ICMEXPO

POSTER SESSION PRESENTERS AND SPEAKERS

Friday, May 20, 2016

Huang Engineering Center, Mackenzie Room

1:00-2:40 P.M.	Poster Session
2:00-2:40 P.M.	HANA Immersive Visualization Environment (HIVE) Demos
2:45-5:00 P.M.	Faculty Vision Talks
5:00-6:30 P.M.	Reception

Faculty Vision Talks

Gunnar Carlsson, Professor of Mathematics, Co-Founder Ayasdi

Gunnar Carlsson is one of the most renowned mathematicians in the world with his undergraduate degree from Harvard University and doctorate from Stanford. Over the past 35 years, Gunnar has taught at University of Chicago, University of California, Princeton University, and since 1991 Gunnar has been a professor of mathematics at Stanford University, where he has been a thought leader in a branch of mathematics called topology, the study of shape. In the theoretical sense, topology has been around since the 1700's, but Gunnar pioneered the applied use of topology to solve complex real world problems starting in the late 1990's. In the early 2000's, this work led to $10M in research grants from the National Science Foundation (NSF) and DARPA to study the application of Topological Data Analysis (TDA) to problems of interest within the U.S. government. In 2008, based on the success of these efforts, Gunnar, along with two other Stanford mathematicians, co-founded Ayasdi. Gunnar is married, has 3 grown boys including two sons who are mathematicians, and lives in Palo Alto.

THE SHAPE OF BIG DATA

There is a lot of discussion around "Big Data", which refers to the large and complex data sets currently being gathered about scientific, engineering, and commercial problems. Although the size of the data is certainly a significant roadblock to the goal of obtaining useful information and knowledge from the data, the complexity of the data is often a more significant hurdle. This means that there is a need for a new organizing principle and modeling mechanism, and it turns out that topology (the mathematical study of shape) can be used to provide such an organizing principle. I will talk about these ideas, with examples coming from the biomedical world and social sciences.

Jenny Suckale, Assistant Professor of Geophysics

Jenny Suckale is an assistant professor of geophysics in the School of Earth, Energy & Environmental Sciences and an affiliated faculty member at the Institute for Computational & Mathematical Engineering and the Woods Institute for the Environment. Prior to joining graduate school at MIT and the Harvard Kennedy School, Suckale worked as a scientific consultant for different international organizations aiming to reduce the impact of natural and environmental disasters in vulnerable communities.

Faculty Vision Talks

HIGH PERFORMANCE COMPUTING FOR NATURAL DISASTER REDUCTION

The motivation behind my research is to create knowledge that reduces the risks associated with natural disasters like volcanic eruptions, induced earthquakes, tsunamis and ice-sheet collapse. The common denominator of what at first glance might seem like disparate systems is multiphase flow. The dynamic interactions between multiple solid and fluid phases, such as ice and melt-water; rocks and wastewater; lava and gas; vegetation and waves, give rise to drastic nonlinearities that govern abrupt change. My group develops original computational methods for advancing our fundamental understanding and predictive capabilities of the multiphase nonlinearities responsible for abrupt changes in many Earth systems.

Johan Ugander,
Assistant Professor of Management Science and Engineering

Johan Ugander is an Assistant Professor of Management Science & Engineering. His research develops algorithmic and statistical frameworks for analyzing social networks, social systems, and other large-scale data-rich contexts. He obtained his Ph.D. in Applied Mathematics from Cornell University in June 2014. From 2010-14 he also held an affiliation with the Facebook Data Science team. For the 2014-15 academic year he was a post-doctoral researcher at Microsoft Research. He joined Stanford in September 2015.

Faculty Vision Talks

EXPERIMENTS WITH NETWORK EFFECTS

A/B testing is a standard approach for evaluating online experiments; the goal is to estimate the "average treatment effect" of a new feature or condition by exposing a sample of the overall population to it. A drawback with A/B testing is that it is poorly suited for experiments involving social interference, when the treatment of individuals spill over to neighboring individuals along an underlying social network. I will discuss recent work on this problem, including how randomization techniques that administer treatments at the level of graph clusters can significantly reduce the bias and variance of experiments in highly networked settings.

Marco Pavone, Assistant Professor of Aeronautics and Astronautics

Dr. Marco Pavone is an Assistant Professor of Aeronautics and Astronautics at Stanford University, where he is the Director of the Autonomous Systems Laboratory. Before joining Stanford, he was a Research Technologist within the Robotics Section at the NASA Jet Propulsion Laboratory. He received a Ph.D. degree in Aeronautics and Astronautics from the Massachusetts Institute of Technology in 2010. His main research interests are in the development of methodologies for the analysis, design, and control of autonomous systems, with an emphasis on autonomous aerospace vehicles and large-scale robotic networks. He is a recipient of an NSF CAREER Award, a NASA Early Career Faculty Award, a Hellman Faculty Scholar Award, and was named NASA NIAC Fellow in 2011. His work has been recognized with best paper nominations or awards at the Field and Service Robotics Conference (2015), at the Robotics: Science and Systems Conference (2014), and at NASA symposia (2015). He is currently serving as an Associate Editor for the IEEE Control Systems Magazine.

Faculty Vision Talks

THE FUTURE OF AUTONOMOUS SYSTEMS: FROM DRIVERLESS CARS TO SPACE ROBOTS

Mobile autonomous systems are poised to transform several sectors, from transportation and logistics all the way to space exploration. In this talk I will briefly review major computational challenges in endowing autonomous systems with fast and reliable decision-making capabilities, and discuss recent advances made at the Stanford Autonomous Systems Laboratory.

Margot Gerritsen,
Director ICME
Stanford University

Margot has been the Director of ICME since 2010. She received her Ph.D. in Scientific Computing and Computational Mathematics at Stanford in 1997. After five years as faculty member at the University of Auckland, she returned to Stanford in 2001. Her primary appointment is in Energy Resources Engineering. Margot specializes in computational modeling of fluid flow processes, with emphasis on reservoir simulation. She teaches several of the ICME core and service courses in numerical analysis and linear algebra, as well as courses in renewable energy and reservoir simulation.

Faculty Vision Talks

ICME OVERVIEW AND WELCOME

Margot will be speaking about the upcoming activities at ICME (summer seminars, innovative student and faculty research, WiDS conference, and ICME Xtend) as well as plans for 2016-17.

Markus Pelger,
Assistant Professor of Management Science and Engineering

Markus Pelger is an Assistant Professor at the Management Science & Engineering Department at Stanford University. His research interests are in statistics, financial econometrics, asset pricing and risk management. Markus received his Ph.D. in Economics from the University of California, Berkeley. He has a Diploma in Mathematics and a Diploma in Economics from the University of Bonn in Germany.

Faculty Vision Talks

ESTIMATING ASSET PRICING FACTORS

Statistical factor analysis based on Principal Component Analysis (PCA) finds factors that can explain most of the covariance structure in the data but has problems identifying factors with a small variance that are important for asset pricing. We propose a modification to the PCA analysis that finds factors with a high Sharpe ratio that can explain both the expected return and covariance structure. We derive the statistical properties of the new estimator. Factors that are weak, i.e. their variance is below a critical value, cannot be estimated with PCA even if a large amount of data is available. However, our estimator can reliably detect all factors that are important for asset pricing. Applying the approach to portfolio and stock data we find factors with Sharpe ratios twice as large as those based on conventional PCA.

Reza Zadeh, Consulting Professor at Stanford, Founder and CEO at Matroid

Reza Bosagh Zadeh is a Consulting Professor in the Institute for Computational & Mathematical Engineering at Stanford University and Founder CEO at Matroid. His work focuses on Machine Learning, Distributed Computing, and Discrete Applied Mathematics. Reza received his PhD in Computational Mathematics from Stanford University under the supervision of Gunnar Carlsson. For his PhD work in Distributed Machine Learning, Reza received the Gene Golub Outstanding Thesis Award. He serves on the Technical Advisory Board of Microsoft and Databricks.

Faculty Vision Talks

SCALABLE MACHINE LEARNING

The emergence of large distributed clusters of commodity machines has brought with it a slew of new algorithms and tools. Many fields such as Machine Learning and Optimization have adapted their algorithms to handle such clusters. This 10 minute talk will touch upon widely used systems and applications of scaled machine learning.

Faculty Vision Talks

Wing Wong,
Professor of Statistics and Biomedical Data Science

Dr. Wong is Professor of Statistics and Professor of Biomedical Data Science at Stanford University. His current research is motivated by problems from personalized medicine and systems biology. He is developing statistical methods and computational solutions to these problems. In the past his group has contributed a number of widely used bioinformatics tools, and technologies from his group have led to the formation of several companies in the space of genomics data analysis and personalized prognostics.

INTERPRETATION OF GENETIC VARIANTS

Every genome carries millions of genetic variants. Many of these variants are expected to have profound implications on health and diseases. Currently we have reasonable confidence in interpreting some of the variants that affect gene-coding regions of the genome. However, the overwhelming majority of the variants are located in non-coding parts of the genome and current methods to interpret such variants are woefully inadequate. In this talk I will discuss strategies that may help to close this gap in interpretation. Advances in this direction will be critical for the routine incorporation of genome sequence information to support health care decisions.

Yinyu Ye,

Professor of Management Science and Engineering

Yinyu Ye is currently the K.T. Li Chair Professor of Engineering in the Department of Management Science and Engineering and Institute of Computational & Mathematical Engineering, Stanford University. He received the B.S. degree in System Engineering from the Huazhong University of Science and Technology, China, and the M.S. and Ph.D. degrees in Engineering-Economic Systems and Operations Research from Stanford University. His current research interests include Continuous and Discrete Optimization, Data Science and Application, Algorithm Design and Analysis, Computational Game/Market Equilibrium, Metric Distance Geometry, Dynamic Resource Allocation, and Stochastic and Robust Decision Making, etc. He has received several academic awards including the 2009 John von Neumann Theory Prize for fundamental sustained contributions to theory in Operations Research and the Management Sciences.

Faculty Vision Talks

RECENT PROGRESS AND EXPLORATIONS OF LINEAR PROGRAMMING ALGORITHMS

We describe recent algorithmic progress on linear programming as well as explorations of alternative algorithms. The topics include strongly polynomial simplex and policy iteration methods for the Markov Decision-Game Process, faster interior-point algorithms for network flows, the first-order potential reduction algorithm for convex optimization, the iteration complexity of cyclic coordinate descent method, and convergence of the multi-block alternating method of multipliers for constrained optimization.

Poster Session Presenter

Adam Backer

ENHANCED DNA IMAGING USING SUPER-RESOLUTION MICROSCOPY AND SIMULTANEOUS SINGLE-MOLECULE ORIENTATION MEASUREMENTS

Single-molecule orientation measurements provide unparalleled insight into a multitude of biological and polymeric systems. We report a simple, high-throughput technique for measuring the azimuthal orientation and rotational dynamics of single fluorescent molecules, which is compatible with localization microscopy. Our method involves modulating the polarization of an excitation laser, and analyzing the corresponding intensities emitted by single dye molecules and their modulation amplitudes. To demonstrate our approach, we use intercalating and groove-binding dyes to obtain super-resolved images of stretched DNA strands through binding-induced turn-on of fluorescence. By combining our image data with thousands of dye molecule orientation measurements, we develop a means of probing the structure of individual DNA strands, while also characterizing dye-DNA interactions. This approach may hold promise as a method for monitoring DNA conformation changes resulting from DNA-binding proteins.

ABOUT THE PRESENTER:

Adam is an ICME PhD candidate currently working in the Moerner Laboratory here at Stanford. His interests include fluorescence microscopy, single-molecule microscopy, and next-generation computational imaging platforms. Over his graduate studies, Adam's work has been supported by the National Defense Science and Engineering Graduate Fellowship (NDSEG), and the Simon's Fellowship. He received a BS in Engineering and Physics in 2008 from Brown University, and as a Craig Fellow, received an MPhil in Engineering from Cambridge University in 2009.

Poster Session Presenter

Aekaansh Verma

AUTOMATED OPTIMIZATION FOR FLOW SIMULATIONS IN CARDIOVASCULAR GEOMETRIES

In recent years, computational techniques have become a vital asset in medical device design and surgical planning. However, due to noisiness or non-availability of gradient information of the objective function, conventional optimization methods are often unsuitable in such problems. Furthermore, as fluid flow simulations are computationally expensive, there is a need for efficient methods that limit number of function evaluations. The objective of this study is to develop an automated framework that couples components that model and analyze parametrized cardiovascular geometries with a suitable optimization algorithm. This includes geometry and mesh generation, flow simulation, cost function calculation and interfacing with the optimization component. The optimization algorithm chosen for this study is the Surrogate Management Framework (SMF), a derivative free optimization technique with an established convergence theory which has been previously used for shape optimization of cardiovascular geometries. To address the issue of improving the fixed budget performance of the method, two key causes of performance deterioration have been identified: 1) ill-conditioning of surrogate for the objective function due to pile-up of previous evaluations in design space; and 2) over-refinement of search grid. Variations of the local search in the SMF algorithm are proposed to mitigate the impacts of these deterrents. These variations are then compared on a representative cardiovascular geometry shape optimization problem subject to a fixed computational budget.

ABOUT THE PRESENTER:

Aekaansh is a second year PhD student in the Mechanical Engineering department and he is a part of the Cardiovascular Biomechanics Computation Lab under the guidance of Dr. Alison Marsden. Aekaansh is interested in topics related to Parallel Computing, Incompressible Flow and Optimization.

Poster Session Presenters

Alfredo Lainez and Luke de Oliveira

RECURRENT CONVOLUTIONAL ARCHITECTURES FOR GENERIC TEXT CLASSIFICATION

Text classification has always been an important topic for both the natural language processing and machine learning communities. In this work, we compare traditionally used methods with simple deep learning architectures that take advantage of pre-trained distributional representations of words in a semantic space, also known as word vectors. We then introduce a novel and more complex hierarchical architecture that makes use of compositional properties of language. In particular, we find a general and performant character-level model that relies on a two level hierarchy and does not require any unsupervised pre-training of word vectors.

ABOUT THE PRESENTERS:

Alfredo Lainez is a 2nd year Master's student in ICME at Stanford. His areas of interest include Deep Learning, especially applied to Natural Language Processing, and Recommendation Systems. Prior to Stanford, he has worked as a software engineer in several companies in Spain. He also developed a recommendation system for Shazam which is currently in use.

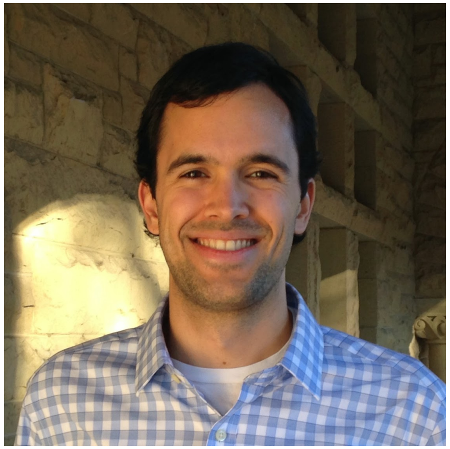
Luke de Oliveira is a second year MS student in the Data Science track. He is interested in Deep Learning, NLP, CV, and Software Architecture, and has a keen interest in making deep learning deployable and scalable. For the past 4 years, he has been a consulting researcher at CERN and has worked with SLAC on building deep learning systems for high energy particle physics at the LHC. Luke also is an industry consultant, helping companies architect, train, and deploy both deep learning and generically data-driven products across a variety of domains.

Poster Session Presenter

Anil Damle

SPARSE REPRESENTATIONS AND FAST ALGORITHMS FOR KOHN-SHAM ORBITALS

Kohn-Sham density functional theory is the most widely used electronic structure theory for molecules and systems in condensed phase. The Kohn-Sham orbitals (a.k.a. Kohn-Sham wavefunctions) are eigenfunctions of the Kohn-Sham Hamiltonian and are generally delocalized, i.e., each orbital has significant magnitude across the entire computational domain. Given a set of Kohn-Sham orbitals from an insulating system, it is often desirable to build a set of localized basis functions for the associated subspace. We present a simple, robust, and parallelizable algorithm to construct a set of (optionally orthogonal) localized basis functions known as the selected columns of the density matrix (SCDM). In addition, we discuss recently developed variants of the SCDM algorithm that drastically reduce the computational cost while maintaining the quality of the basis.



ABOUT THE PRESENTER:

Anil is currently a graduate student in the Institute for Computational & Mathematical Engineering at Stanford University. His research interests include numerical linear algebra, matrix analysis, and fast algorithms for structured matrices. A significant portion his current work is focused on computational quantum chemistry and, more specifically, on the development of efficient algorithms for localizing basis functions resulting from Kohn-Sham density functional theory calculations. Anil is also spends a portion of his time on a wide variety of teaching activities.

Poster Session Presenter

Apaar Sadhwani

(this is joint work with Justin Sirginano and Kay Giesecke)

DEEP LEARNING FOR MORTGAGE RISK

We analyze mortgage risk at loan and pool levels using an unprecedented data set of over 120 million prime and subprime mortgages originated across the United States from 1995 to 2014, which includes the individual characteristics of each loan/borrower, monthly updates on loan performance over the life of a loan, and a number of time-varying economic variables at zip-code level. We develop, estimate, and test dynamic machine learning models for mortgage prepayment, delinquency, and foreclosure. The basic building block is a deep neural network which addresses the significantly nonlinear relationships between explanatory variables and loan performance in the data. Our likelihood estimators, which are implemented using GPU clusters, indicate the importance for mortgage risk of local economic factors such as unemployment and foreclosure rates in a zip code. The out-of-sample predictive performance of our model is a significant improvement over that of other available models.

ABOUT THE PRESENTER:

Apaar Sadhwani is a PhD student in the Department of Management Science and Engineering, advised by Kay Giesecke (MS&E) and Lawrence Wein (GSB). Prior to this, he obtained B.Tech. in Mechanical Engineering from Indian Institute of Technology, Delhi, and M.S. in Operations Research from Stanford University. His research interests span quantitative finance, operations research, machine learning, and computer vision. In his doctoral thesis, he pursues applications of mathematical models and machine learning to biometrics, healthcare, and finance. He has also developed a state of the art deep learning system to detect eye diseases in fundus images. Previously, he has worked in probability theory on simulation of stochastic differential equations.

Poster Session Presenter

Austin Benson

Scalable Methods for Nonnegative Matrix Factorizations of Near-separable Tall-and-skinny Matrices

Numerous algorithms are used for nonnegative matrix factorization under the assumption that the matrix is nearly separable. In this paper, we show how to make these algorithms scalable for data matrices that have many more rows than columns, so-called "tall-and-skinny matrices." One key component to these improved methods is an orthogonal matrix transformation that preserves the separability of the NMF problem. Our final methods need to read the data matrix only once and are suitable for streaming, multi-core, and MapReduce architectures. We demonstrate the efficacy of these algorithms on terabyte-sized matrices from scientific computing and bioinformatics.

ABOUT THE PRESENTER:

Austin Benson is a fourth-year PhD student in ICME and a Stanford Graduate Fellow. Austin's research focuses on matrix-based approaches to network science and data mining. He is advised by Jure Leskovec in the Computer Science department. Prior to Stanford, Austin received a BS in Computer Sciences and Engineering and a BA in Applied Mathematics from UC-Berkeley. Outside of the university, he has interned with Google (three times), Sandia National Laboratories, and HP Labs.

Poster Session Presenter

Brad Nelson

KLEIN BOTTLE MODELS FOR IMAGE PATCHES

In "On the Local Behavior of Spaces of Natural Images", Carlsson et al. introduced a model with the topology of the Klein bottle for the highest density 3 x 3 image patches in natural images. In this poster we show that there is a general class of models with the Klein bottle topology that can be considered for image patches.

ABOUT THE PRESENTER:

Brad Nelson is a second year PhD student at ICME exploring applications of geometry and topology in data analysis. He is also interested in high performance computing and recently taught a short course on scientific computing in Julia. Previously, he studied applied mathematics at Dartmouth and worked as a software developer.

Poster Session Presenter

Carlos Riquelme

ONLINE ACTIVE LINEAR REGRESSION VIA THRESHOLDING

We consider the problem of online active learning to collect data for regression modeling. Specifically, we consider a decision maker that faces a limited experimentation budget but must efficiently learn an underlying linear population model. Our goal is to develop algorithms that provide substantial gains over passive random sampling of observations. To that end, our main contribution is a novel threshold-based algorithm for selection of observations; we characterize its performance and related lower bounds. We also apply our approach successfully to regularized regression. Simulations suggest the algorithm is remarkably robust: it provides significant benefits over passive random sampling even in several real-world datasets that exhibit high nonlinearity and high dimensionality — significantly reducing the mean and variance of the squared error.

ABOUT THE PRESENTER:

Carlos's work is focused around sequential decision making with partial feedback, and reinforcement learning algorithms. In particular, he wants to understand how active learning and adaptive data collection techniques can improve model selection in settings where labeling observations is expensive.

Poster Session Presenters

Celso Ferreira, Sergio Villanueva Maldonado, and Simone Marras

UNDERSTANDING THE PROTECTIVE ROLE OF COASTAL ECOSYSTEMS

Natural habitats such as marshes, mangroves or coral reefs, may protect coastal communities from natural hazards such as storm waves and surge, erosion and tsunamis. An accurate quantification of the protective role of coastal ecosystems remains a standing challenge for the scientific community. The complex interaction between fluid, sediment and vegetation requires a multi-disciplinary approach ranging from fluid dynamics to biomechanics. We propose to address the problem via integrating a combination of high- and low-order numerical models to field data. We are currently developing numerical models specifically targeted to address this phenomena. The inter-comparison of their performance, when validated against field data, can lead to an modeling framework that optimizes the trade-offs between complexity and uncertainty. We expect these models can support policy makers and stakeholders undertake well-informed nature-based decisions for coastal management.

ABOUT THE PRESENTERS:

Dr. Celso Ferreira is a visiting scholar at Stanford University. He is an Assistant Professor of Water Resources Engineering in the Civil, Infrastructure and Environmental Engineering Department of George Mason University. He is also an Associate Researcher at the USGS National Research Program. He has a PhD from Texas A&M University in Civil Engineering. His current research interests are associated to flood hazards and its impacts to civil engineering.

Dr. Sergio Villanueva's work relates to the quantification of the protective services provided by coastal habitats, within the framework of The Natural Capital project. He joined Stanford (Department of Geophysics) in January, after completing a PhD in Environmental Fluid Dynamics at The University of Edinburgh, UK. I hold a MSc in Coastal Engineering from Universidad Nacional Autonoma de Mexico, and a BEng in Mechanical Engineering from Tecnologico de Monterrey, Mexico.

Dr. Simone Marras an aerospace engineer by training, with a M.S. from Politecnico di Milano. He received his doctorate degree in computational mechanics from the Universitat Politecnica de Catalunya, Spain, jointly with the Barcelona Supercomputing Center. Before joining Stanford University as a research associate, he was a National Research Council research associate at the Naval Postgraduate School for two years. His main field of expertise is dynamic diffusion methods for element-based Galerkin methods in computational fluid dynamics.

Poster Session Presenter
Chao Chen

MASSIVELY PARALLEL HIERARCHICAL LINEAR SOLVERS

The poster presents two parallel hierarchical linear solvers. The first solver uses the hierarchical off-diagonal low-rank (HODLR) structure of the matrix, and the algorithm is implemented in Legion, which is novel runtime system targeting at distributed heterogeneous architectures. The second solver uses the low-rank property of the far-field, and the algorithm is implemented in MPI. Both solvers are black-box algorithms and have iso-efficiency P=N/log(N). The poster shows preliminary results on thousands of cores.

ABOUT THE PRESENTER:

My research interests are numerical linear algebra and parallel computing. I am currently developing portable high performance linear solvers targeting at large-scale distributed heterogeneous architectures. My research project is part of the Exascale Computing Engineering Center Predictive Science Academic Alliance Program (PSAAP) at Stanford.

Poster Session Presenter

Dangna Li (Joint work with Henry Li)

PAC: PARTITION-ASSISTED CLUSTERING APPLICATION TO MULTI-SAMPLE HIGH-DIMENSIONAL SINGLE-CELL ANALYSIS

At the frontier of the cytometry field, fluorescence flow cytometry (FFC) is being replaced by mass cytometry (CyTOF), which enables the routine measurement of more than 45 markers per cell. This dramatic increase in dimension opens the door to new scientific inquiries; however, it poses new data analysis challenges. Compared to FFC datasets, CyTOF datasets are not suitable for hand-gating analysis because the human researcher would need several months to gate and annotate the data. In addition, computational methods created for FFC data analysis are no longer efficient due to speed and dimensionality issues. Leading analysis tools for cytometry datasets are SPADE and flowMeans. SPADE was created for CyTOF data analysis; it down-samples the points and outputs a projection/ distortion of a high-dimensional clustering tree structure. flowMeans was created for the accurate analysis of flow cytometry datasets; unfortunately, it cannot handle most CyTOF datasets. These methods are either inaccurate due to data down-sampling or too slow/impossible to run for moderately large CyTOF datasets. We introduce a new data analysis pipeline based on recursive partitioning the data space. The data space is first partitioned using either Bayesian Sequential Partition or Discrepancy Sequential Partition. Afterward, the partitions are used as informed initiation for downstream clustering. We use k-means in our implementation.

ABOUT THE PRESENTER:

I am a third year PhD student from ICME. My interests include statistical machine learning and computational biology.

Poster Session Presenter

Daniele Schiavazzi

ASSIMILATION AND PROPAGATION OF CLINICAL DATA UNCERTAINTY IN CARDIOVASCULAR MODELING

Availability, consistency and variability of clinical measurements affect the predictive performance of cardiovascular models, suggesting the need to include such uncertainty in data assimilation practices. Starting from an analysis of the structural and practical, local and global parameter identifiability in lumped circulation models with applications in pediatric and adult diseases monitoring, we combine multi-level Bayesian estimation, adaptive Markov chain Monte Carlo and multi-resolution uncertainty propagation to learn model parameters and quantify confidence in numerical predictions.

ABOUT THE PRESENTER:

Dr. Schiavazzi is a Post-Doctoral fellow in the department of Pediatrics at Stanford. He graduated with honors from the University of Padova and pursued a Ph.D. degree in applied mathematics from the same University. During his Ph.D. thesis, he was awarded a visiting researcher appointment at the Stanford UQ Lab. He is the recipient of an American Heart Association Postdoctoral Fellowship for 2015-2016 and will be joining the faculty at University of Notre Dame as an Assistant Professor in summer of 2016.

Poster Session Presenter

Danielle Maddix

SPARSE MATRIX VECTOR MULTIPLICATION USING THE MERGE PATH

This is work from my summer internship at NVIDIA in Joe Eaton's group on a method to perform sparse matrix vector multiplication on GPUs efficiently. We present a load-balanced approach with particular application to power law graphs present in social networks. Applications arising from various semi-rings include optimized pagerank algorithms, single source shortest path, and maximum flow.

ABOUT THE PRESENTER:

Danielle is a third year PhD student in ICME. Her research interests include computational fluid dynamics, numerical PDEs, numerical linear algebra, numerical optimization and high performance computing. Danielle works on developing robust and accurate numerical methods for multiphase fluid flow problems.

Poster Session Presenter

Dave Deriso

INVERSE APPROXIMATIONS FOR ELECTROCARDIOGRAPHY

According to the World Health Organization, more people die annually from Cardiovascular diseases (CVD) than from any other cause. In 2012, an estimated 17.5 million people died from CVDs, representing 31% of all global deaths. The electrocardiogram (ECG) has been the primary clinical tool for diagnosing cardiac arrhythmias and other cardiac dysfunctions, however localizing sources of irregular electrical behavior given an observed distribution of time-varying ECG data is an ongoing challenge known as the "inverse problem of electrocardiography." This poster outlines existing numerical optimization approaches for inverse modeling using "activation-based" source models and the "potential-based" source models, preliminary experimental results of an ongoing research project, and future research directions.

ABOUT THE PRESENTER:

Dave teaches Interactive Data Visualization (CME151) at Stanford University's Department of Computational & Mathematical Engineering. Before that he was a Marie Cure fellow at Oxford, NSF undergraduate researcher at UCSD, and blogger for Nature.

Poster Session Presenter

Dustin Gerrard

TOPOLOGY OPTIMIZATION OF THERMO-ELASTICALLY DAMPED MEMS RESONATORS

Microelectromechanical systems (MEMS) resonators are widely used in applications such as inertial navigation and electronic timing. The major energy loss mode in many of these resonators is thermo-elastic dissipation (TED), with air damping, anchor loss, and surface loss as secondary energy loss modes. Thus by using finite element (FE) to mathematically optimize the resonator geometry for TED quality factor (QTED) we are able to increase QTOTAL. We use FE models in the software COMSOL that are able to accurately model TED in resonators. The eigenfrequency of the system is a complex number which provides information on the natural frequency (f) and the Q of the resonator. We calculate the sensitivity of this complex eigenfrequency with respect to FE mesh elements by extracting linear mass-spring-damper matrices from the system and calculating element sensitivity using an expression that does not require re-solving the FE problem. By doing this we are able to obtain sensitivities several orders of magnitude faster than re-solving, and we are the first to demonstrate that this can be done for damped systems. The element sensitivities are used to perform topology optimization on the device via method of moving asymptotes (MMA) which results in resonators with highly unique topologies and improved performance specifications.

ABOUT THE PRESENTER:

Dustin Gerrard is a fourth year mechanical engineering PhD student in the Microstructures and Sensors Lab at Stanford advised by Tom Kenny. Dustin earned a double BS degree in mechanical engineering and mathematics in 2012 from Brigham Young University. He also earned master's degrees from Stanford University in Mechanical Engineering and ICME in 2013 and 2015 respectively.

Poster Session Presenter
Eileen Martin

DIRT CHEAP SURVEYS: NEAR-SURFACE MONITORING WITH AMBIENT SEISMIC NOISE COLLECTED BY DAS

Distributed acoustic sensing (DAS) is a technology which re-purposes a standard fiber optic cable as a low-cost seismic monitoring device. We are particularly interested in developing a network to continuously monitor permafrost melt. DAS can easily collect huge amounts of data over a long distance, but that data is challenging to utilize due to a low SNR, and insensitivity to waves at certain frequencies and incoming angles. Following a successful small-scale pilot test in 2014, a 640 m trenched DAS array was installed alongside a road in a patchy permafrost zone in Fairbanks, AK. We used this array to monitor the speed of waves in the near subsurface using only ambient seismic noise generated primarily by traffic. We show how traffic noise can generate coherent artifacts, and strategies we developed to mitigate the effect of these artifacts on wave speed estimates. Our wave speed estimates are compared with an electrical resistivity tomography study of the survey region.

ABOUT THE PRESENTER:

Eileen Martin is a fourth year Ph.D. student in ICME with B.S. degrees in mathematics and computational physics. She is advised by Biondo Biondi in the geophysics department. Her research is primarily motivated by seismic imaging applications, including some work on fast algorithms. She is supported by the Department of Energy Computational Science Graduate Fellowship, and as part of this program she spent summer 2014 working at Lawrence Livermore National Lab. She interned at Shell Projects & Technology in summer 2015.

Poster Session Presenter

Fayadhoi Ibrahima

AN EFFICIENT DISTRIBUTION METHOD FOR NONLINEAR TRANSPORT PROBLEMS IN HIGHLY HETEROGENEOUS MULTIDIMENSIONAL STOCHASTIC POROUS MEDIA

Because geophysical data is inexorably sparse and incomplete, stochastic treatments of simulated responses are crucial to explore possible scenarios and assess risks in subsurface problems. In particular, nonlinear two-phase flows in porous media are essential, yet challenging, in reservoir simulation and hydrology. Adding highly heterogeneous and uncertain input, such as the permeability and porosity fields, transforms the estimation of the flow response into a tough stochastic problem for which computationally expensive Monte Carlo (MC) simulations remain the preferred option. We propose an alternative approach to evaluate the probability distribution of the (water) saturation for the stochastic Buckley-Leverett problem when the probability distributions of the permeability and porosity fields are available. We give a computationally efficient and numerically accurate method to estimate the one-point probability density (PDF) and cumulative distribution functions (CDF) of the (water) saturation. This surrogate distribution method draws inspiration from a Lagrangian approach of the stochastic transport problem and expresses the saturation PDF and CDF essentially in terms of a deterministic mapping and the distribution and statistics of scalar random fields. We provide examples and comparisons with MC simulations to illustrate the performance of the method.

ABOUT THE PRESENTER:

Fayadhoi is a final year PhD candidate in the Institute for Computational & Mathematical Engineering (ICME) working under the supervision of professor Hamdi Tchelepi in the ERE department on uncertainty propagation of multiphase flow in highly heterogeneous porous media and in multiple dimensions. This is of interest in oil recovery, contaminant spreading or flow in biological tissue for instance. He is a math lover and especially enjoys numerical analysis and applied probability and statistics. Prior to starting the PhD, he obtained master degrees in applied mathematics at the University Pierre and Marie Curie (UPMC), the Ecole Centrale Paris (ECP) [both in France] and Stanford University. He was born and raised in (southern) France, but his parents are from the Comoros Islands (very small islands in the Indian ocean). He enjoys traveling, teaching and running.

Poster Session Presenter
Gabriel Maher

CARDIOVASCULAR EDGE DETECTION FOR EFFICIENT SEGMENTATION FOR PATIENT-SPECIFIC MODELING

To study the effectiveness of cardiovascular simulations it is necessary to perform patient-specific simulations on large patient cohorts. A bottleneck that prevents performing simulations at a large scale is the construction of accurate 3D patient-specific cardiovascular models. Currently researchers must manually construct 3D models using CAD software. In this work, a neural network based approach that forms a step towards automating the construction of accurate 3D cardiovascular models is proposed.

ABOUT THE PRESENTER:

Gabriel Maher is currently a PhD student at the Institute for Computational & Mathematical Engineering at Stanford University. For his research Gabriel is investigating the use of deep learning techniques for cardiovascular segmentation of medical volumetric scan data at the Cardiovascular Biomechanics Computation lab.

Poster Session Presenter

Gianluca Geraci

A MULTIFIDELITY CONTROL VARIATE APPROACH FOR THE MULTILEVEL MONTE CARLO TECHNIQUE

In the last few years efficient algorithms have been proposed for the propagation of uncertainties through numerical codes. The propagation of uncertainties for real case applications remains challenging due to the presence of discontinuous responses, high-gradients or in general high non-linearity. Very often all these features are associated with a high dimensionality of the parameter space. In such situations it appears difficult to efficiently apply established deterministic approaches for UQ as, for instance, stochastic collocation or generalized Polynomial Chaos approaches. More recently, statistical sampling methods have (re)gained popularity. In particular, the multilevel Monte Carlo method (MLMC) has emerged as a novel technique able to retain the robustness of the Monte Carlo method, but increasing also its efficiency. The MLMC is a good candidate for many applications which actually are intractable by other UQ methods, however the number of simulations required remains high. This poster introduces a first effort to reduce the computational burden through the use of a variety of hierarchies of models.

ABOUT THE PRESENTER:

Gianluca Geraci is a Postdoctoral Fellow in the Flow Physics and Computational Engineering group (Stanford University). He received his PhD in Applied Mathematics and Scientific Computing in 2013 from the French Institute for Research in Computer Science and Automation (INRIA) and University of Bordeaux, France. In 2010 he received a Master degree in Aeronautical Engineering from the Politecnico di Milano, Italy with a major in aerodynamics. His current research interests are in the area of Uncertainty Quantification for fluid flow problems.

Poster Session Presenter

Henry Ehrenberg

(joint work with Alex Ratner, Chris De Sa, Professor Chris Ré, and other members of HAZY)

DATA PROGRAMMING WITH DDLITE

Given the success of automatic feature generation methods, learning-based approaches for building knowledge-bases have recently shifted focus from manual feature engineering to obtaining large volumes of labels for training data. Rather than manual creating a highly curated set of labels, we introduce a new approach called data programming, in which users generate large amounts of noisy training labels with simple, heuristic programs. Our work also addresses how best to model and denoise this noisy evidence in learning representations over the features. Using this technique, we are able to construct high-quality knowledge-base systems much more rapidly than before. Since the ability to quickly prototype, evaluate, and debug these programs is a key component of this paradigm, we also introduce DeepDive Lite, an interactive development framework for data programming.

ABOUT THE PRESENTER:

Henry is a first year ICME masters student on the Data Science track working with Chris Ré and the Hazy Research group in the Computer Science department. His interests include machine learning, scalable algorithms, and expensive ice cream.

Poster Session Presenter

Hongyang Zhang

APPROXIMATE PERSONALIZED PAGERANK ON DYNAMIC GRAPHS

We propose and analyze two algorithms for maintaining approximate Personalized PageRank (PPR) vectors on a dynamic graph, where edges are added or deleted. Our algorithms are natural dynamic versions of two known local variations of power iteration. One, Forward Push, propagates probability mass forwards along edges from a source node, while the other, Reverse Push, propagates local changes backwards along edges from a target. In both variations, we maintain an invariant between two vectors, and when an edge is updated, our algorithm first modifies the vectors to restore the invariant, then performs any needed local push operations to restore accuracy. For Reverse Push, we prove that for an arbitrary directed graph in a random edge model, or for an arbitrary undirected graph, given a uniformly random target node t, the cost to maintain a PPR vector to t of additive error Îµ as k edges are updated is O(k + d/Îµ), where d is the average degree of the graph. This is O(1) work per update, plus the cost of computing a reverse vector once on a static graph. For Forward Push, we show that on an arbitrary undirected graph, given a uniformly random start node s, the cost to maintain a PPR vector from s of degree-normalized error Îµ as k edges are updated is O(k + 1/Îµ), which is again O(1) per update plus the cost of computing a PPR vector once on a static graph.

ABOUT THE PRESENTER:

Hongyang is a PhD student in the theory group at Computer Science, where he is advised by Ashish Goel. He is interested in the design and analysis of algorithms. Hongyang has been working on scalable algorithms and data structures on large graphs at Stanford. He is also interested in optimization. Before coming to Stanford, he dabbed at game theory.

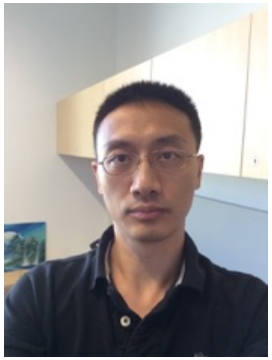
Poster Session Presenter

Hongzhi Lan

SIMVASCULAR: AN OPEN SOURCE PIPELINE FOR IMAGE-BASED CARDIOVASCULAR SIMULATION

SimVascular (www.simvascular.org) is currently the only fully open source software package that provides a complete pipeline from medical image based modeling to patient specific blood flow simulation and analysis. It was initially released in 2007 and has contributed to numerous advances in fundamental hemodynamics research, surgical planning, and medical device design. However, early versions had several major barriers preventing wider adoption by new users, large-scale application in clinical and research studies, and educational access. In the past years, SimVascular has made significant progress by integrating open source alternatives for the expensive commercial libraries previously required for anatomic modeling, mesh generation and the linear solver. In addition, it simplified the across-platform compilation process, improved the graphical user interface and launched a comprehensive documentation website. Many enhancements and new features have been incorporated for the whole pipeline, such as 3D segmentation, Boolean operation for discrete triangulated surfaces, and multi-scale coupling for closed loop boundary conditions. In this poster we will briefly overview the modeling/simulation pipeline and advances of the new SimVascular.

ABOUT THE PRESENTER:



Hongzhi Lan received his PhD in Biomedical Engineering from Tulane University. Currently he is a postdoctoral scholar in Dr. Marsden's Cardiovascular Biomechanics Computation Lab at Stanford University and focuses on the SimVascular open source project, flow solver and cardiovascular blood flow simulation. Prior to PhD, he received his bachelor's degree in Mechanical Engineering from Tsinghua University in China.

Poster Session Presenters

Jiyan Yang and Peng Xu

SUB-SAMPLED NEWTON METHODS WITH NON-UNIFORM SAMPLING

We consider the problem of finding the minimizer of a convex function F(x) = sum_{i=1}^n f_i(w) + R(w) where a low-rank factorization of the Hessian of f_i(w), H_i(w) is readily available. We propose randomized Newton-type algorithms that exploit non-uniform sub-sampling of H_i's, as well as inexact updates, as means to reduce the computational complexity. Two non-uniform sampling distributions based on block norm squares and block partial leverage scores are considered. Under certain assumptions, we show that our algorithms inherit a linear-quadratic convergence rate in w and achieve a lower computational complexity compared to similar existing methods. In addition, we show that our algorithms exhibit more robustness and better dependence on problem specific quantities, such as the condition number. We numerically demonstrate the advantages of our algorithms in recovering the minimizer with generalized linear models (GLMs) on several real datasets.

ABOUT THE PRESENTERS:

Jiyan Yang received the B.S. degree in mathematics from Nanjing University, Nanjing, China, in 2011 and the M.S. degree in computational mathematics from Stanford University, Stanford, CA, USA, in 2013, where he is currently working toward the Ph.D. degree in computational mathematics. His research interests include numerical linear algebra and optimization, large-scale machine learning, and randomized algorithms.

Peng Xu is currently a PhD candidate in computational mathematics at Stanford University advised by Michael Mahoney and Chris Re. My research interests are in randomized linear algebra, machine learning theory and optimization.

Poster Session Presenter

Joongyeub Yeo

RISK CONTROL OF MEAN-REVERSION TIME IN STATISTICAL ARBITRAGE

We propose a new online algorithm that reduces the risk on mean-reversions in statistical arbitrage. Stocks with faster mean-reverting speed are selected to form portfolios, from the observation that each residual carries different reliability in terms of mean-reversions strategy. We also restrict trading activity if the estimation error for the signal is large. With real data, we show that this control method significantly upgrades the performance and the reliability of mean-reversion based strategy. This procedure has more benefits than mere testing the hypothesis of mean reversion, since it directly seeks for faster mean-reversion which minimizes the crucial risk of the strategy. We also deal with the dynamic asset allocation problem in the market- and dollar- neutral strategy, and present its transformation to a non-linear constrained optimization, which can be solved numerically with ease. It has been found that the performance of this investment framework is robust to market fluctuations. Lastly, we provides an answer to a puzzled relation among the number of factors, lengths of estimation window, and transaction costs, which are major parameters that have direct impacts on the strategy. Our study further explores variations from other components, such as regimes, portfolio sizes, and capitalization.

ABOUT THE PRESENTER:

Joongyeub is a PhD candidate of ICME, advised by Prof. George Papanicolaou. His research includes factor models, random matrix theory, and computational finance. He received MS degree in Financial Mathematics at Stanford, and BS/MS degree in Physics at Seoul National University, Korea. For fun, he enjoys basketball, baduk, and photography.

Poster Session Presenter

Justin Tran

A FRAMEWORK FOR AUTOMATED TUNING AND UNCERTAINTY QUANTIFICATION IN MULTISCALE CORONARY FLOW SIMULATIONS

Computational simulations of coronary flow can provide non-invasive information on hemodynamics that can aid in clinical treatment planning and disease progression research. In this study, patient-specific geometries are constructed and combined with finite element flow simulations using the open source software SimVascular. Lumped parameter networks (LPNs), consisting of circuit representations of hemodynamic behavior, can be used as coupled boundary conditions for the flow solver to model the global hemodynamic behavior of the vasculature downstream of the 3D model. There are, however, hurdles in applying this to model large numbers of patients. Users previously had to manually tune the parameters of the LPN to match patient targets, a laborious and time consuming process. Additionally, results are typically reported as deterministic without any uncertainty, even though the clinical, physiologic, and image data used in the model have uncertainty associated with their measurement. To address these concerns, we propose a framework to automate the parameter estimation process, and provide tools to perform uncertainty quantification. This Bayesian framework utilizes parallel Markov Chain Monte Carlo to iteratively sample parameter values that produce results consistent with the patient's data, yet also respects their associated uncertainty. These parameter samples can then be used to quantify confidence bounds on 3D simulation results. Having the capability to perform these functions within the same framework significantly improves repeatability of the modeling process, enables simulations on larger cohorts of patients, and produces confidence intervals on simulation results for the first time.

ABOUT THE PRESENTER:

My name is Justin Tran, and I am a PhD student in Mechanical Engineering in my 3rd year. My research interests include uncertainty quantification and fluid-structure interaction in cardiovascular flow simulations. In my spare time, I like to nap, browse reddit, and play video games on my smartphone. I have two pet chinchillas and they are really cute.

Poster Session Presenter

Karianne Bergen

UNSUPERVISED APPROACHES FOR POST-PROCESSING IN COMPUTATIONALLY EFFICIENT WAVEFORM-SIMILARITY-BASED EARTHQUAKE DETECTION

Seismic sensors collect massive quantities of data that contain a wealth of information about processes within the earth. Seismologists are increasingly adopting data mining and machine learning techniques to identify previously unknown earthquakes in large seismic data sets. New scalable earthquake detection methods, such a Fingerprint and Similarity Thresholding (FAST), use waveform similarity to detect earthquakes (Yoon et al. 2015). FAST leverages locality sensitive hashing (LSH), a data mining technique for efficiently identifying similar items in large data sets, to detect similar waveforms without templates. As algorithms make it possible to process larger data sets, manual inspection of the detection results will become infeasible, and new tools will be required to process the output of similarity-based detection. We propose extensions to FAST for automating the removal of false alarms and improving the detection sensitivity.

ABOUT THE PRESENTER:

Karianne Bergen is a Ph.D. student in Computational & Mathematical Engineering at Stanford University. She holds a B.Sc. in Applied Mathematics from Brown University and a M.Sc. in Computational and Mathematical Engineering from Stanford University. Prior to starting her graduate studies, Karianne was an assistant technical staff member in the Biological and Chemical Defense Systems group at MIT-Lincoln Laboratory.

Poster Session Presenter

Lan Huong Nguyen

METHODS FOR DIFFERENTIAL ABUNDANCE ESTIMATION FOR MICROBIOME DATA

The field of microbial ecology has undergone a major transformation with the advent of highly parallel next generation sequencing (NGS) technologies. The advances have facilitated the discovery of many microbial species. These improvements are of course accompanied by many challenges. DNA sequencing data now consists of discrete, highly skewed counts of sequence reads. Thus, the matured statistical methods designed for continuous microarray data are no longer applicable. We propose a new framework based on a Zero-Inflated Negative Binomial model. The method makes use of the existing normalization and regularization tools to correct for uneven sampling depths and help alleviate over-parametrization, which are issues shared by both microbial and RNA-seq data. We compare our method with the existing DE RNA-seq packages, edgeR and DESeq2 and another software developed specifically for microbiome data, metagenomeSeq, which is based on a Zero-Inflated-Gaussian model.

ABOUT THE PRESENTER:

Lan is a PhD student in Computational Mathematics advised by Prof. Susan Holmes. Her research focus is statistics in application to biology. In particular she works on developing statistical and computational methods for analyzing microbiome data.

Poster Session Presenter

Leopold Cambier

FAST, AN SDDP TOOLBOX FOR MATLAB

Multistage and Nested Stochastic Linear Programs are challenging problems to solve because of the potentially exponential dimension of the search space. The SDDP (Stochastic Dual Dynamic Programming) algorithm provides an efficient and scalable way to solve these problems. Unfortunately, implementation can be quite complex. In this work, we developed FAST (Finally An SDDP Toolbox), an open-source SDDP toolbox in Matlab. FAST handles both the modeling and the algorithmic part. Subproblems are then solved using any commercial or free solver. Using such an algorithm allows us to solve problems potentially involving billions of variables efficiently and with as few as 20 lines of code.

ABOUT THE PRESENTER:

Leopold is currently a 1st year Ph.D. student at ICME. He graduated in 2015 with a Master from Universite Catholique de Louvain (Belgium) where he studied Mathematical Engineering, with an emphasis on optimization and linear algebra. In his Master's thesis, he tackled the problem of Robust Low-Rank Matrix Completion using Riemannian Optimization. On the side, he also started developing FAST for other projects, mostly in Economics.

Poster Session Presenter

Lluis Jofre

TOWARD SIMULTANEOUS EXECUTION OF ENSEMBLE COMPUTATIONS: EXPLORATORY ANALYSIS ON ELLIPTIC PDES

As computing power continues to increase, the engineering community has focused more attention on developing computational tools that will facilitate not only the analysis and understanding of physical phenomena, but that will also play an important role in the design process of industrial applications. This typically involves massive number of samples in order to compute quantities of interest. Fortunately, upcoming superscale computers will significantly make its contribution by presenting 1-10k times increased node-level floating-point capacity. Since data management seems to be one of the a priori bottlenecks for scalability, this work aims at exploring the concept of data-shared strategies in order to efficiently execute multiple samples simultaneously. In many application codes the time spent in the linear solver represents over 50% of the overall cost. Therefore, the focus is placed on analyzing elliptic PDEs. In particular, the PSAAP II overarching problem -- radiated turbulent particle-laden flow -- serves as a test case to demonstrate that the difference among a set of similar samples is reduced as data is coarsened in a hierarchical representation. This result suggests that parts of data within an ensemble of computations can be potentially shared, and thus, encourages continuing exploring akin strategies.

ABOUT THE PRESENTER:

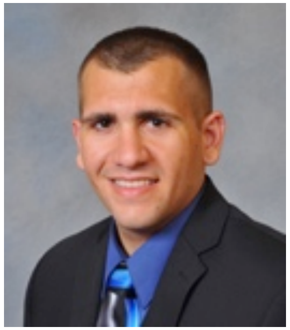
Lluis Jofre is a postdoctoral researcher in the Center for Turbulence Research at Stanford University. He graduated from Polytechnic University of Catalonia - BarcelonaTech (Spain) in conjunction with KTH - Royal Institute of Technology (Sweden), and obtained a PhD with honors in mechanical engineering from the same University. His main research interests are advanced numerical methods for multiphysics applications, modeling and computational studies of turbulent multiphase flow, and high-performance computing. He is part of the Stanford's PSAAP II Exascale Center working on Predictive Simulations of Particle-Laden Turbulence in a Radiation Environment.

Poster Session Presenter
Matthew J. Zahr

EFFICIENT PDE-CONSTRAINED OPTIMIZATION USING ADAPTIVE MODEL REDUCTION

Optimization problems constrained by partial differential equations are ubiquitous in modern science and engineering. They play a central role in optimal design and control of multiphysics systems, as well as nondestructive evaluation and execution, and inverse problems. Methods to solve these optimization problems rely on, potentially many, numerical solutions of the underlying equations. For complicated physical interactions taking place on complex domains, these solutions will be computationally expensive in terms of both time and resources to obtain, rending the optimization procedure difficult or intractable. This situation is exacerbated when a risk-averse solution is sought that accounts for the uncertainties inherent in any physical system. This work introduces a framework for accelerating optimization problems governed by deterministic and stochastic partial differential equations by leveraging adaptive sparse grids and model reduction. Adaptive sparse grids perform efficient integration approximation in a high-dimensional stochastic space and reduced-order models reduce the cost of objective and gradient evaluations by decreasing the complexity of primal and dual PDE solves. A globally convergent trust-region framework accounts for these two levels of inexactness in the objective and gradient.

ABOUT THE PRESENTER:



Matthew is a PhD candidate in Computational & Mathematical Engineering at Stanford University, with minors in Mechanical Engineering and Aeronautics/Astronautics, under the advisement of Prof Charbel Farhat. He received his BSc in Civil and Environmental Engineering, with a minor in Mathematics, from UC Berkeley in 2011. In October 2016, he will be the Luis W. Alvarez Postdoctoral Fellow at Lawrence Berkeley National Laboratory.

Poster Session Presenter

Matthias Cremon

SPE COMPARATIVE SOLUTION PROJECT 11: OPTIMIZATION OF NET PRESENT VALUE UNDER UNCERTAINTY

SPE's Comparative Solution Projects aim to compare numerical results from both industrial and academic partners on critical topics. SPE10, the latest SPE comparative solution project released in 2001, focused on upgridding and upscaling approaches for a waterflood case. We are developing SPE11. Its primary focus is optimization of Net Present Value (NPV) under uncertainty, which has turned into a central topic following the increase in computational capabilities. To facilitate a common effort in studying and solving this problem, two main steps are needed: design a multi-tier benchmark and provide a suite of reference solutions. We use real reservoir data from a major company, namely a corner-point grid, well logs and seismic data, and from this we generate an ensemble of realizations that are consistent with all the data available. We then run multiphase flow simulations on this ensemble using several commercial simulators, and aim at optimizing the Net Present Value. We are currently focusing on designing the optimization problem, testing it and further developing a distributed stochastic optimization code.

ABOUT THE PRESENTER:

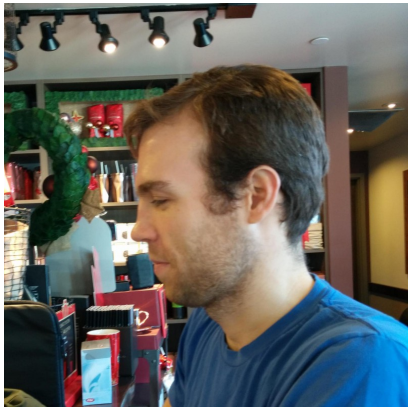
My current interests are centered on oil production optimization under geological uncertainty. We are trying to develop a new benchmark to compare different routines of optimization under uncertainty, using both synthetic and real reservoir data (in the spirit of SPE's Comparative Solutions Project). This project mainly involves geostatistics, global and gradient-free optimization and reservoir simulation, and is conducted as part of a Master of Science in Computational & Mathematical Engineering at Stanford University. Prior to that work, I studied Fluid Mechanics in France at the INP-ENSEEIHT where I obtained a Diplôme d'Ingénieur, after attending a Classe Préparatoire for 2 years.

Poster Session Presenter

Nick Henderson

THREAD DIVERGENCE IN A GPU MONTE CARLO RADIOTHERAPY SIMULATOR

We present a high performance Monte Carlo simulator for radiation therapy using CUDA and NVIDIA GPUs. The code is based on the core algorithm and physics models used in Geant4, a particle physics simulation toolkit. We show that results from the GPU code matches output from Geant4 in several test configurations. In runtime tests we find that the GPU code is between 185 and 250 times faster than single-CPU Geant4 in different test setups. Our continued research focuses on improving the performance of the GPU code. We find that thread divergence due to the stochastic nature of the simulation is the key limitation. We use a simplified surrogate Monte Carlo simulation to show that speedup may be obtained via a particular sorting strategy.



ABOUT THE PRESENTER:

Nick Henderson is a Research Associate with Institute for Computational & Mathematical Engineering (ICME) and the CUDA Center of Excellence at Stanford University. He received a PhD at ICME in 2012. His interests involve high performance physics simulation using GPUs, numerical optimization, and scientific computing in general.

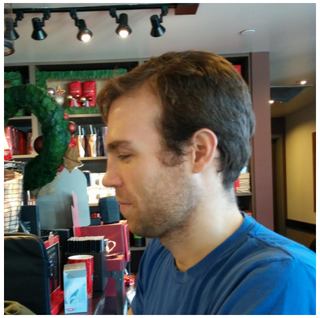
Poster Session Presenter

Nick Henderson

(joint work with Ding Ma, Yuekai Sun, and Professor Michael Saunders)

CONSERVATION ANALYSIS OF GENOME-SCALE BIOCHEMICAL NETWORKS

To model biochemical networks, systems biologists are generating increasingly large stoichiometric matrices, whose rows and columns correspond to chemical species and chemical reactions. An important step toward evaluating drug targets and analyzing transient behavior of the network is called conservation analysis, which reduces to finding the rank of the matrix and the nullspace of the transpose. SVD is not practical, but with care, sparse QR or sparse LU factors can serve both purposes. On some large authentic examples, the sparse QR package of Tim Davis has proved remarkably efficient (with Q stored in sparse product form). However, the QR factors are sometimes significantly less sparse than the original matrix. As an alternative, we find that LUSOL's LDU factors can be significantly more sparse than QR, but also significantly more expensive to compute. We therefore study and present results for a few alternative uses of LUSOL.



ABOUT THE PRESENTER:

Nick Henderson is a Research Associate with Institute for Computational & Mathematical Engineering (ICME) and the CUDA Center of Excellence at Stanford University. He received a PhD at ICME in 2012. His interests involve high performance physics simulation using GPUs, numerical optimization, and scientific computing in general.

Poster Session Presenter

Raphael Townshend

DEEP LEARNING ON PROTEIN COMPLEXES

The dramatic rise in the number of macromolecules that have been structurally characterized has opened up the field of structural biology as a whole new area of application for deep learning algorithms. As structure is fundamental to determining function, such algorithms allow us to explore the inner workings of the cell in completely novel ways. In this work, we tackle the problem of protein complex prediction through the use of siamese networks. This method allows us to better understand the details of interactions between proteins as well as predict new interactions.

ABOUT THE PRESENTER:

Raphael Townshend is a 2nd year PhD student in Artificial Intelligence at Stanford in Professor Ron Dror's lab, applying machine learning methods to problems in Structural Biology. He has experience working in the fields of Computer Vision, Computer Architecture, Genomics, and Statistical Inference.

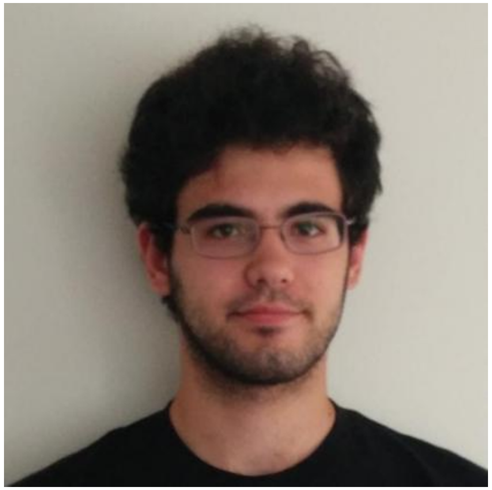
Poster Session Presenter

Ron Estrin

FROM QR FACTORIZATION TO WIRELESS COMMUNICATION

Despite the numerous applications of lattice reduction (e.g., in MIMO systems, algebraic number theory and lattice-based cryptography) and the close connection between the Lenstra-Lenstra-Lovasz (LLL) algorithm and column-pivoted QR factorizations (through generalizing from permutations to unimodular transforms), the well-established techniques for exploiting level 3 BLAS within Householder QR have not yet been extended. We therefore propose a novel right-looking variant of a Householder-based LLL which both accumulates Householder transformations where possible and avoids redundant applications when the Lovasz condition requires column swaps. We provide benchmarks of the new scheme, a tree-based, recursive extension, against the corresponding implementations from the popular NTL and FPLLL libraries.

ABOUT THE PRESENTER:



Ron Estrin is a second year Ph.D. student in ICME, working with Professors Jack Poulson and Michael Saunders on projects utilizing high performance numerical linear algebra. Prior to studying at Stanford, he received his B.S. in Mathematics and Computer Science from the University of British Columbia. He is interested in numerical linear algebra and Krylov subspace methods, and in particular their applications to fields such as optimization and number theory. Outside of academics, he enjoys practicing martial arts and playing tennis.

Poster Session Presenter

Sebastien Dubois

LEARNING FROM ELECTRONIC HEALTH RECORDS TO FIND UNDIAGNOSED CASES OF FAMILIAL HYPERCHOLESTEROLEMIA (FH)

Familial Hypercholesterolemia (FH) is known as one of the most common and under-diagnosed, life-threatening genetic disorders around the world. While it can be effectively treated, more than 90% of individuals with FH are not accurately diagnosed in the US. This study uses Electronic Health Records (EHR) of over 2 million patients at Stanford Hospital, with only 89 identified cases of FH, to design a tool able to flag potential undiagnosed patients. Building upon the traditional questionnaire used by clinicians, we show how we overcome the complexity of EHR data to learn a new scoring method, through features inspired by this test and a Logistic Regression model. We finally compare with other Machine Learning approaches to conclude that we achieve similar accuracy while being consistent with domain knowledge, thus providing better candidates of undiagnosed FH.

ABOUT THE PRESENTER:

Sebastien Dubois is a Master's student in Data Science at the Institute for Computational & Mathematical Engineering at Stanford University. He is a research assistant in Professor Nigam Shah's group, where his work focuses on automating machine learning pipelines and applying such techniques to address problems in healthcare. Before coming to Stanford, he studied Computer Science and Applied Mathematics at the Ecole Polytechnique in France, graduating with a Diplome d'Ingenieur. There, he received the Research Internship Award in Computer Science for work performed during an internship at MIT with Prof. Kalyan Veeramachaneni.

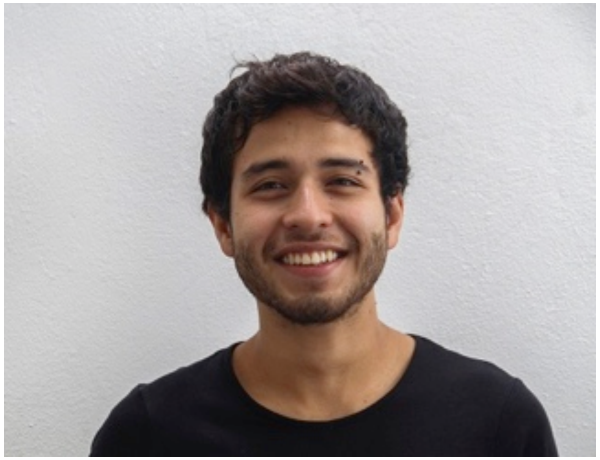
Poster Session Presenter
Sergio Camelo

NEAREST NEIGHBORS METHODS FOR SUPPORT VECTOR MACHINES

A key issue in the practical applicability of the support vector machine methodology is the identification of the support vectors in very large data sets. In this work we propose methods based on sampling and nearest neighbors, that allow for an efficient implementation of an approximate solution to the classification problem and, at least in some problems, will help in identifying a significant fraction of the support vectors in large data sets at low cost. The performance of the proposed method is evaluated in different examples and some of its theoretical properties are discussed.

ABOUT THE PRESENTER:

Sergio is a first year PhD student in ICME. Sergio's interests cover the broad fields of convex optimization and statistics, particularly machine learning and approximation algorithms.



Poster Session Presenter

Sven Schmit

(joint work with Carlos Riquelme, Vijay Kamble and Ramesh Johari)

HUMAN INTERACTION WITH RECOMMENDATION SYSTEMS

More and more, recommendation systems assist users in making choices. These systems rely on historical user data to provide suggestions. However, the interaction between humans and algorithm is largely ignored in the literature. We propose an explicit, though simplified, model for the interaction between the users and recommendation system, and relate this model to the multi-armed bandit literature. First, we show that this interaction leads to a bias in naive estimators due to selection effects. This bias leads to suboptimal outcomes if left unaddressed. Second, though agents are myopic, agents' heterogeneous preferences lead to sufficient exploration of alternatives. This model provides new and practical insights relevant to a wide range of systems designed to help users make decisions based on user data.

ABOUT THE PRESENTER:

Sven hails from the Netherlands, where he obtained a bachelor's degree in Econometrics and Operations Research. He then obtained a master's degree in Mathematics from the University of Cambridge. Currently, he is a fourth year PhD student at ICME advised by Prof. Johari.

Poster Session Presenter

Timothy Anderson

EFFICIENT BRAIN MRI SEGMENTATION FOR 3D PRINTING APPLICATIONS

Advances in computing power and additive manufacturing (3D printing) have now made possible the efficient simulation, optimization, and replication of patient-specific procedures and prosthetics. One very promising application is generation of patient-specific brain analogues, but the primary issue is in the efficient segmentation of brain imaging data. This study puts form a novel algorithm for brain MRI segmentation, and applies this algorithm to create a 3D printed brain model. The given algorithm exploits structural characteristics of the brain to pre-process the images and combines deterministic and statistical image segmentation techniques to deliver a reliable segmentation routine. Specifically, we iterative employ an active contour model with nonlinear image filtering to perform initial segmentation, use k-means clustering to further segment the image data, and perform morphological post-processing to improve the image quality for 3D printing. This study demonstrates that active contour models are very well- suited for brain MRI segmentation. The pre-processing successfully removed the skull and most of the dura in all images. The active contour successfully captured the outer contour of the brain and k-means efficiently segmented the inner brain folds. Patient- specific brain analogues have many far-reaching applications in neurosurgery, and it is our hope that the algorithm presented in this paper will bring these closer into reach.

ABOUT THE PRESENTER:

I am a fourth year undergraduate in electrical engineering and coterminal masters student in ICME. My research interests are in signal processing and inverse modeling and their applications to problems in neuroengineering and computational neuroscience. A specific area I am interested in is the crossover between image processing and additive manufacturing, and the applications of this area to neurosurgery and neuromechanical modeling. In my free time, I enjoy reading, listening to music, playing the tuba, and collecting rubber ducks.

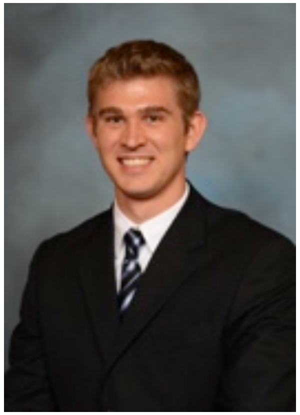
Poster Session Presenter

Victor Minden

(joint work with Anil Damle, Ken L. Ho, and Professor Lexing Ying)

FAST SPATIAL GAUSSIAN PROCESS MAXIMUM LIKELIHOOD ESTIMATION VIA SKELETONIZATION FACTORIZATIONS

Maximum likelihood estimation for parameter-fitting given observations from a kernelized Gaussian process in space is a computationally-demanding task that restricts the use of such methods to moderately-sized datasets. We present a framework for unstructured observations in two spatial dimensions that allows for evaluation of the log-likelihood and its gradient (i.e., the score equations) in O(n^{3/2}) time under certain assumptions, where n is the number of observations. Our method relies on the skeletonization procedure described by Martinsson and Rokhlin (2005) in the form of the recursive skeletonization factorization of Ho and Ying (2015). Combining this with an adaptation of the matrix peeling algorithm of Lin et al. (2011), our method can be used in the context of any first-order optimization routine to quickly and accurately compute maximum-likelihood estimates.



ABOUT THE PRESENTER:

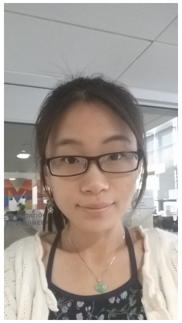
Victor is currently pursuing a Ph.D. in Computational & Mathematical Engineering, where his research interests concern scientific computing and numerical algorithms. His current research interests include fast algorithms for linear algebra problems coming from physical applications such as discretized differential equations, discretized boundary integral equations, or spatial statistics. Most recently, Victor has been working on applying these techniques to parameter estimation problems for kernelized Gaussian processes.

Poster Session Presenter

Xiaotong Suo

TIME SERIES FORECASTING FROM HISTORICAL DATA

Seasonal influenza epidemics have a severe and detrimental impact on society in terms of economic and health costs every year. According to Centers for Disease Control and Prevention, estimated flu-associated death ranges from 3,000 to 49,000 between 1976 and 2006 in the United States. Therefore, it is of great importance for policy makers to have access to accurate and reliable forecasts of flu activity each season. During 2013 - 2014 flu season, CDC hosted "Predict the Influenza season challenge" with the task of predicting key epidemiological measures in real-time throughout the season. Motivated by the challenge and by the losses caused by seasonal influenza every year, we propose a novel framework for predicting epidemic trajectories in real-time through a methodology of weighted smoothing splines and show how to compute these trajectories efficiently. In addition, we provide a posterior distribution over predicted trajectories of the current season. In contrast with previously proposed forecasting frameworks, we only use the final version of surveillance data provided by CDC. Our model is not restricted to seasonal influenza and can be applied to any infectious disease with a history of seasonal outbreaks.



ABOUT THE PRESENTER:

I am a fourth year Ph.D. student in ICME. I am interested in statistical learning and machine learning. My advisor is Robert Tibshirani in the department of Statistics.

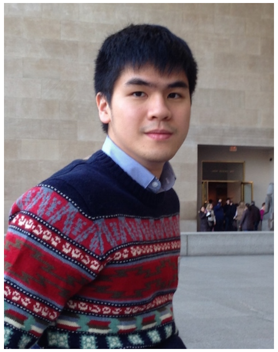
Poster Session Presenter

Yi-Chun Chen

LEARNING DISCRETE BAYESIAN NETWORKS FROM CONTINUOUS DATA

Real data often contains a mixture of discrete and continuous variables, but many Bayesian network structure learning and inference algorithms assume all random variables are discrete. Continuous variables are often discretized, but the choice of discretization policy has significant impact on the accuracy, speed, and interpretability of the resulting models. This poster introduces a principled Bayesian discretization method for continuous variables in Bayesian networks with quadratic complexity instead of the cubic complexity of other standard techniques. Empirical demonstrations show that the proposed method is superior to the state of the art. In addition, this paper shows how to incorporate existing methods into the structure learning process to discretize all continuous variables and simultaneously learn Bayesian network structures. This work is under review by Machine Learning (Springer journal) and is also available as an arXiv preprint (http://arxiv.org/abs/1512.02406).

ABOUT THE PRESENTER:



Yi-Chun Chen is a graduate student in the Institute for Computational & Mathematical Engineering (ICME) at Stanford University. He is currently in Stanford Intelligent System Lab, and works on machine learning and decision making problems. Prior to joining Stanford, Yi-Chun received his undergraduate degree in physics from National Taiwan University.

Poster Session Presenter

Yinbin Ma

TIME-LAPSE FULL-WAVEFORM INVERSION IN ACOUSTIC MEDIA

We implement multiparameter full-waveform inversion (FWI) in the isotropic acoustic media with the nonlinear conjugate gradient (CG) method. The performance of the FWI is evaluated using different combinations of acoustic parameters, including velocity, density and acoustic impedance. We apply FWI to a time-lapse seismic inverse problem, and show gradient based methods cannot efficiently resolve velocity and density change simultaneously because of the crosstalk between parameters. We show a second order method is promising in time-lapse FWI by applying the approximated inverse of the Hessian to the gradient.

ABOUT THE PRESENTER:

I am a PhD student in the Institute for Computational & Mathematical Engineering (ICME) at Stanford working with Prof. Biondo Biondi. I joined Stanford Exploration Project (SEP) in the summer 2014. My research interests include seismic imaging, numerical optimization, and high-performance computing.

Poster Session Presenter

Yingzhou Li

(joint work with Haizhao Yang, Duke University)

INTERPOLATIVE BUTTERFLY FACTORIZATION

Interpolative butterfly factorization (IBF) is a nearly optimal implementation of several transforms in harmonic analysis, when their explicit formulas satisfy certain analytic properties and the matrix representations of these transforms satisfy a complementary low-rank property. A preliminary interpolative butterfly factorization is constructed based on interpolative low-rank approximations of the complementary low-rank matrix. A novel sweeping matrix compression technique further compresses the preliminary interpolative butterfly factorization via a sequence of structure-preserving low-rank approximations. For an $N\times N$ matrix, the resulting factorization is a product of $O(\log N)$ sparse matrices, each with $O(N)$ non-zero entries. With the efficient algorithm, it can be constructed and applied in $O(N\log N)$ operations. Numerical results are provided to demonstrate effectiveness of the algorithm

ABOUT THE PRESENTER:

Yingzhou Li is a Ph.D. candidate in ICME advised by Lexing Ying. His research interests lie in hierarchical matrix and its applications, fast algorithms for elliptic PDEs, butterfly algorithms for Fourier integral operations, distributed-memory algorithms and computational biology.