dimensional, traditional statistical methodology for formally comparing mean vectors of two populations often fails, especially when the dimensionality of the data exceeds the sample size. Building upon the methodology of Munk et al (2008), we propose a novel nonparametric methodology for testing the approximate equality of two mean vectors. The key to our approach is the use of a neighborhood hypothesis, which relaxes the standard hypothesis of means being equal to each other means instead of only being within a predetermined distance from each other. We present test statistics for checking such claims using both asymptotic results and the nonparametric bootstrap. Finally, we use simulation studies to illustrate the performance of our method in various scenarios, including when the sample size is less than the dimensionality of the data. This is joint work with L. Ellingson.

Dulanjalee Devage Dona, Leif Ellingson
Texas Tech University
dulanjalee-damitha.karunaratne@ttu.edu,
leif.ellingson@ttu.edu

PP1

Data Based Learning Symmetric Hamiltonian

Lagrangian and Hamiltonian systems are commonly used

to model mechanical systems. For complex systems, modeling a Hamiltonian requires large amounts of specific domain knowledge and expertise. In the last decades, technological improvements have allowed for increasingly complex systems and thus required an increasing amount of modeling effort. Likewise, the ability of gathering, refining and using data increased during that time, too. For a beneficial use of this data, deep learning methods are commonly used that take into account the structural properties of the Hamilton or Lagrange function, respectively. These, however often neglect symmetries that mechanical systems often possess. In our research, we extend the classical Hamiltonian Neural Networks [Greydanus et al., 2019] or Lagrange Neural Networks [Cranmer et al., 2020], respectively, to discover and preserve additional symmetries. If a system's symmetry is known, we introduce this specific knowledge as an additional condition to the Hamiltonian function during the learning process. However, if this specific knowledge is not available, a matching symmetry group action can be learned simultaneously with learning the Hamiltonian or Lagrangian function, respectively. For this purpose, we learn a basis of the underlying Lie algebra of the symmetry. We evaluate the proposed approach by various mathematical examples, such as the inverted pendulum and the Kepler problem.

Eva Dierkes
Center for Industrial Mathematics
University of Bremen, Germany
eva.dierkes@uni-bremen.de

Kathrin Flaßkamp
Systems Modeling and Simulation
Saarland University, German
kathrin.flasskamp@uni-saarland.de

PP1

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

PP1

Vemos - Visual Explorer for Metrics of Similarity

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

PP1

Sample Efficiency of Data Augmentation Consistency Regularization

Data augmentation is popular in the training of large neural networks; currently, however, there is no clear theoretical comparison between different algorithmic choices on how to use augmented data. In this paper, we take a step in this direction - we first present a simple and novel analysis for linear regression, demonstrating that data augmentation consistency (DAC) is intrinsically more efficient than empirical risk minimization on augmented data (DA-ERM). We then propose a new theoretical framework for analyzing DAC, which reframes DAC as a way to reduce function class complexity. The new framework characterizes the sample efficiency of DAC for various non-linear models (e.g., neural networks). Further, we perform experiments that make a clean and apples-to-apples comparison (i.e., with no extra modeling or data tweaks) between ERM and consistency regularization using CIFAR-100 and WideResNet; these together demonstrate the superior efficacy of DAC.

Yijun Dong
Oden Institute
The University of Texas at Austin
ydong@utexas.edu

Shuo Yang
The University of Texas at Austin
yangshuo_ut@utexas.edu

Rachel Ward
Department of Mathematics
University of Texas at Austin
rward@math.utexas.edu

Inderjit S. Dhillon
University of Texas at Austin
inderjit@cs.utexas.edu

Sujay Sanghavi
Electrical and Computer Enigneering
University of Texas at Austin
sanghavi@mail.utexas.edu

Qi Lei
Princeton Unversity
qilei@princeton.edu

PP1

Long Range Dependence in Earthquake Occurrence Rates

Low-frequency earthquakes (LFEs) are small magnitude (less than 2) earthquakes, with reduced amplitudes at frequencies greater than 10 Hz relative to ordinary small earthquakes. They are usually grouped into families of events, with all the earthquakes of a given family originating from the same small patch on the plate interface and recurring more or less episodically in a bursty manner. First, for each LFE family, I translate the list of earthquake occurrence times into a discrete time series defined by the number of events per unit of time. I then look for evidence of long-range dependence in the event rate time series. For each family of LFEs, I compute the value of the fractional index d, which represents how fast the variance in the number of LFEs increases with the length of the time window considered. For most of the LFE families studied, I find that 0 ¡ d ¡ 0.5, which is characteristic of long-range dependence in the time series. Second, I consider all the LFEs associated with a single family as a point process. I then use recurrent neural networks to model the conditional intensity function associated with each point process and I attempt to forecast future occurrences of LFEs given the timings of past events.

Ariane Ducellier
University of Washington
ducela@uw.edu

PP1

Amortized Learning in Physical Models: A Cryo-Em Test Case

Variational autoencoders (VAEs) have become a popular generative model due their ability to learn efficient maps from samples to proposed distributions. Recently, VAEs have been used for amortized learning of latent variables in physical models. Amortized learning allows for significant computational speed-up by learning a deep neural network to approximate these distributions for all samples instead of performing variational inference on each sample. Cryogenic electron microscopy (cryo-EM) is a technique to reconstruct 3D biomolecules using hundreds of thousands of tomographic projection images with unknown viewing directions. We will present a test case in cryo-EM and discuss the advantages of VAEs and challenges in using them in various applications.

Daniel G. Edelberg, Roy Lederman
Yale University
daniel.edelberg@yale.edu, roy.lederman@yale.edu

PP1

Simple Control for Complex Pandemics

The COVID-19 pandemic began over two years ago, yet schools, businesses, and other organizations are still struggling to keep the risk of disease outbreak low while returning to (near) normal functionality. Observations from these past years suggest that this goal can be achieved through the right balance of mitigation strategies, which may include some combination of mask use, vaccinations, viral testing, and contact tracing. The choice of mitigation measures will be uniquely based on the needs and available resources of each organization. This article presents practical guidance for creating these policies based on an analytical model of disease spread that captures the combined effects of each of these interventions. The resulting guidance is tested through simulation across a wide range of parameters and used to discuss the spread of disease on college campuses.

Sarah C. Fay, Dalton Jones
Massachusetts Institute of Technology
scfay@mit.edu, daltonj@mit.edu

Munther Dahleh
MIT
dahleh@mit.edu

Anette "Peko" Hosoi
Massachusetts Institute of Technology
peko@mit.edu

PP1

Analysis of Goal, Feedback and Rewards on Sustained Attention via Ml

Sustained attention is an umbrella term used in the field of cognitive psychology generally referring to a subjects readiness to detect unpredictably occurring signals over prolonged periods. In a recent experiment, reaction times (RTs) and pupillometry data were recorded from 350 subjects in a 30-minute vigilance task. Subjects were also presented with different types of goal, feedback, and reward. In this study, we revisit this experimental data and solve three families of machine learning problems: (i) RT-regression problems, to predict subjects' RTs using all available data, (ii) RT-classification problems, to classify responses more broadly as attentive, semi-attentive, and inattentive, and (iii) to predict the subjects' experimental conditions from physiological data. After establishing that regressing RTs is a difficult task, we achieve better results classifying them in broader categories. We also successfully disambiguate subjects who received goals and rewards from those who did not. Finally, we quantify changes in accuracy when coarser features (averaged throughout trials) are used. Interestingly, the the machine learning pipeline selects different features depending on their resolution, suggesting that predictive physiological features are also resolution-specific. We use state-of-the-art machine learning pipelines for data curation/imputation/standardization, feature selection, and hyperparameter tuning along with a large variety of classifiers and regressors.

Thiwanka N. Fernando
University of Texas at Arlington
nethali.fernando@uta.edu

PP1

Gauss-Legendre Features for Gaussian Process Regression

Gaussian processes provide a powerful probabilistic kernel learning framework, which allows learning high qual-

ity nonparametric regression models via methods such as Gaussian process regression. Nevertheless, the learning phase of Gaussian process regression requires massive computations which are not realistic for large datasets. In this paper, we present a Gauss-Legendre quadrature based approach for scaling up Gaussian process regression via a low rank approximation of the kernel matrix. We utilize the structure of the low rank approximation to achieve effective hyperparameter learning, training and prediction. Our method is very much inspired by the well-known random Fourier features approach, which also builds low-rank approximations via numerical integration. However, our method is capable of generating high quality approximation to the kernel using an amount of features which is poly-logarithmic in the number of training points, while similar guarantees will require an amount that is at the very least linear in the number of training points when using random Fourier features. Furthermore, the structure of the low-rank approximation that our method builds is subtly different from the one generated by random Fourier features, and this enables much more efficient hyperparameter learning. The utility of our method for learning with low-dimensional datasets is demonstrated using numerical experiments.

Paz Fink Shustin, Haim Avron
Tel Aviv University
pfink22@gmail.com, haimav@tauex.tau.ac.il

PP1

A Population Approach to the Calibration of a Cell Volume Regulation Model Using Qualitative Information

Mammalian cells have water permeable membranes and therefore their volume is constantly challenged even by their own physiological activity. Cells have therefore developed powerful mechanisms for regulating their volume, which involve in particular the active and passive transport of ions and water across their plasma membrane. Various dynamical models describing the coupling of water and ions have been proposed to describe cell volume regulation. However, due to their specific constraints and difficulties, these models of ion and water homeostasis are generally not calibrated with experimental data, leaving many key parameters hand-tuned. Inferring those parameters from data could help describe important biophysical properties of specific membrane elements involving in volume regulation. In this work, we show how we can get around the difficulties inherent to this inference problem such as a lack of identifiability and heterogeneity of such biophysical properties. We integrate the qualitative information of regulated physiological variables as soft constraints and add a regularization penalty to the calibration problem to identify key parameters. Furthermore, a multi-experiment approach is adopted to combine the many ways cell volume can be challenged. This population approach allows for the estimation of the expected value of key parameters related to water and ion transport and subsequently to cell volume regulation.

Pierre-Louis Gagnon, Nicolas Doyon, Elsa Rousseau, Pierre Marquet
Université Laval
pierre-louis.gagnon.1@ulaval.ca, nicolas.doyon@mat.ulaval.ca, elsa.rousseau@ift.ulaval.ca, pierre.marquet@neuro.ulaval.ca

PP1

Operator Inference for Non-Intrusive Model Reduction with Quadratic Manifolds

Traditional model reduction techniques often fail to identify a low-dimensional linear subspace for approximating the solution to many physics-based simulations. In this work we propose a novel approach for learning a data-driven quadratic manifold from high-dimensional data, then employing the quadratic manifold to derive efficient physics-based reduced-order models. The key ingredient of the approach is a polynomial mapping between high-dimensional states and a low-dimensional embedding. This mapping comprises two parts: a representation in a linear subspace (computed in this work using the proper orthogonal decomposition) and a quadratic component. The approach can be viewed as a form of data-driven closure modeling, since the quadratic component introduces directions into the approximation that lie in the orthogonal complement of the linear subspace, but without introducing any additional degrees of freedom to the low-dimensional representation. Combining the quadratic manifold approximation with the operator inference method for projection-based model reduction leads to a scalable non-intrusive approach for learning reduced-order models of dynamical systems. Applying the new approach to transport-dominated systems of partial differential equations illustrates the gains in efficiency that can be achieved over approximation in a linear subspace.

Rudy Geelen
Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin
rudy.geelen@austin.utexas.edu

PP1

PARTICLEDA.JL: A Real-Time Data Assimilation Software Platform

Digital twins of physical and human systems informed by real-time data, are becoming ubiquitous across weather forecasting, disaster preparedness, and urban planning, but researchers lack the tools to run these models effectively and efficiently, limiting progress. One of the current challenges is to assimilate observations in highly nonlinear dynamical systems, as the practical need is often to detect abrupt changes. The RADDISH (Real-time Advanced Data assimilation for Digital Simulation of numerical twins on HPC) group has developed a software platform to improve the use of real-time data in nonlinear system representations. Particle filtering data assimilation (DA) techniques have been implemented within an user-friendly software platform in Julia - ParticleDA.jl. To ensure the applicability of the developed platform in realistic scenarios, emphasis has been placed on numerical efficiency, scalability and optimisation for high performance computing frameworks. Furthermore, the platform has been developed to be forward model agnostic, ensuring that it is applicable to a wide range of modelling settings. In this presentation an interactive demo showcasing the platforms integration with a tsunami model and a numerical weather prediction model will be shown. Details on the steps involved in linking alternative numerical models will be discussed.

Daniel Giles
University College London

d.giles@ucl.ac.uk

Devaraj Gopinathan
Department of Statistical Science
University College London
d.gopinathan@ucl.ac.uk

Matthew Graham, Mosè Giordano, Tuomas Koskela, Alexandros Beskos, Serge Guillas
University College London
m.graham@ucl.ac.uk, m.giordano@ucl.ac.uk, t.koskela@ucl.ac.uk, a.beskos@ucl.ac.uk, s.guillas@ucl.ac.uk

PP1

Automated Data Versioning Using Statistical Machine Learning

In research settings, it is common to generate slightly modified versions of the same original dataset due to different preprocessing strategies or the selection of certain subsets of observations. Hence, there is a need for automated and objective data and metadata comparison (and versioning) strategies. We propose a dataset comparison strategy based on the parameters derived from Principal Component Analysis (PCA) models to trigger automated versioning mechanisms, integrated as part of new metadata standards. A PCA approach is used to map the relevant information of the datasets to be compared to a space of reduced dimensionality. The statistical features of the resulting PCA models, namely the correlation between pairs of homologous loading vectors, are then used as quantitative metrics to compute distances between data versions. This approach has been assessed under several scenarios where dataset comparison would be of benefit: imputing missing cell values, deleting records from an original data set, and applying non-linear transformations to the data. An ANOVA (analysis of variance) test on these parameters shows stability on the selected parameters when comparing the original dataset and modified versions with up to 50% of missing cells and up to a removal of 80% of rows. The sensitivity of the approach to different changes to the original dataset has been tested with two open data sets.

Luke A. Mcguinness, Alba González-Cebrián, Adriana E. Chis, Michael Bradford, Horacio González-Vélez
National College of Ireland
Luke.mcguinness@ncirl.ie, alba.gonzalez-cebrian@ncirl.ie, adriana.chis@ncirl.ie, michael.bradford@ncirl.ie, horacio.gonzalez-velez@ncirl.ie

PP1

A Deep Learning Approach to Probabilistic Forecasting of Weather

We discuss an approach to probabilistic forecasting based on two chained machine-learning steps: a *dimensional reduction* step that learns a reduction map of predictor information to a low-dimensional space in a manner designed to preserve information about forecast quantities; and a *density estimation* step that uses the probabilistic machine learning technique of normalizing flows to compute the joint probability density of reduced predictors and forecast quantities. This joint density is then renormalized to produce the conditional forecast distribution. In this method, probabilistic calibration testing plays the role of a regularization procedure, preventing overfitting in the second step, while effective dimensional reduction from the first step is the source of forecast sharpness. We verify the method using a 22-year 1-hour cadence time series of Weather Research and Forecasting (WRF) simulation data of surface wind on a grid.

Carlo Graziani
Argonne National Laboratory
cgraziani@anl.gov

Nick Rittler
University of California at San Diego
Department of Statistics
nrittler@ucsd.edu

Jiali Wang, Rao Kotamarthi
Argonne National Laboratory
jialiwang@anl.gov, vrkotamarthi@anl.gov

PP1

Multi-Dimensional Data Analysis of High-Performance Computing Metrics

I will present a data visualization technique created in Python and optimized for multi-dimensional data to explore the general relationships between power, temperature, time, and location of high-performance computing system components.

Kristen Lee Hallas
The University of Texas Rio Grande Valley, U.S.
kristen.hallas01@utrgv.edu

PP1

A Nonlinear Numerical Transient Compact Photocurrent Model

The dynamical behavior of excess charge carriers within semiconductor materials is often well-described by the Ambipolar Diffusion Equation (ADE), which can be derived from the Drift-Diffusion Equations by applying some minimal assumptions. However, the ADE is a nonlinear, time-dependent partial differential equation that is not immediately amenable to analytic solution methods. As a result, the development of traditional compact device models based on the ambipolar charge transport model generally requires physical approximations to render the ADE solvable in closed form. Violation of the assumptions underpinning these approximations (which may occur if the range of operating points of interest is large) can compromise the model accuracy. Alternatively, compact models may be obtained directly by applying Model Order Reduction (MOR) techniques to the discretized ADE, which avoids needing to apply any physical simplifications. In this work, we show that a projection-based MOR method combined with polynomialization of the discretized nonlinear ADE yields a robust and effective compact model for stimulated photocurrent in a pn-junction.

Joshua Hanson
University of Illinois at Urbana-Champaign
jmh4@illinois.edu

Charles Hembree
Sandia National Laboratories
cehembr@sandia.gov

PP1

Data Compression Techniques for Large-Scale

Memory-Bound Finite Element Applications

Many large-scale high-performance finite element applications are memory-bound for unstructured meshes due to the storage of various element-based quantities. Some of these quantities can be calculated on-demand, but this is often prohibitively expensive. In this presentation, we discuss an approach that utilizes data-compression for these element-based quantities to drastically reduce memory constraints while maintaining, and sometimes improving, efficiency for high-performance memory-bound finite element applications. Our approach is based on the observation that the Pareto principle, which postulates that the majority of the variance comes from a minority of members, holds for many unstructured meshes. By leveraging this observation, we construct data-compression algorithms to minimize the storage of redundant information, such as basis functions at quadrature points and local Jacobians. We demonstrate that this simple approach can lead to drastic memory compression ratios on unstructured meshes used in practical applications.

Graham Harper
Computational Mathematics Dept.
Sandia National Laboratories
gbharpe@sandia.gov

Tim Wildey
Optimization and Uncertainty Quantification Dept.
Sandia National Laboratories
tmwilde@sandia.gov

PP1

Non-Uniform Sampling Methods for Testing Link Prediction Methods on Networks

Link prediction problems are important for helping to recover missing data in networks and have applications in recommendation systems. Most state-of-the-art tests for link prediction methods have only considered uniformly missing data, e.g., removing edges uniformly from the original network. Uniform sampling of random edges is inherently efficient for constructing a representative subgraph. However, uniform sampling might lead to biased outcomes for link prediction because missing data in many real-world scenarios are probably not uniform. For example, there might be missing data centered around specific individuals or communities, so a link prediction scheme developed under uniform sampling may not achieve its optimal performance in another setting. In order to address this shortcoming, we consider a variety of different sampling methods for mimicking real missing data. We apply these sampling methods to over 500 real-world network data sets to generate tests for link prediction algorithms. These networks include social, biological, transportation, information, economic, and technology networks. We show that under different sampling conditions, various link prediction algorithms can yield different levels of accuracy on different types of networks.

Xie He
Dartmouth College
xie.he.gr@dartmouth.edu

Amir Ghasemian
Yale University
amir.ghasemian@yale.edu

Eun Lee
Pukyong National University
eunlee@pknu.ac.kr

Peter Mucha
Dartmouth College
peter.j.mucha@dartmouth.edu

PP1

Optimally Weighted PCA for High-Dimensional Heterogeneous-Quality Data

Principal component analysis (PCA) is a foundational data science technique, but how should PCA be performed in modern applications where data are often high-dimensional and have heterogeneous quality? This poster considers how to perform PCA for high-dimensional data when the noise is heterogeneous across samples, i.e., when some samples are noisier than others. Such heterogeneity naturally arises, e.g., when data are combined from diverse sources or sensors. A natural way to account for this heterogeneity is to give noisier blocks of samples less weight in PCA by using the leading eigenvectors of a weighted sample covariance matrix. We consider the problem of choosing weights to optimally recover underlying (i.e., latent) components. In general, one cannot know these optimal weights since they depend on the underlying components we seek to estimate. However, we show that under some natural statistical assumptions the optimal weights converge to a simple function of the signal and noise variances for high-dimensional data. Surprisingly, the optimal weights are not the inverse noise variance weights commonly used in practice. We demonstrate the theoretical results through numerical simulations and comparisons with existing weighting schemes. We also explain how estimated signal and noise variances can be used when the true variances are unknown, and we illustrate the optimal weights on real data from astronomy. For details see: https://arxiv.org/abs/1810.12862

David Hong
University of Pennsylvania, U.S.
dahong67@wharton.upenn.edu

Fan Yang
University of Pennsylvania
fyang75@wharton.upenn.edu

Jeffrey Fessler
University of Michigan
fessler@umich.edu

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

PP1

On Estimating the Distribution of the Maximum Likelihood Estimator in Exponential Families

Maximum likelihood (ML) estimation in exponential families, especially for data with complex dependencies such as point patterns, often suffers from an intractable likelihood. As a consequence, we are concerned with theoretical and numerical obstacles. Asymptotic normality of the ML estimator is generally not proven, in fact one may expect a nonstandard asymptotic behavior. And although an ML estimate can be obtained by using Markov chain Monte Carlo methods [Geyer, Ch. J., & E. A. Thompson.

Constrained Monte Carlo maximum likelihood for dependent data. J R Stat Soc Series B, 54.3 (1992): 657-699.], the situation is still unsatisfactory due to the computational intensity. Thus, performing a parametric bootstrap to evaluate the uncertainty of an ML estimate may be prohibitively expensive. We present a simple estimator for the distribution of the ML estimator in exponential families. This estimator uses computationally efficient alternative estimation methods based on pseudo-likelihoods [e.g. Baddeley, A. et al. Logistic regression for spatial Gibbs point processes. Biometrika 101.2 (2014): 377-392.] or variational equations [e.g. Baddeley, A. & D. Dereudre. Variational estimators for the parameters of Gibbs point process models. Bernoulli 19.3 (2013): 905-930] and can be easily implemented by using standard software for linear regression. A simulation study illustrates the proposed method for the ML estimator in a Lennard-Jones Gibbs point process model.

Henning Höllwarth
TU BA Freiberg
henning.hoellwarth@math.tu-freiberg.de

PP1

On Tensor-Based Training of Neural Networks

Neural networks have been successfully used in many applications. However, there are also cases of failure, for instance, in image classification tasks, when a hardly perceptible perturbation of an image leads to an incorrect classification. In this work, by resorting to the continuous 'model' of a shallow neural network, we might shed some light on this issue of the success and failure by observing the ill-posedness of the underlying continuous problem. Moreover, we present a novel approach to the training of neural networks that is based on a suitable approximate solution of a Fredholm integral equation of the first kind. Here, we concentrate on a least-squares collocation method (generalised Galerkin method), which is combined with functional tensor train format and alternating ridge regression. This allows us to reconstruct a shallow neural network with the help of appropriate quadrature rules. Efficiency and reliability of this approach is confirmed by numerical experiments.

Patrick Gelß, Aizhan Issagali
FU Berlin
p.gelss@fu-berlin.de, aizhan.issagali@fu-berlin.de

Ralf Kornhuber
Freie Universität Berlin, Germany
ralf.kornhuber@fu-berlin.de

PP1

Statistical Modelling of Communication Through Violence During Election Periods in Africa

Do elections draw political violence? Elections are sensitive events with meaningful political and economic implications for many. It represents an opportunity for the redistribution of power and a shift in priorities for a nation. As such, it invites extreme responses that sometimes can be translated to political violence. Examining a longitudinal sample of African countries election cycles between the years 1997-2020, we assess the association between elections and political violence. We utilize the Armed Conflict Location and Event Data and fit negative binomial regression models and logistic regression models to examine the links between different types of political violence before and after an election and in comparison, to the general levels of political violence during the non-election period. Our results indicate that an election develops a politically critical period which is on average more violent than any other period. Also, the statistical results show that elections attract low levels of violence, namely riots and violence against civilians, while reducing the likelihood of extreme levels of violence, such as battles and the use of explosives. Accordingly, not all violence event types have equal incidence around the election day. These findings add to the discussion on electoral violence and violent democracy, offering a unified theory to different manifestations of political violence around elections.

Sithma Jayawardena, Dimitri Volchenkov, Ori Swed
Texas Tech University
sithma.jayawardena@ttu.edu, dimitri.volchenkov@ttu.edu, ori.swed@ttu.edu

PP1

Impact Assessment of Expanded Access to HIV Treatment in Lesotho Between 2014 and 2016 Using Integrated Data from Multiple Surveys and Scraped Data

In 2016, Lesotho became the first sub-Saharan African country to adopt a Treat All policy, with universal treatment of all individuals with HIV regardless of CD4 level. As a country with large HIV prevalence in the population, we are interested in understanding the impacts of the adoption of the program. We used data from two population-based surveys to assess the impact of expanded access to HIV treatment in Lesotho by assessing the national prevalence of HIV viral suppression before and after policy implementation. The two surveys, the Lesotho Demographic Health Survey (DHS) in 2014 and Population-based HIV Impact Assessment (PHIA) in 2016, will be combined to evaluate the distribution of HIV viral load in HIV positive patients. Calibration data to convert the two measurements will be scraped from a published scatterplot of viral load between the two methods in the same group of subjects. Direct comparison of HIV viral load between DHS and PHIA is challenging, because DHS used dried blood method (DBS), but PHIA used plasma method for measuring viral load. Plasma is considered a gold standard measure, and DBS tends to involve error in measurement whilst being a cheaper and less intensive method. Our approach uses a multiple imputation method with non-differential measurement error to estimate HIV viral load suppression across both surveys, where the difference in measurement between the two methods is considered as error.

Kate Johnson
University of California Davis
kj864ltj@gmail.com

McCord Murray
University of Massachusetts, Dartmouth, U.S.
dmurray3@umassd.edu

PP1

Uncertainty Quantification and Optimization with Bayesian Hybrid Models for Chemical Sciences, Engineering, and Beyond

In chemical engineering, multiscale computer models are used to guide discovery and systems optimization. The complex nature of multiscale modeling, however, often requires approximations to reduce computational cost

thereby introducing uncertainty. In order to harness the full predictive potential of these models, comprehensive quantification of uncertainty is key. At the intersection of glass and black-box models are Kennedy-OHagan (KOH) models [M.C. Kennedy and A. OHagan, Bayesian calibration of computer models] which augment full-physics formulations with a surrogate model to quantify epistemic uncertainty. While there exists a corpus of literature utilizing KOH models, it remains unclear how to best perform optimization under uncertainty for decision-making. Extending the KOH framework, we develop Bayesian hybrid models of the form:

$$y_i = \eta(x_i;\theta) + \delta(x_i;\phi) + \epsilon_i$$

Here, observation y_i is modeled with three components: a white-box model $\eta(\cdot,\cdot)$ with global parameters θ and inputs x_i for experiment i, a stochastic discrepancy function $\delta(\cdot)$, and random measurement error $\epsilon \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{I}\boldsymbol{\sigma}^{\mathbf{2}})$. Using a reaction engineering case study, we show how this workflow learns unknown reaction kinetics and accurately predicts the duration and temperature of a reaction, thus highlighting its potential to accelerate decision science.

Kyla D. Jones, Elvis Eugene, Alexander Dowling
University of Notre Dame
kjones29@nd.edu, eeugene@nd.edu, adowling@nd.edu

PP1

Statistical Approach to Study Various Challenges in Data Science

The role of data science is of great importance to extract useful information for decision making in science, engineering, business and society. A very large collection of data is generated every day due to modern information systems, artificial intelligence, digital technologies etc. Statistical analysis of such data is needed and is the area of quite concern. The present paper, mainly aims to explore the new challenges, to provide scientific tools, and to study the impact of statistics with the help of appropriate examples and diagrams. It provides a platform to explore statistical techniques and tools at numerous stages. It also opens a new horizon for researchers to develop ideas and tackle scientific challenges emerged from data storage, capturing, retrieval, analysis, optimization and visualization.

Rajinder Kaur
Trinity College, Jalandhar
raj.kahlon2003@gmail.com

PP1

Finding Higher Order Interactions Using Local CorEx

In applications such as financial markets, social networks, and gene expression data, the variables often interact in complex ways. Yet accurately characterizing pairwise variable interactions can be a difficult task, let alone efficiently characterizing complex higher-order interactions, which is an unsolved problem. This difficulty is exacerbated when variable interactions change across the data. For example, gene interactions in single-cell RNA-sequencing (scRNA-seq) data will typically differ from one cell type to another. To solve these problems, we propose a new method called Local Correlation Explanation (CorEx). Local CorEx captures higher-order variable interactions at a local scale by first clustering data points based on their proximity on the data manifold. We then use a multivariate version of mutual information, called the total correlation, to construct a latent factor representation of the data within each cluster to learn the local variable interactions. We compare Local CorEx with other methods and show that it performs favorably on both synthetic and real data including images and scRNA-seq data

Thomas J. Kerby
Utah State University
Department of Mathematics and Statistics
thomas.kerby@usu.edu

Kevin Moon
Department of Mathematics and Statistics
Utah State University
kevin.moon@usu.edu

PP1

A Stochastic Inverse Solution to Thermo-Fluid Problems

Inverse problems find their applications in various thermo-fluid systems. Finding a solution to inverse problems has been challenging due to their ill-posed nature. A solution to inverse problems is a promising area of research that could help find various unknown model parameters that are otherwise difficult to obtain. In the past few decades, different deterministic and stochastic methods have been developed to find the solution to inverse problems. The present work aims to develop a computational framework using fast Bayesian inference, which leads to forward uncertainty propagation in various thermo-fluid models and solves the corresponding inverse problems. This work addresses inverse problems in thermo-fluid systems with high dimensional parameter spaces. The framework leverages the polynomial chaos expansions (PCEs) to generate a computationally efficient and statistically equivalent surrogate model of the computationally expensive forward model and dimensionality reduction based on Karhunen-Loeve (K-L) expansion. The applicability of the framework is established by validating different model problems in thermo-fluid systems, such as in heat and mass transfer and subsurface flows in petroleum engineering. The combination of PC-based surrogate modeling and K-L-based dimensionality reduction leads to over multiple fold speed up of a numerical solution of inverse problems.

Sufia Khatoon
Indian Institute of Technology, Delhi
sufia3304@gmail.com

Supreet Bahga, Jyoti Phirani
Indian Institute of Technology Delhi
bahga@mech.iitd.ac.in, jphirani@chemical.iitd.ac.in

PP1

Efficient Approximations of the Fisher Matrix in Neural Networks Using Kronecker Product Singular Value Decomposition

Several studies have shown the ability of natural gradient descent to minimize the objective function more efficiently than ordinary gradient descent based methods. However, the bottleneck of this approach for training deep neural networks lies in the prohibitive cost of solving a large dense linear system corresponding to the Fisher Information Matrix (FIM) at each iteration. This has motivated various approximations of either the exact FIM or the empirical one. The most sophisticated of these is KFAC [Martens & Grosse, Optimizing Neural Networks with Kronecker-

factored Approximate Curvature, PMLR 37: 2408–2417, 2015], which involves a Kronecker-factored block diagonal approximation of the FIM. With only a slight additional cost, a few improvements of KFAC from the standpoint of accuracy are proposed. The common feature of the four novel methods is that they rely on a direct minimization problem, the solution of which can be computed via the Kronecker product singular value decomposition technique. Experimental results on the three standard deep auto-encoder benchmarks showed that they provide more accurate approximations to the FIM. Furthermore, they outperform KFAC and state-of-the-art first-order methods in terms of optimization speed.

Abdoulaye Koroko
IFP Energies nouvelles
Université Paris-Saclay
abdoulaye.koroko@ifpen.fr

Ani Anciaux-Sedrakian
IFP Energies nouvelles
ani.anciaux-sedrakian@ifpen.fr

Ibtihel Ben Gharbia
IFP Energies nouvelles, France
ibtihel.ben-gharbia@ifpen.fr

Valérie Garès
Univ Rennes, INSA, CNRS, IRMAR
valerie.gares@insa-rennes.fr

Mounir Haddou
Institut National des Sciences Appliquées de Rennes
France
mounir.haddou@insa-rennes.fr

Quang Huy Tran
IFP Energies nouvelles
quang-huy.tran@ifpen.fr

PP1

New Evaluation Framework for Data Quality

We live in a revolutionary era of big data: almost every e-commerce company leverages data, builds data science models, collects and produces large volumes of digitized information. Data quality goes hand in hand with big data. The field of statistics and its techniques are used to validate model performance and data quality. RMSE, p-values and ROC are usual candidates to evaluate model performance. In this paper, we developed a new framework of validation methods, thereby bridging data applications and combinatorial techniques closer. Leveraging a simple but very power combinatorial foundation, we explicitly computed number of sequences of 10,8,6,4-digits, starting from 0000 and up to 9999 in a case of 4-digit numbers, in which the sum of the first half of the digits is equal to the sum of the second half. With explicit formulas, we can reduce noise/bias in data. Using the formulas, the number of "lucky" sequences- for 4-digit sequence 670, for six, 55252, for eight, 4816030, and for ten, 432457640. Thus, we have an explicit technique to assess how probable it is to see a certain numerical sequence as an input/output in a model. For example, $670/10^4$ shows how probable it is to account for sequences of 4 digits where the sum of the first 2 equals the last 2. If there are 10^4 data points, and there are significantly more or less than 670 "lucky" sequences, then the data can be categorized as biased. The value of the formulas is the fast evaluation of data quality.

Ellen J. Kozlov
Calabasas High School
e.kozlov.05@gmail.com

PP1

Use the Kolmogorov Superposition Theorem to Approximate High Dimensional Data and Functions

We explain that we can use the Kolmogorov superposition theorem to overcome the curse of dimensionali1y when approximating high dimensional functions and data. Numerical experimental results in 2D and 3D will be demonstrated that we are able to use a few hundred basis functions to approximate very complicated functions reasonably well.

Ming-Jun Lai
University of Georgia
Department of Mathematics
mjlai@math.uga.edu

PP1

Total Variation Denoising with Slack Variables

Total variation (TV) regularization is a well-known method to remove noise while simultaneously preserving the sharp edges. The traditional L1-regularized optimization problem is particularly effective when used to remove Gaussian noise from a distinct image, but it performs poorly when given a noisy image with a smooth background and sharp well-localized transitions (calcifications). Therefore, we propose a new method to address this issue by incorporating the calcifications into the traditional 1D TV Denoising model to attempt to improve its output for images with calcifications using slack variables which may be to solve this constraint.

Nhi Yen Le
Rice University, U.S.
nyl1@rice.edu

PP1

Physics Informed Deep Learning: Application to Well Opening and Closing Events

Numerical simulations provide the solution to the multiphase flow equations that model the behavior of the CO2 injection site. They are a valuable tool to decide either or not to exploit a potential carbon storage site and to monitor the operations. However, numerical simulations of fluid flow in porous media are computationally demanding: it can take up to several hours on a HPC cluster in order to simulate one injection scenario for a large CO2 reservoir. This becomes a limiting issue when performing a large number of simulations. More specifically, well events cause important numerical difficulties due to their instant impact on the system. This forces a drastic reduction of the time step size to be able to solve the non-linear system of equations resulting from the discretization of the continuous mathematical model. However, these specific well events are rather alike across space and time. Recent interest in machine learning applied to the prediction of physical processes has fueled the development of "Physics Informed Deep Learning" (PIDL), where machine learning models either replace or complement traditional numerical algorithms while preserving the inherent constraints from

the physical model. Therefore, we propose to adapt recent advances in PIDL so as to alleviate the impact of well events in the numerical simulation of multi-phase flow in porous media. More precisely, we complement the implicit numerical solver by learning Newton's method initialization.

Antoine Lechevallier
Laboratoire Jacques-Louis Lions
IFP Energies Nouvelles
lecheval7u@gmail.com

Frédéric Nataf
CNRS and UPMC
nataf@ann.jussieu.fr

Thibault Faney
IFPEN, Department of Applied Mathematics
thibault.faney@ifpen.fr

Sylvain Desroziers
Michelin
sylvain.desroziers@michelin.com

Eric Flauraud
IFP Energies Nouvelles
eric.flauraud@ifpen.fr

PP1

Advances in Solving HJB Equations Arising in Stochastic Optimal Control

We present a neural network approach for approximately solving high-dimensional stochastic control problems. Our network design and the training problem leverage insights from optimal control theory. We approximate the value function of the control problem using a neural network and use the stochastic Pontryagin maximum principle and Dynamic Programming principle to express the optimal control (and therefore the sampling) in terms of the value function. Our training loss consists of a weighted sum of the objective functional of the control problem and penalty terms that enforce the Hamilton Jacobi Bellman equations along the sampled trajectories. As a result, we can obtain the value function in the regions of the state space traveled by optimal trajectories to avoid the curse of Dimensionality, and therefore solving the problem in a semi-global fashion. Importantly, training is unsupervised in that it does not require solutions of the control problem. Our approach reduces to the method of characteristics as the system dynamics become deterministic. In our numerical experiments, we compare our method to existing solvers for a more general class of semilinear PDEs. Using a two-dimensional toy problem, we demonstrate the importance of the stochastic PMP to inform the sampling. For a 100-dimensional benchmark problem, we demonstrate that approach improves accuracy and time-to-solution.

Deepanshu Verma, Xingjian Li
Emory University
dverma4@emory.edu, xingjian.li@emory.edu

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

PP1

Least-Squares Neural Network for Inverse Medium Problems

In this talk, we present a novel neural network approach for the reconstruction of the medium coefficients in several different model settings by taking into account the observed data and physical modeling concepts. Our approach combines the concepts of convolutional neural networks in deep learning and physics-based training arising from differential equations to retrieve high-resolution reconstructions. We demonstrate numerical results of our approach on acoustic wave scattering.

Ying Liang
The Chinese University of Hong Kong
liang402@purdue.edu

PP1

Optimal Recovery from Inaccurate Data in Hilbert Spaces: Regularize, But What of the Parameter?

In Optimal Recovery, the task of learning a function from observational data is tackled deterministically by adopting a worst-case perspective tied to an explicit model assumption made on the functions to be learned. Working in the framework of Hilbert spaces, this article considers a model assumption based on approximability. It also incorporates observational inaccuracies modeled via additive errors bounded in ℓ_2. Earlier works have demonstrated that regularization provide algorithms that are optimal in this situation, but did not fully identify the desired hyperparameter. This article fills the gap in both a local scenario and a global scenario. In the local scenario, which amounts to the determination of Chebyshev centers, the semidefinite recipe of Beck and Eldar (legitimately valid in the complex setting only) is complemented by a more direct approach, with the proviso that the observational functionals have orthonormal representers. In the said approach, the desired parameter is the solution to an equation that can be resolved via standard methods. In the global scenario, where linear algorithms rule, the parameter elusive in the works of Micchelli et al. is found as the byproduct of a semidefinite program. Additionally and quite surprisingly, in case of observational functionals with orthonormal representers, it is established that any regularization parameter is optimal.

Chunyang Liao, Simon Foucart
Texas A&M University
liaochunyang@tamu.edu, foucart@tamu.edu

PP1

Towards Closed-Loop Deep Brain Stimulation via Optimal Control

Recent advances in machine learning and optimal control have opened new avenues to tackle learning problems in computational neuroscience. In this work, we consider deep brain stimulation (DBS), a highly efficacious treatment for various neurological disorders including medically refractory Parkinsons disease and epilepsy. While conventional open-loop DBS remains an effective treatment strategy, its success is limited by trial and error parameter selection by clinicians and its inability to adapt and personalize to a patients evolving clinical state. Closed-loop neurostimula-

tion has the potential to mitigate these issues and maximize the therapeutic benefit of stimulation while reducing stimulation side effects. Focusing on the Hodgkin-Huxley neuronal model, we formulate the problem of finding an optimal neurostimulation strategy as a control problem. We state and prove the optimality conditions of the Hamilton-Jacobi-Bellman equations and derive the value function from which an optimal control can be recovered. We lay down the mathematical foundation for designing closed-loop treatment strategies and empirically demonstrate the utility of this approach.

Malvern Madondo
Emory University
malvern.madondo@emory.edu

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

Nicholas Au Yong
Emory University
nicholas.au.yong@emory.edu

PP1

Change Detection in High-Dimensional Covariance Using Random Matrix Theory

Time series data often violates stationarity assumptions that are crucial to the success of most analysis methods. In this paper, we introduce a novel test of stationarity based on random matrix theory. The test uses the spectral distribution of the sample covariance matrix to detect deviations from second-order stationarity under the single change point model: the covariance matrix changes once over the observation interval. Given a sample of size T of an N-dimensional random process, we consider the presence of one change point $1 \leq r \leq T$ such that the observations indexed $1, , r$ are i.i.d. and have a different covariance matrix than the observations indexed $r+1, , T$, also assumed to be i.i.d. Using random matrix theory and free probability theory, we develop a statistic that is asymptotically zero for all indices in the proportional growth limit $N/T \to \gamma \in (0, \infty)$ and $r/T \to \xi \in (0, 1]$ as $N, r, T \to \infty$. Furthermore, if a change is present, we show that the statistic is asymptotically maximized at r with high probability. This yields a procedure by which we can detect a change in covariance by thresholding and subsequently estimate the change point by maximizing the statistic over the observed indices. Due to the universality of random matrix theory, these results are shown under mild regularity conditions. Finally, we discuss convergence rates in probability and the statistical performance of the associated test.

Robert Malinas
University of Michigan
rmalinas@umich.edu

Benjamin Robinson
Air Force Research Laboratory
benjamin.robinson.8@afresearchlab.com

Alfred Hero
University of Michigan
hero@umich.edu

PP1

Dictionary Learning for Inverse Heat Transfer Problems

Let us consider the estimation of hidden corrosion profiles from infrared thermographic data, a nonlinear inverse problem. A classic method would be to define a quadratic cost function and to use some regularized nonlinear optimization method. This implies to iteratively linearize the problem around the current estimate of parameters, that at the beginning may be very far from the real values, and this brings usually to a local minimum. If we interpret the Hessian matrix, that arise in a typical inexact/quasi-Newton step, as a column-wise dictionary, we realize immediately that, at early steps, this is really poor at describing the prediction error. On the other hand, it is possible to define a dictionary even capable of representing exactly the prediction error that could arise in an arbitrary real instance, i.e. to solve the inverse problem in a single step, but this implies to build a very huge matrix, e.g. 10^{50} columns, with an intractable cost. In this talk we propose and analyze intermediate dictionaries that give a better trade-off between the computational cost required to build the dictionary and the efficiency of dictionary predictions performed by sparse recovery algorithms. In particular, we will show an a-priori dictionary initialized from simulations, then optimized with techniques of dictionary learning. We will present some results with an ad-hoc sparse recovery algorithm for NNLS and also clues to generalize at other heat inverse problems.

Fabio Marcuzzi, Sabrina Carpanese
University of Padova
marcuzzi@math.unipd.it, marcuzzi@math.unipd.it

PP1

Multifidelity Covariance Estimation Via Regression on the Manifold of Symmetric Positive Definite Matrices

In this poster we introduce a multifidelity covariance estimator defined as the solution to a regression problem on tangent spaces to product manifolds of symmetric positive definite (SPD) matrices. Given a set of high- and low-fidelity sample covariance matrices, which we view as a sample of a product-manifold-valued random variable, we estimate the underlying true covariance matrices by minimizing a notion of squared Mahalanobis distance, defined via intrinsic statistics on the SPD manifold, between the data and a model for its variation about its mean. The resulting estimates are guaranteeably positive definite and the intrinsic Mahalanobis distance which they minimize has desirable properties, including tangent-space agnosticism and affine-invariance. Mahalanobis distance minimization can be carried out using unconstrained gradient-descent methods when a reparametrization in terms of SPD matrix square roots is employed, and we introduce a new Julia package providing a convenient API for solving these multifidelity covariance regression problems. Using its machinery, we demonstrate that our estimator can provide significant reductions in MSE over single-fidelity covariance estimators in forward uncertainty quantification problems.

Aimee Maurais
Massachusetts Institute of Technology

maurais@mit.edu

PP1

A Simplicial Pooling Layer

For deep learning problems on graph-structured data, pooling layers are important for down sampling, reduced computational cost, and to minimize overfitting. We consider data structured as simplicial complexes, which are generalizations of graphs that include of higher-dimensional simplices beyond vertices and edges; this structure allows for greater flexibility in modeling higher-order relationships. DiffPool (Ying et al, 2018) use a hierarchical clustering scheme for graph coarsening, where a cluster assignment matrix is learned at each GNN layer and used to coarsen the graph for input to the subsequent layer. We propose a simplicial pooling layer built upon DiffPool, which generates hierarchical representations of simplicial complexes. This is a generalized coarsening method for simplicial complexes that collapses information in a learned fashion. The simplicial pooling method builds on learned vertex cluster assignments and extends to coarsening of higher dimensional simplices in a deterministic fashion. While in practice, the pooling operations are computed via a series of matrix operations, its topological motivation is based on unions of stars of simplices.

Sarah Mcguire, Elizabeth Munch, Matthew Hirn
Michigan State University
mcgui176@msu.edu, muncheli@msu.edu, mhirn@msu.edu

PP1

Data Analysis of Flow Cytometry Data: A Probabilistic Approach

Flow cytometry (FC) is a technique for quantifying biomarker expression levels of cells, which is useful for diagnosing and managing diseases such as cancer. FC experiments produce on the order of 10^6 events, corresponding to measurements of each cells size, homogeneity, or biomarker expression level. To distinguish between populations (e.g., cancer vs healthy cells) in a sample, a user manually gates or clusters data, which is often represented in terms of histograms or scatter plots. While interpretation of data is critical to the measurement process, gating is not only time-consuming and subjective but suffers from difficult to quantify uncertainties. We developed a methodology that (1) identifies a population described by its probability density function (PDF) and (2) uses constrained bilevel optimization to subtract off one populations PDF to estimate the unknown populations distribution. Though the construction of the PDFs includes some uncertainty, we can quantify this uncertainty directly from our algorithm. We apply our algorithm to testing data obtained from stained cells to differentiate populations. Our preliminary results suggest our method can efficiently separate populations comparable to that of traditional gating methods while removing initial subjectivity in the data analysis.

Danielle J. Middlebrooks, Megan Catterton
National Institute of Standards and Technology
danielle.middlebrooks@nist.gov,
megan.catterton@nist.gov

Anthony J. Kearsley
Applied and Computational Mathematics Division
National Institute for Standards and Technology
anthony.kearsley@nist.gov

Geoffrey McFadden, Gregory Cooksey
National Institute of Standards and Technology
geoffrey.mcfadden@nist.gov, gregory.cooksey@nist.gov

Paul Patrone
NIST Gaithersburg, Information Technology Lab
Applied and Computational Mathematics Division
paul.patrone@nist.gov

PP1

Efficient Topological Features for Machine Learning

Topological Data Analysis is an emerging area rooted in theories from Algebraic Topology, which enables researchers to extract discriminating geometric and topological features from data. We give an overview of some of the popular methods of extracting features from point-cloud data, which first requires one to construct a 1-parameter family of spaces on the data using the geometry of the cloud. We demonstrate their benefits and shortcomings and introduce a new, more efficient construction that we name the Delaunay-Rips Filtration. Aided by intuitive examples, we also provide an empirical run-time comparison of the two existing methods with our new algorithm on the computation of the persistence diagrams of some synthetic data sets. Finally, we demonstrate the utility of our algorithm in a Machine Learning classification problem on some noisy, synthetic datasets. We conclude that our algorithm not only is computationally faster, but its results are also comparable to those of using a standard, slower algorithm.

Amish Mishra, Francis C. Motta
Florida Atlantic University
amishra2019@fau.edu, fmotta@fau.edu

PP1

Diffusion Transport Alignment

The integration of multimodal data presents a challenge in cases when the study of a given phenomena by different instruments or conditions generates distinct but related domains. In this paper, we propose Diffusion Transport Alignment (DTA) a semi-supervised manifold alignment method that exploits prior correspondence knowledge between distinct data views. DTA finds a bijection between two domains, which by assumption, share a similar geometrical structure coming from the same underlying data generating process. We empirically demonstrate the effectiveness of our method to integrate multimodal data, as well as how it can improve the performance of machine learning tasks, otherwise less effective when only one of the domains is considered.

Kevin Moon, Andres Duque
Department of Mathematics and Statistics
Utah State University
kevin.moon@usu.edu, andres.duque@usu.edu

Guy Wolf
Department of Mathematics and Statistics
University of Montreal
guywolf@umontreal.ca

PP1

Modern Tensor Factorizations and Applications to

Longitudinal 'Omics Data Analysis

Many real-world data are inherently of multi-way structure. However, multidimensional data is often processed as two-dimensional arrays (matrices), thus, ignoring the inherent higher-order structure therein. Arguably, matrization of higher-order data is such a common practice due to the ubiquitousness and strong theoretical foundations of matrix algebra. Recently, developments in multilinear algebra have established an Eckart-Young like best rank approximation for the tensor tubal singular value decomposition (tSVDM), providing theoretical justification for the superiority of tensor-based approximations over matrization for the first time. In this work, we utilize the tSVDM to construct a principal component analysis (PCA) analog for 3rd order tensors which we refer to as the M-product based Tensor Component Analysis (TCAM). We derive optimality for the TCAM, namely, the maximization of variance and distortion minimization for the TCAM embedding. These theoretical guarantees put TCAM as a promising utility for multi-way data analysis tasks. We explore this utility in the context of analyzing high dimensional so-called "omics' data, which is collected in longitudinal biological experiments.

Uria Mor
Tel Aviv University
Weizmann Institute of Science
uriamor@mail.tau.ac.il

PP1

Scale Dependencies and Self-Similarity Through Wavelet Scattering Covariance

We introduce a scattering covariance which provides non-Gaussian models of time series having stationary increments. A complex wavelet transform computes signal variations at each scale. Dependencies across scales are captured by the joint covariance across time and scales of complex wavelet coefficients and their modulus. This covariance is nearly diagonalized by a second wavelet transform, which defines the scattering covariance. We show that this set of moments characterizes a wide range of non-Gaussian properties of multiscale processes. This is analyzed over fractional Brownian motions, Poisson, multifractal random walks and Hawkes processes. We prove that self-similar processes have a scattering covariance which is scale invariant. This property can be estimated numerically and defines a class of wide-sense self-similar processes. We build maximum entropy models conditioned by scattering covariance coefficients, and generate new time-series with a microcanonical sampling algorithm. Applications are shown over financial and turbulence time-series.

Rudy Morel
École Normale Supérieure, Center for Data Science
rudy.morel@ens.fr

Gaspar Rochette, Roberto Leonarduzzi
École Normale Supérieure
Center for Data Science
gaspar.rochette@ens.fr,
roberto.fabio.leonarduzzi@gmail.com

Jean-Philippe Bouchaud
Capital Fund Management
jean-philippe.bouchaud@cfm.fr

Stéphane Mallat
Collège de France
École Normale Supérieure
stephane.mallat@ens.fr

PP1

A Framework to Characterize Surrogate Models in Scientific Machine Learning

We present a novel and unifying framework of characterizing surrogate models for the emerging field of scientific machine learning (SciML). This framework rests upon the relationships between robustness, accuracy, interpretability, scalability, and efficiency (RaISE). We apply this framework to methods of both function and operator approximation to create a comprehensive and intuitively accessible catalogue.

Zachary Morrow
Sandia National Laboratories
zbmorro@sandia.gov

PP1

The SoCal Data Science Program

According to LinkedIns 2017 U.S. Emerging Jobs Report, data scientists rank among the top emerging jobs and California is expected to be a hub. It is essential that data science courses for undergraduates, particularly those with no or limited statistical or programming background, are developed and accessible for all scientific disciplines. It has been shown that creating a social community to support students can improve retention and mitigate some challenges faced by underrepresented minority students in STEM. The aim of this poster is to showcase the SoCal Data Science Program from the perspective of a data science fellow, a second-year math major at California State University Fullerton (CSUF). This program is a novel collaboration between CSUF, University of California Irvine, and Cypress College that is composed of students from all STEM majors who receive training in various aspects of data science and statistical software. Students will come together to participate in hands-on applications alongside data scientists and will present their work at the SoCal Data Science Symposium. The SoCal Data Science Fellows will gain the experience necessary for successful advancement into data science careers in related fields that interest them. As these students continue their undergraduate degree, they will serve as mentors to incoming data science fellows and foster a community for future data scientists.

Emily Murphy, Jessica Jaynes, Sam Behseta
California State University, Fullerton, U.S.
eamvh@csu.fullerton.edu, jjaynes@fullerton.edu, sbehseta@fullerton.edu

PP1

Markov Chain Generative-Adversarial Neural Networks for Solving Bayesian Inverse Problems

The Bayesian approach for solving inverse problems in scientific computing is very popular as it provides an estimate as well as uncertainty of the solution in shape of a posterior distribution. However, the computational costs associated with high-dimensional problems and expensive forward maps often make it infeasible to solve within a satisfying timeframe. Furthermore, choosing an appropriate prior distribution is crucial while not always an easy choice. To alleviate these problems, we present a

new approach referred to as Markov Chain Generative Adversarial Networks (MCGANs). Generative Adversarial Networks (GANs) are a suitable framework to aid in the solution of Bayesian inference problems, as they are designed to generate samples from complicated high-dimensional distributions. By training a GAN to sample from a low-dimensional latent space and then embedding it in a Markov Chain Monte Carlo method, we can efficiently sample from the posterior, by replacing both the high-dimensional prior and the expensive forward map. It can be shown that the proposed methodology converges to the true posterior in the Wasserstein-1 distance and that sampling in the latent space is equivalent to sampling in the high-dimensional space in a weak sense. The method is showcased on various nonlinear, high-dimensional, and dynamic problems such as leak detection in water transport pipelines.

Nikolaj T. Mücke
Centrum Wiskunde & Informatica (CWI) and Utrecht University
nikolaj.mucke@cwi.nl

PP1

Recent Advances for Efficiently Computing Conformal Prediction Sets

If you predict a label y of a new object with $\hat{y}$, how confident are you that $y = \hat{y}$? The conformal prediction method provides an elegant framework for answering such a question by establishing a confidence set for an unobserved response of a feature vector based on previous similar observations of responses and features. This is performed without assumptions about the distribution of the data. While providing strong coverage guarantees, computing conformal prediction sets requires adjusting a model to an augmented dataset considering all possible values that the unobserved response can take, and proceeding to select the most likely ones. For a regression problem where y is a continuous variable, it typically requires an infinite number of model fits; which is usually infeasible. In general, the computation of such distribution-free confidence sets is still considered an open problem. However, remarkable advances have been obtained recently by introducing more or less weak assumptions on the prediction model in order to have feasible calculations. In this lecture, I will present an overview of recent successes, highlighting their strengths and weaknesses in terms of complexity and practical performance. Finally, we will propose a discussion on the state of the art of the computational problems still open and future research to obtain software able to compute these confidence sets, at very large scale.

Eugene Ndiaye
Telecom-ParisTech, CNRS LTCI
Université Paris-Saclay
endiaye3@gatech.edu

PP1

An Inverse Time-Dependent Source Problem with Neumann Boundary Conditions

We consider an inverse time-dependent source problem for a parabolic partial differential equation with Neumann boundary conditions and subject to an integral constraint in a domain of $\mathbb{R}^n$, $n \geq 1$. We show well-posedness and regularity of solutions in Hlder spaces. The proof of the existence and uniqueness of solutions yields an algorithm that we employed to approximate solutions of the inverse problem using a finite element discretization in space and the backward Euler scheme in time. Due to instability in the reconstruction, Tikhonov regularization is applied. The errors resulting from our experiments show that the proposed scheme is an accurate approach to approximate solutions of this inverse problem.

Sedar Ngoma
SUNY Geneseo
ngoma@geneseo.edu

PP1

Actual Events vs. Actual Reporting: Modeling Firm Performance under Environmental Uncertainty Using Machine Learning

Not all companies respond the same to natural disaster events. This study investigates two ways that natural disasters affect firm performance: actual events vs. actual reporting. In this study, I consider the billion-dollar weather and climate disasters in the United States as the actual events and the number of words related to natural disasters in the Management Discussion and Analysis (MD&A) section in Form 10-Ks filing by the U.S. public companies as the actual reporting. This study also aims at comparing the performances of classification and regression trees (CART) and neural networks with the benchmark model-based linear regression model in predicting the performance of the U.S. public companies under environmental uncertainty. The results show that both actual events and actual reporting of natural disasters in year t negatively affects return on assets (ROA) in year $t+1$. Also, the actual natural disasters in year t negatively affects sales growth in year $t + 1$. Even though environmental uncertainty variables have some effects on firm performance, they are much less important than the traditional financial statement variables in predicting firm performance. Comparing among CART, neural networks, and linear regression models, I find that CART and neural networks outperform linear regression models in predicting firm performance. This result is robust to any given firm performance criteria, split ratios, and prediction errors.

Minh Nguyen
University of Hawaii at Manoa
duyminh@hawaii.edu

PP1

Mixed Models and Its Application to the Analysis of Covid-19 Data

Despite efforts by various stakeholders involved in COVID-19 management, the virus remains a global pandemic. The virus has been shown to be highly contagious in recent cases, and it continues to mutate despite a decrease in the number of deaths, hospitalizations, and Intensive Care Unit (ICU) cases in some countries. The goal of this article is to use mixed models to study the dynamics of the pandemic at different stages based on different determinants in order to benchmark our previous research. To begin, we used a generalized additive model and then a generalized mixed model on our datasets. The findings are consistent with previous findings and provide valuable insights (that is, there is a relationship between socio-economic, demographic and epidemiology variables) into the modeling of the COVID-19 outbreak.

Kayode Oshinubi
Université Grenoble Alpes

Université Grenoble Alpes
kayode.oshinubi@univ-grenoble-alpes.fr

Jacques Demongeot, Mustapha Rachdi
Université Grenoble Alpes
jacques.demongeot@univ-grenoble-alpes.fr,
mustapha.rachdi@univ-grenoble-alpes.fr

PP1

Streamrak - a Streaming Multi-Resolution Adaptive Kernel Algorithm

Kernel ridge regression (KRR) is a popular kernel method for non-linear non-parametric learning. Yet, existing implementations of KRR require that all data is stored in-memory, which severely limits the use of KRR in contexts where data size far exceeds in-memory size. One approach to address this issue is the streaming model of computation. Furthermore, in high dimensions, KRR suffers from the "curse of dimensionality" requiring large amounts of training data to converge. Moreover, the computational complexity, memory requirement, and the number of parameters to learn grow unbounded with the number of training samples, a drawback known as the "curse of kernelization". In the context of streaming, prospects of unbounded data streams makes these shortcomings even more detrimental. In this talk we present StreaMRAK - a streaming multi-resolution version of KRR. StreaMRAK uses a streaming computational model to reduce the in-memory requirements. Furthermore, a new sub-sampling scheme is proposed to address the curses of dimensionality and kernelization. The sub-sampling is adapted to a multi-resolution scheme, which once combined with a localized kernel, gives frequency and location-based discretization, similar to the wavelet framework that has shown great success in numerous applications. The presented numerical studies show that the proposed algorithm is fast and accurate. The Abstract is partially adapted from [A. Oslandsbotn et al, StreaMRAK, arXiv 2108.10411].

Andreas Oslandsbotn
Simula Research Laboratory
andreaso@simula.no

Zeljko Kereta
University College London
z.kereta@ucl.ac.uk

Valeriya Naumova
Simula Research Laboratory
valeriya@simula.no

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

Yoav Freund
University of California San Diego
yfreund@eng.ucsd.edu

PP1

Tensor Train Approximation in Semi-Global Optimal Control and Mean Field Games

The Tensor Train decomposition provides a powerful tool in the approximation of high-dimensional functions. Their manifold structure allows to use the tools of Riemannian optimization. In the context of semi-global feedback stabilization and machine learning in high-dimensional mean field games some adjoint methods based on the Pontryagin maximum principle have been established. We will combine these adjoint methods with optimization methods on the Tensor Train manifold and show numerical results.

Mathias Oster
Technical University Berlin
Oster@math.tu-berlin.de

Reinhold Schneider
Technische Universitat Berlin
schneidr@math.tu-berlin.de

PP1

Physics-Informed Machine Learning for Underground Reservoir Pressure Management with Heterogeneity

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.

Aleksandra Pachelieva
Los Alamos National Laboratory, U.S.
apachalieva@lanl.gov

PP1

Nonparametric Methods for Multivariate Statistical Process Control Using Sequential Normal Scores

Many methods exist for detecting shifts in the means of data arising from both univariate and multivariate normal distributions, such as CUSUM/EWMA control charts. In this project, we develop a new method to detect small shifts in mean vectors from arbitrary multivariate distributions using sequential normal scores based on the work of "Conover et al (2018)" for univariate distributions. We use simulation studies to compare the performance of our methodology to the parametric methods suggested by "Qiu (2013)".

Chathuri N. Perera, William Conover, Leif Ellingson
Texas Tech University
Nuwanthika.Perera@ttu.edu, jay.conover@ttu.edu,
leif.ellingson@ttu.edu

PP1

Nested Smoothing Algorithms for Joint Parameter and State Estimation of Heterogeneous Multi-Scale State-Space Systems

Multi-scale problems, where variables of interest evolve in different time-scales and live in different state-spaces, can be found in many fields of science where complex series of data have to be analyzed. Here, we introduce a new recursive methodology for Bayesian inference that aims at estimating the static parameters and tracking the dynamic variables of these kind of systems. Although the proposed approach works in rather general multi-scale systems, for clarity we analyze the case of a heterogeneous multi-scale model with 3 time-scales (static parameters, slow dynamic state variables and fast dynamic state variables). The proposed scheme, based on the nested filtering methodology of [S. Prez-Vieites, I. P. Mario, J. Mguez. Probabilistic scheme for joint parameter estimation and state predic-

tion in complex dynamical systems. Physical Review E, 98(6), 063305, 2017], combines three intertwined layers of filtering techniques that approximate recursively the joint posterior probability distribution of the parameters and both sets of dynamic state variables given a sequence of noisy data. We explore the use of sequential Monte Carlo schemes and Gaussian filtering techniques in the different layers of computation. Some numerical results are presented for a stochastic two-scale Lorenz 96 model with unknown parameters.

Sara Pérez-Vieites, Harold Molina-Bulla, Joaquín Míguez
Universidad Carlos III de Madrid
saraperezvieites198@gmail.com, hmolina@tsc.uc3m.es, jmiguez@tsc.uc3m.es

PP1

Similarity Based Lexicographic Method for Hierarchical Multiobjective Linear Programs

We develop a new lexicographic method for solving a hierarchical multiobjective linear program (h-MOLP). It exploits the structure of the underlying hierarchical model by monitoring the changes that occurred in the input parameters and leverages re-optimization when solving the objectives in the hierarchy. We define a similarity measure between intermediate linear programs appearing while solving the model and use it to decide whether we should solve the current LP from scratch or use the available feasible solution obtained from the previous LP solve. We show the computational effectiveness of our approach by comparing it with the standard lexicographic method for some h-MOLP.

Devanand R
Indian Institute Of Technology Bombay
devanand@iitb.ac.in

Ashutosh Mahajan, Nandyala Hemachandra
IIT Bombay
amahajan@iitb.ac.in, nh@iitb.ac.in

PP1

Exploring Rideshare Usage Pattern Using K Protype Segmentation Analysis for the City of Chicago in the Era of Pandemic

Taking a page from Soria et. al. we seek to use data science algorithms to analyze customer ride-sharing data for the City of Chicago. Recent legislation has given us access to ridesharing data on a level never before seen for large cities, most prominently Chicago. Previous authors were therefore limited by the available data and the tools they had at their disposable (some authors literally rode on Uber to gain data). This access to such data enables us to make use of sophisticated data science algorithms, yielding quantitatively and qualitatively different data. My coauthors and I focus on the year of 2020. Not only do we seek to cluster an entire year's worth of data and data during the pandemic, the clustering we do uses data on several orders of magnitude than have been attempted previously, especially by Soria et. al. We have gained surprising insights into the ridesharing behavior of Chicago residents (aggregated at a census tract level), that could have implications for better policy making. In particular the interface between employment density, median income, and density of rideshares raises important questions about the inequality of access to transportation in the Chicago area.

Shakil Rafi, Anindya Kishore Debnath, Arna Nishita Nithila
University of Arkansas
sarafi@uark.edu, akdebnat@uark.edu, annithil@uark.edu

PP1

Machine Leaning Models for the Prediction of US House Election Outcome Using District Level Social and Economic Predictors

The democratic process serves as a core function in the redistribution of power within societies. The winner gains the opportunity to redefine priorities and the reallocation of resources. The loser loses access to power and decision making. As such, it is a sensitive and critical process with multiple stakeholders attempt to predict its outcome and result. The most common type of predicting elections outcome is based on polls. In this study we explore an alternative approach, using statistical models to predict election outcomes. Drawing from a body of theory from sociology and political science we suggest a new predictive model using the Support Vector Machines (SVM) and Random Forests machine learning approaches. Our model utilizes longitudinal socioeconomic, demographic, and political data. The proposed model was trained with US house election outcomes and district level socio-economic data from 2000 to 2016. Election cycle data for 2018 and 2020 were used to validate the developed model.

Kusal Rathnayake
TExas Tech University
kusal.rathnayake@ttu.edu

Dimitri Volchenkov, Ori Swed
Texas Tech University
dimitri.volchenkov@ttu.edu, ori.swed@ttu.edu

PP1

Mini-Max Entropy Super Resolution by Wasserstein Metric

Super resolution is an essential tool in optics, especially on interstellar scales, due to physical laws restricting possible imaging resolution. We propose using optimal transport and entropy for super resolution applications. We prove that the reconstruction is accurate when sparsity is known and noise or distortion is small enough. We prove that the optimizer is stable and robust to noise and perturbations. We compare this method to a state of the art convolutional neural network and get similar results for much less computational cost and greater methodological flexibility.

Michael G. Rawson
Pacific Northwest National Lab
michael.rawson@pnnl.gov

Jakob Hultgren
University of Maryland at College Park
hultgren@umd.edu

PP1

Interlaced Characterization and Calibration: Online Bayesian Optimal Experimental Design for Constitutive Model Calibration

Engineering decisions have become increasingly reliant on

computational simulations, necessitating a high confidence in model calibration and predictions. Importantly, this reliance has also meant adopting models with increasing complexity, requiring larger experimental calibration campaigns and time-consuming processes. Traditionally, simple specimens with homogeneous loading states have been used for data-generation. However, oftentimes such tests are not representative of real-world loading conditions and limit the stress-states that may be probed. These factors emphasize the importance of model development, the careful design of experiments for model calibration to minimize cost, and the use of techniques to quantify and minimize parameter uncertainty. This work reconsiders the calibration process of material constitutive models for solid mechanics simulations, which simulate material behavior under different loading conditions and process history. The models are calibrated with dimension-reduced full-field Digital Image Correlation data generated under heterogeneous loading conditions. Additionally, an Interlaced Characterization and Calibration (ICC) framework is proposed. This uses Bayesian Experiment Design to iteratively perform inference on model parameters and determine the next best experiment to conduct to minimize parameter uncertainty in a real-time feedback loop. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA000352

Denielle Ricciardi
Sandia National Laboratories
dericci@sandia.gov

Tom Seidl
Sandia National Lab
dtseidl@sandia.gov

Brian Lester, Amanda Jones, Elizabeth Jones
Sandia National Laboratories
btleste@sandia.gov, ajones1@sandia.gov, emjones@sandia.gov

PP1

Harfe: Hard-Ridge Random Feature Expansion

We propose a random feature model for approximating high-dimensional sparse additive functions called the hard-ridge random feature expansion method (HARFE). This method utilizes a hard-thresholding pursuit based algorithm applied to the sparse ridge regression (SRR) problem to approximate the coefficients with respect to the random feature matrix. The SRR formulation balances between obtaining sparse models that use fewer terms in their representation and ridge-based smoothing that tend to be robust to noise and outliers. In addition, we use a random sparse connectivity pattern in the random feature matrix to match the additive function assumption. We prove that the HARFE method is guaranteed to converge with a given error bound depending on the noise and the parameters of the sparse ridge regression model. Based on numerical results on synthetic data as well as on real datasets, the HARFE approach obtains lower (or comparable) error than other state-of-the-art algorithms.

Esha Saha
University of Waterloo
esaha@uwaterloo.ca

Hayden Schaeffer
Carnegie Mellon University
schaeffer@cmu.edu

Giang Tran
University of Waterloo
giang.tran@uwaterloo.ca

PP1

Multi-Objectivizing General Single Nonlinear Continuous Optimization Problems

Within the framework of continuous single-objective nonlinear optimization. Ideally, one is interested in finding the global solution(s) to the problem in question which is indeed a tremendously complex task to be achieved via purely deterministic methods, although theoretically permissible, it is unbelievably costly from the computational perspective. In practice, seldom when we find a good approximation to a local optimum in a reasonable amount of time and we often settle for a good approximation to a KKT point or a critical point to the problem in question. This problem stems from the fact that it is impossible to infer global information about the landscape of the function in question given the local information that is available to the method. Thus, in the pursuit of attempting to overcome the addressed intrinsic yet crucial difficulty while at the same time not sacrificing the element of computational efficiency, we take advantage of the theory of multi-objective optimization in multi-objectivizing the general nonlinear single-objective problem at hand, while benefiting from the structure of stochastic algorithms, in particular, we use a variation of the genetic algorithm so to benefit from its parallelstic approach in navigating the feasible set, and at the same time, we strive to increase the probability of it reaching a tight approximation to the global solution(s).

Ehsan Shams
Master's student
Faculty of Science - Alexandria University
ehsan.ali@alexu.edu.eg

PP1

A Compressed Sensing Based Least Squares Approach to Semi-Supervised Local Cluster Extraction

A least squares semi-supervised local clustering algorithm based on the idea of compressed sensing is proposed to extract clusters from a graph with known adjacency matrix. The algorithm is based on a two-stage approach similar to the one in [Ming-Jun Lai and Daniel Mckenzie, Compressive sensing approach to cut improvement and local clustering, SIAM Journal on Mathematics of Data Science, 2 (2020), pp. 368395]. However, under a weaker assumption and with less computational complexity than the method proposed by Lai and Mckenzie, our algorithm is shown to be able to find a desired cluster with high probability. The "one cluster at a time" feature of our method distinguishes it from other global clustering methods. Several numerical experiments are conducted on the synthetic data such as stochastic block model and real data such as MNIST, political blogs network, ATT and YaleB human faces data sets to demonstrate the effectiveness and efficiency of our algorithm.

Zhaiming Shen, Ming-Jun Lai
University of Georgia

zs98306@uga.edu, mjlai@uga.edu

PP1

Dimension-Independent Markov Chain Monte Carlo on the Sphere

We consider Bayesian analysis on high-dimensional spheres with angular central Gaussian priors. These priors model antipodally-symmetric directional data, are easily defined in Hilbert spaces and occur, for instance, in Bayesian binary classification and level set inversion. In this talk we derive efficient Markov chain Monte Carlo methods for approximate sampling of posteriors with respect to these priors. Our approaches rely on lifting the sampling problem to the ambient Hilbert space and exploit existing dimension-independent samplers in linear spaces. By a push-forward Markov kernel construction we then obtain Markov chains on the sphere, which inherit reversibility and spectral gap properties from samplers in linear spaces. Moreover, our proposed algorithms show dimension-independent efficiency in numerical experiments.

Bjoern Sprungk
TU Freiberg
bjoern.sprungk@math.tu-freiberg.de

Han Cheng Lie
University of Potsdam
han.lie@uni-potsdam.de

Daniel Rudolf
University of Passau
daniel.rudolf@uni-passau.de

Tim Sullivan
University of Warwick
t.j.sullivan@warwick.ac.uk

PP1

Using Synthetic Data to Train Neural Networks for an Automated Detection of Rub in Aeroderivative Gas Turbines

Rotor-casing rub is one of the most common failures in gas and steam turbines. In most cases, rub can be easily detected. However, in aero-derivative gas turbines, the only available source of information about the phenomenon is a noisy indirect signal generated by accelerometers on the casing. The volume of real data from faulty turbines under rub is insufficient to train deep learning algorithms since these faults are often destructive when heavy enough. This is the case, particularly with accelerometer signals from aero-derivative gas turbines. As a solution to this problem, we propose in this work to train automated rotor-casing contact detection systems based on Deep Neural Networks with data generated by a Finite Element Model of the rotating machine of interest. The resulting network is tested on casing acceleration data from an experimental rotating machine under a single-point rub to verify if our synthetic data-trained networks would be able to generalize the information of the training set. The samples are pre-processed by applying synchronous resampling based on machine rotations and the discrete Fourier transform in real and numerical simulated datasets. The best-performing network has shown accuracy rates up to 93% during the test on real experimental data. Results show that the algorithm can generalize the information contained in the training set, obtaining satisfactory results in a real-case scenario.

Antonio Squicciarini
Universidad Politecnica de Madrid
Departamento Matemática Aplicada a la Ingeniería Industrial
a.squicciarini@alumnos.upm.es

Carlos Eduardo Gonzalez Gguillen, Alejandro Zarzo Altarejos, Alejandro Silva
Departamento Matemática Aplicada a la Ingeniería Industrial
Universidad Politécnica de Madrid
carlos.gguillen@upm.es, alejandro.zarzo@upm.es, alejandro.silva@upm.es

PP1

Radiomics on Spatial-Temporal Manifolds via Fokker-Planck Dynamics

Purpose: To develop a new radiomics paradigm for sparse, time-series imaging data, where features are extracted from a spatial-temporal manifold modeling the time evolution between images. Methods: Radiomics is a high-throughput approach to image phenotyping which extracts mineable quantitative features from medical images using data analytics and deep learning. We developed an equilibrium-driven deformation algorithm to model the temporal evolution between two images acquired at t=0 and t¿0. Images serve as no-flux boundary conditions of the Fokker-Planck equation. This generates a 4D spatial-temporal manifold from which radiomic features are extracted. First, our approach was numerically verified by stochastically sampling dynamic Gaussian processes of monotonically decreasing noise. The noise transformation was compared between Fokker-Planck estimation and simulated ground-truth. Second, we conducted a patient study to estimate early metabolic response of patients undergoing definitive radiotherapy for oropharyngeal cancer. Results: Numerical results confirmed our technique can recover image noise characteristics given sparse input data as boundary conditions. Compared to ground-truth, the estimated impulse response of energy and entropy achieved cross-correlations of 0.82 and 0.94, respectively. Conclusion: We developed, verified and applied a new approach to sparse, time-series image characterization via data assimilation of radiomics with partial differential equations.

Jack B. Stevens, Jihyeon Je
Duke University
jack.stevens@duke.edu, jihyeon.je@duke.edu

Yuan Gao
Department of Mathematics
Purdue University, USA
gao662@purdue.edu

Chunhao Wang
Duke University
chunhao.wang@duke.edu

Yvonne Mowery, David Brizel
Duke University Medical Center
yvonne.mowery@duke.edu, david.brizel@duke.edu

Fang-fang Yin, Jian-guo Liu, Kyle Lafata
Duke University
fangfang.yin@duke.edu, jian-guo.liu@duke.edu,

kyle.lafata@duke.edu

PP1

Satellite Image Analysis for Forest Fire Prediction

Satellite Image Analysis for forest fire prediction.

Alfredo Suarez
Southwestern College, U.S.
alfredo.suarez@sckans.edu

PP1

Transport Transforms for 1-D Signal Analysis and Classification in Machine Learning

Classification and estimation problems are at the core of machine learning. In this talk we will see a new mathematical signal transform that renders data easy to classify or estimate by decoding information related to non rigid signal displacements. The idea stems from a very old theory of transportation that was started by Monge. We will learn about the existing Cumulative Distribution Transform [**?**] and extend to a more general measure theoretic framework, to define the new transform (Signed Cumulative Distribution Transform), that works for arbitrary (signed) signals on
RE. We will look at both forward (analysis) and inverse (synthesis) formulas for the transform, and describe several of its properties including translation, scaling, convexity, linear separability and others. Finally, we will demonstrate the application of the transform in classifying (detecting) signals under random displacements.

Sumati Thareja
Vanderbilt University
sumati.thareja@vanderbilt.edu

PP1

A Case Study on Artifacts in Manifold Learning for Cryo-EM

Cryo-electron microscopy (Cryo-EM) is a technique that allows mapping the 3D structure of biomolecules by embedding them in a layer of vitreous water and imaging them with an electron microscope, which results in hundreds of thousands of 2D noisy projections of instances of a macromolecular structure with random unknown orientations and in-plane shifts. Obtaining the 3D structure that generated these highly noisy projections involves solving an inverse problem where the measurement operator itself is unknown, namely the unknown rotations and in-plane shifts corresponding to each projection. Recently, manifold learning algorithms have been used to estimate the conformational landscape of a molecule from its 2D projections. This relies on the implicit assumption that the distribution of 2D particles on the low-dimensional manifold that such algorithms compute reflects the physical conformational landscape of the (3D) molecule. In this work, we show how this may not always be the case by considering the problem of estimating the manifold of 3D objects from their 2D projections in a simplified setting that captures the main characteristics of the cryo-EM problem.

Bogdan Toader, Roy Lederman
Yale University
bogdan.toader@yale.edu, roy.lederman@yale.edu

PP1

Synthesis Sparse Dictionary Learning for Ptychographic Probe Recovery

Ptychography is a technique which uses a partially coherent electromagnetic wavefield to probe a sample with complex valued index of refraction at high spatial resolution, which when combined with tomographic methods results in 3D nanoscale images. It solves a constrained optimization problem to recover quantitative images of the complex valued sample as well as magnitude and phases of the probing wavefield, which is decomposed into orthogonal spatial coherence probe modes (SCPMs). Typically the problem is solved iteratively with total complex valued degrees of freedom (image pixels) in the millions, potentially billions. It would be useful for increased solution robustness and quality (and decreased computational effort) if as many constraints about the sample, probing wavefield, and experimental arrangement were used in the iterative solution process to reduce the allowable degrees of freedom of the optimization problem. In furtherance of this, we will discuss the use of Zernike Polynomials to solve a synthesis sparse dictionary learning problem to drastically reduce the degrees of freedom for the SCPMs. This will significantly accelerate the solution of these SCPMs, which in turn will affect the solution of the sample images, resulting in a far more numerically robust and computationally cheaper ptychography image reconstruction algorithm. We will present benchmarking results as well as ideas to further reduce the degrees of freedom for this problem.

Ashish Tripathi
Mathmatics and Computer Science Division
Argonne National Laboratory
atripath@anl.gov

PP1

Towards Reinforcement Learning-Driven Mesh Adaptivity for Second Order Elliptic Problems

Adaptive mesh reinforcement techniques have become an indispensable tool in achieving accurate and efficiently computed solutions to problems which require impractically fine uniform meshes to obtain an accurate approximation. The adaptive algorithm involves a recursive application of SOLVE-ESTIMATE-MARK-REFINE steps where in particular the step 'ESTIMATE' involves computing a posteriori error estimator based on only the numerical solution and the data of the problem. Over the years, several a posteriori error estimators have been developed and successfully applied but often times, the choice of estimator is ill suited for the problem at hand. In this research, we present two estimators namely the residual-based estimator and the gradient recovery type estimator and illustrate the suitability of these estimators for different problems at hand. We also provide the foundation for data-driven adaptive mesh refinement strategies based on Reinforcement learning (RL) with a focus on the Q-learning algorithm which is a fundamental learning algorithm in RL.

Augustine Twumasi
University of Texas at El Paso

atwumasi@miners.utep.edu

PP1

Correlation Metrics for Image Assessment in Single Particle Cryo-Em

Single particle electron cryo-microscopy (cryo-EM) is an imaging technique capable of producing high-resolution structures of biomolecules from many noisy projection images.Prior to computing the 3D structure, a common step is to assess the quality of data by clustering the projection images (referred to as 2D class averages), followed by manual filtering by an expert. The standard metric used to estimate resolution in cryo-EM is the Fourier shell correlation (FSC), which measures the normalized cross-correlation between corresponding shells in Fourier space from two half sets of the data. Although the FSC was developed from its 2D analogue, the Fourier ring correlation (FRC), the FRC is not commonly used to assess image quality in the steps prior to 3D structure determination.Here we show that the radially averaged power spectrum in combination with a modified FRC computed from subsamples of the same image can be used to rank the quality of images automatically.

Eric J. Verbeke
Princeton University
ev9102@princeton.edu

PP1

Learning in High-Dimensional Feature Spaces Using Anova-Based Fast Matrix-Vector Multiplication

Kernel matrices are crucial in many learning tasks and typically dense and large-scale. Depending on the dimension of the feature space even the computation of all its entries in reasonable time becomes a challenging task. For such dense matrices the cost of a matrix-vector product scales quadratically with the dimensionality, if no customized methods are applied. In basically all kernel methods, a linear system must be solved. Our approach exploits the computational power of the non-equispaced fast Fourier transform (NFFT), which is of linear complexity for fixed accuracy. The ANOVA kernel has proved to be a viable tool to group the features into smaller pieces that are then amenable to the NFFT-based summation technique. Multiple kernels based on lower-dimensional feature spaces are combined, such that kernel-vector products can be realized by this fast approximation algorithm. Based on a feature grouping approach this can be embedded into a CG solver within a learning method and we nearly reach a linear scaling. This approach enables to run learning tasks using kernel methods for large-scale data on a standard laptop computer in reasonable time without or very benign loss of accuracy. It can be embedded into methods that rely on kernel matrices or even graph Laplacians. Examples are support vector machines or graph neural networks that can then benefit from having the fast matrix-vector products available.

Theresa Wagner, Franziska Nestler
TU Chemnitz
theresa.wagner@math.tu-chemnitz.de,
franziska.nestler@math.tu-chemnitz.de

Martin Stoll
Department of Mathematics,
Technische Universitaet Chemnitz
martin.stoll@math.tu-chemnitz.de

PP1

Testing Hyperbolicity and Finding Determinantal Representation from Sample Points

Learning varieties from sample points is a fundamental problem at the interface of data science and algebraic geometry. Hypersurface defined by a hyperbolic polynomial is a special class of varieties. Hyperbolic polynomials play an important role in PDE, convex optimization and numerical algebra. Helton and Vinnikov made a brilliant contribution to find the connection between the hyperbolicity of a homogeneous polynomial and the existence of a real symmetric determinantal representation. For a known hyperbolic polynomial, Plaumann and Vinzant gave an algorithm of computing its Hermitian determinantal representation. We consider a further problem: given a set of real zero points of a degree d homogeneous polynomial $p(x, y, z)$, we use tools from numerical algebraic geometry to determine whether p is hyperbolic and find its real symmetric determinantal representation.

Zinan Wang, Jose Rodriguez
University of Wisconsin-Madison
zwang894@math.wisc.edu, jose@math.wisc.edu

PP1

PYOMO.DOE: An Open-Source Package for Model-Based Design of Experiments in Python

Predictive mathematical models are a cornerstone of science and engineering. Yet calibrating and validating said models need a large amount of information from physical and computer experiments. Model-based design of experiments (MBDoE) provides a systematic framework to maximize information gain from experiments while minimizing resource costs. Unlike classical 'black-box' DoE methods (e.g., factorial), MBDoE leverages science-based models and facilitates efficient gradient-based optimization. Despite its rich history of success, MBDoE remains limited to niche application areas, in part because practitioners must integrate expertise in statistics, optimization, and modeling. To reduce this barrier, we present Pyomo.DOE, the first open-source package for MBDoE. Pyomo.DOE uses a nonlinear sensitivity analysis code k_aug to approximate Fisher information matrix, reducing the computational time by over one order of magnitude. Pyomo.DOE also leverages a new two-stage stochastic programming abstraction to facilitate the gradient-based optimization framework. Case study 1 considers a kinetic model with highly-correlated parameters, showing how Pyomo.DOE eliminates model unidentifiability. Case study 2 considers CO_2 adsorption as the first application of MBDoE to fixed-bed breakthrough experiments, and reveals that MBDoE provides a principled approach to estimate the value of additional measurements or modifications before changing experimental campaigns in the laboratory.

Jialu Wang, Alexander Dowling
University of Notre Dame
jwang44@nd.edu, adowling@nd.edu

PP1

Structure-Preserving Learning of Mechanical Sys-

tems

Data-driven reduced-order modeling is an essential tool in constructing high-fidelity compact models to approximate physical phenomena when explicit models, state-space formulations with access to internal variables, are not available yet abundant input/output data are. In the case of mechanical processes, data is often available in the frequency domain, where the systems' input-to-output behavior is described by rational functions rather than differential equations. While classical frequency domain approaches like the Loewner framework, vector fitting and AAA are available, these methods result in unstructured models, which lack any physical interpretation and structure-inherited properties. In this work, we present a new framework for the learning of mechanical phenomena in the frequency domain, while enforcing the mechanical system structure in the model description.

Steffen W. R. Werner
Courant Institute of Mathematical Sciences
New York University
steffen.werner@nyu.edu

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

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

PP1

Risk-Aware Bayesian Goal-Oriented Optimal Experimental Design

Optimal experimental design (OED) allows us to predict, a priori, when and where to collect data to minimize uncertainty regarding a system. Traditionally, the aim has been to minimize uncertainty related to model calibration. Our work instead aims to develop goal-oriented strategies which minimize uncertainty directly related to our modeling objectives, such as prediction or control objectives. We consider a Bayesian framework for OED to propagate uncertainty measurements through to our respective goals. This goal-oriented Bayesian regime allows us to consider experimental design for high-risk scenarios, determining conservative strategies that appropriately account for risk. In this work, we investigate goal-oriented, risk-averse design for nonlinear mathematical models governed by PDEs. We compare our approach to a standard, risk-neutral approach using a sensor placement problem where the physics are governed by diffusive transport equations.

Rebekah White
Sandia National Laboratories
rebwhit@sandia.gov

PP1

U.S. Rental Moratorium: A Visual Analysis of the Likelihood of Eviction in the Next Two Months

Renters in the U.S. are currently facing possible eviction when the government-mandated rental moratorium ends. In the early phases of the pandemic, millions of Americans lost their primary source of income and were unable to cover rental payments. To protect renters and prevent the spread of COVID-19, the Centers for Disease Control and Prevention (CDC) halted evictions in counties with high transmission. This research aims to show the impact of the COVID-19 pandemic on housing insecurity and test the following hypotheses: 1) state and government acts relating to rental moratorium directly impact renter's attitudes about eviction; and 2) less populated states are at greater risk of being affected due to lack of infrastructure to effectively distribute rental aid. A visual analytics system was developed to visually investigate Phase 2 Phase 3.1 survey data from the U.S. Census Bureau covering a duration of 46-weeks. Results showed a correlation between government policies and renter's feelings of insecurities, and that smaller states had the greatest rates of feelings of insecurities that directly related to lack of infrastructure for distributing relief funds.

Jordan Williams
Grambling State University, U.S.
jordanappl1901@gmail.com

PP1

Reinforcement Learning for PDE-Based Control Problems

The process of identifying optimal control policies plays a key role in a wide range of scientific applications. While many complex systems can be modeled effectively with numerical methods, it is often difficult to determine the best strategy for interacting with such systems based on the forward model alone. In this poster, we explore the potential for leveraging deep reinforcement learning algorithms to automate the decision-making process in the context of scientific applications. In particular, we show that a proximal policy optimization algorithm can successfully identify a near-optimal strategy for positioning sinks within a simplified PDE model of contaminant flow. The generality of the underlying learning scheme, along with the promising results from initial experiments, suggest that reinforcement learning may provide an effective framework for optimizing the decision-making process in a wide range of scientific applications.

Nickolas Winovich
Sandia National Laboratory, U.S.
nwinovi@sandia.gov

PP1

Adversarial Examples in Random Neural Networks with General Activations

A substantial body of empirical work documents the lack of robustness in deep learning models to adversarial examples. Recent theoretical work proved that adversarial examples are ubiquitous in two-layers networks with sub-exponential width and ReLU or smooth activations, and multi-layer ReLU networks with sub-exponential width. We present a result of the same type, with no restriction on width and for general locally Lipschitz continuous activations. More precisely, given a neural network $f(\cdot\,;\theta)$ with random weights θ, and feature vector $\mathbf{x}$, we show that an adversarial example $\mathbf{x}'$ can be found with high probability along the direction of the gradient $\nabla_{\mathbf{x}} f(\mathbf{x};\theta)$. Our proof is based on a Gaussian conditioning technique. Instead of proving that f is approximately linear in a neighborhood of $\mathbf{x}$, we characterize the joint distribution of $f(\mathbf{x};\theta)$ and $f(\mathbf{x}';\theta)$ for $\mathbf{x}' = \mathbf{x} - s(\mathbf{x})\nabla_{\mathbf{x}} f(\mathbf{x};\theta)$.

Yuchen Wu, Andrea Montanari
Stanford University

wuyc14@stanford.edu, montanari@stanford.edu

PP1

Robust Denoising of Cryo-Em Images via β-Gan

The Cryo-Electron Microscopy (Cryo-EM) becomes popular for macromolecular structure determination. However, the 2D images captured by Cryo-EM are of high noise and often mixed with multiple heterogeneous conformations and contamination, imposing a challenge for denoising. In this work, we approach the robust denoising problem for Cryo-EM images by introducing a family of Generative Adversarial Networks (GAN), called β-GAN, which is able to achieve robust estimation of certain distributional parameters under Huber contamination model with statistical optimality. To address the denoising challenges, β-GANs are exploited to enhance the robustness of Denoising Autoencoder. Our proposed method is evaluated by both a simulated dataset on the Thermus aquaticus RNA Polymerase (RNAP) and a real world dataset on the Plasmodium falciparum 80S ribosome dataset (EMPIAR-10028), in terms of Mean Square Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) and 3D Reconstruction as well. Quantitative comparisons show that equipped with some designs of β-GANs and the robust ℓ_1-Autoencoder, one can stabilize the training of GANs and achieve the state-of-the-art performance of robust denoising with low SNR data and against possible information contamination. Our proposed methodology thus provides an effective tool for robust denoising of Cryo-EM 2D images, and helpful for 3D structure reconstruction.

Yin Xian
Hong Kong Applied Science and Technology Research Institute
poline3939@gmail.com

PP1

Overparameterization and Generalization Error: Weighted Trigonometric Interpolation

Motivated by surprisingly good generalization properties of learned deep neural networks in over-parameterized scenarios and by the related double descent phenomenon, this paper analyzes the relation between smoothness and low generalization error in an overparameterized linear learning problem. We study a random Fourier series model, where the task is to estimate the unknown Fourier coefficients from equidistant samples. We derive exact expressions for the generalization error of both plain and weighted least squares estimators. We show precisely how a bias towards smooth interpolants, in the form of weighted trigonometric interpolation, can lead to smaller generalization error in the overparameterized regime compared to the underparameterized regime. This provides insight into the power of overparameterization, which is common in modern machine learning.

Yuege Xie
UT Austin
yuege@oden.utexas.edu

Hung-Hsu Chou
RWTH Aachen University
chou@mathc.rwth-aachen.de

Holger Rauhut
RWTH Aachen University
Lehrstuhl C für Mathematik (Analysis)
rauhut@mathc.rwth-aachen.de

Rachel Ward
Department of Mathematics
University of Texas at Austin
rward@math.utexas.edu

PP1

Spde-NetII: Optimal Stabilization Parameter Prediction with Neural Networks

Singularly Perturbed Partial Differential Equations (SPPDEs) are challenging to solve, and a one-fit-all strategy for the solution of SPPDEs does not yet exist. A blind application of mesh-based numerical methods, such as the Finite Element Method, for this purpose may result in nonphysical oscillations in the numerical solution. Over the years, a number of stabilization techniques have been proposed in order to obtain a stable solution sans spurious oscillations. However, most of the stabilization techniques rely on an optimal value of the stabilization parameter, which unfortunately remains difficult to evaluate. In this work, we propose a deep neural network based approach for approximating the stabilization parameter for an accurate and stable solution of the 2-dimensional convection dominated convection-diffusion equation. In this technique, the stabilization parameter is obtained by minimizing the residual, along with the crosswind term for the particular SPPDE under consideration. We show that this approach yields stable solutions for several benchmark problems involving 2D SPPDEs, while also being more accurate compared to other contemporary neural networks based PDE solvers.

Sangeeta Yadav
Indian Institute of Science
shongmusic@gmail.com

PP1

Optical Neural Networks and Approximation Properties of Complex-Valued Neural Networks

Artificial neural networks are usually considered to be real-valued, i.e., the numbers that propagate through the network are real numbers. In certain applications, such as in applications that work in the frequency domain, the natural domain for the inputs of the network is the space of complex numbers. Also, in some emerging technologies, such as in optical computing, the signals that perform the computation are inherently complex-valued. In this talk, we discuss a recent design for an optical neural network proposed by the authors ('Analysis of a dynamical system modeling lasers and applications for optical neural networks', arXiv:2103.02678 (2021)). We also consider the question of approximation properties of complex-valued neural networks in general. For the real-valued neural networks, the classical universal approximation theorem states that nonpolynomiality of the activation function is essentially a necessary and sufficient condition for the associated neural networks to be capable of uniformly approximating any continuous function. An analogous result for the complex-valued neural networks was proved only recently by Voigtlaender ('The universal approximation theorem for complex-valued neural networks', arXiv:2012.03351 (2020)) for globally defined activation functions, and later extended in connection with optical neural networks to locally defined activation functions by the authors.

Lauri A. Ylinen

University of Helsinki
lauri.ylinen@helsinki.fi

Tuomo Von Lerber, Franko Küppers
Center for Photonics and Quantum Materials
Skolkovo Institute of Science and Technology (Skoltech)
t.vonlerber@skoltech.ru, f.kueppers@skoltech.ru

Matti Lassas
University of Helsinki
matti.lassas@helsinki.fi

PP1

Low-Rank Matrix Recovery: Global Convergence and Riemannian Geometry

Low-rank matrix manifold, the Riemannian manifold of matrices with fixed size and rank, is a powerful tool for low-rank recovery applications in compressed sensing and data science. We propose a unified framework on the low-rank matrix manifold for a class of problems, such as matrix sensing and phase retrieval. We analyze the global behavior of vanilla Riemannian gradient descent (RGD) with random initialization for the population loss function of these problems. We show that under certain assumptions, RGD with global random initialization converges to the global minimum in a nearly linear rate that is almost dimension-free. In other words, it takes $O(log(n) + log(1/\epsilon))$ iterations to reach an ϵ-accurate solution. A previously unknown geometric property of the manifold is the existence of spurious critical points. They are rank-deficient, singular, saddle-like fixed points of RGD. We prove that with high probability, randomly initialized RGD escapes all the spurious critical points in just a few steps. Moreover, for the rank-(r-1) spurious critical points, RGD escapes them almost surely. Our global analysis and convergence guarantee fully explains the numerical observations. It could potentially be extended to more applications with random measurement structures and empirical least squares loss functions.

Ziyun Zhang, Thomas Hou
Applied and Computational Mathematics
California Institute of Technology
zyzhang@caltech.edu, hou@cms.caltech.edu

Zhenzhen Li
Carnegie Mellon University
zhenzhenli2020@gmail.com

PP1

Extracting Dynamical Models from Data

The FJet approach is introduced for determining the underlying model of data from a dynamical system. It borrows ideas from the fields of Lie symmetries as applied to differential equations (DEs), and numerical integration (such as RungeKutta). The technique can be considered as a way to use machine learning (ML) to derive a numerical integration scheme. The technique naturally overcomes the "extrapolation problem", which is when ML is used to extrapolate a model beyond the time range of the original training data. It does this by doing the modeling in the phase space of the system, rather than over the time domain. When modeled with a type of regression scheme, its possible to accurately determine the underlying DE, along with parameter dependencies. Ideas from the field of Lie symmetries applied to ordinary DEs are used to determine constants of motion, even for damped and driven systems. These statements are demonstrated on three examples: a damped harmonic oscillator, a damped pendulum, and a damped, driven nonlinear oscillator (Duffing oscillator). In the model for the Duffing oscillator, its possible to treat the external force in a manner reminiscent of a Greens function approach. Also, in the case of the undamped harmonic oscillator, the FJet approach remains stable approximately 10**9 times longer than 4th-order Runge-Kutta

Michael Zimmer
Neomath, LLC
zim@neomath.com