12^{th} Annual SIAM Front Range Applied Mathematics Student Conference

March 5, 2016

Registration: 8:30 - 9:00

Morning Session I - Room 4017
9:00 - 11:35

9:00 - 9:20	Luke Nelsen *University of Colorado, Denver*	A Peek at the Absorbing Method: When a Little Laziness Helps
9:25 - 9:45	Axel Brandt *University of Colorado, Denver*	Fixing Congressional Dysfunction with Polynomials
9:50 - 10:10	Lauren Nelsen *University of Denver*	Color-blind index of graphs
10:10 - 10:25	15 Minute Break	
10:25 - 10:45	Eric Sullivan *University of Colorado, Denver*	Color degree sequences: a preliminary report
10:50 - 11:10	Kerrek Stinson *Colorado School of Mines*	A randomized algorithm for enumerating zonotope vertices
11:15 - 11:35	Ikuko Saito *University of Colorado, Colorado Springs*	Traceless Matrices that are not Commutators

Morning Session II - Room 4125
9:00 - 11:35

9:00 - 9:20	Yi Meng *University of Colorado, Denver*	Improving the LAPACK test suite: testing, readability, and scalability.
9:25 - 9:45	Ben Sattelberg *Colorado School of Mines*	Plasma Simulation Using the One- Dimensional Particle-in-Cell Method
9:50 - 10:10	Louie J. Long *University of Colorado, Boulder*	Kinematic Dynamos in Low Rossby and Magnetic Prandtl Number Convection
10:10 - 10:25	15 Minute Break	
10:25 - 10:45	Dalton Anderson *University of Colorado, Boulder*	Controlling Wavebreaking in Viscous Fluid Conduits

10:50 - 11:10	Will Farmer *University of Colorado, Boulder*	Parameterization and Analysis of Viscous Fluid Conduit Edges for Dispersive Hydrodynamics
11:15 - 11:35	Alex Masarie *Colorado State University*	A partial differential equation model for fire resource movement

Morning Session III - Room 4119
9:00 - 11:35

9:00 - 9:20	William Van Horn, Samuel Stanton, & Zane Showalter-Castorena *University of Colorado, Denver*	ICM Problem F: Modeling Refugee Migration
9:25 - 9:45	Scott Wurst, Nicolas Pinkowski, & Joseph Arehart *University of Colorado, Boulder*	ICM Problem F: A Bipartite 'Push-Permit' Model for Refugee Flow
9:50 - 10:10	Nathan Neri *Colorado School of Mines*	Modeling Acute Stage HIV with Homeostasis
10:10 - 10:25	15 Minute Break	
10:25 - 10:45	Rachel Neville *Colorado State University*	Classification of the Linked Twist Map
10:50 - 11:10	Subrata Paul *University of Colorado, Denver*	Development of Polygenic Risk Scores for Vitiligo from a Genome-Wide Association Study
11:15 - 11:35	Harrison Pielke-Lombardo *University of Colorado, Boulder*	Expected fraction of bases in Human Chromosome 21 that may be edited using CRISPR-Cas9

Break: 11:35 - 11:45

Lunch: 11:45 - 12:25, Room 2500

Plenary Address: 12:30 - 1:30, Room 2600

Dr. Chun Liu
Department of Mathematics
Pennsylvania State University

An Energetic Variational Approach for Ionic Fluids and Ion Channels

Group Photographs at 1:30

Afternoon Session I - Room 4017
1:45 - 3:20

1:45 - 2:05	Devon Sigler *University of Colorado, Denver*	Notions of Optimality in Multicriteria Optimization Under Uncertainty
2:10 - 2:30	James Folberth *University of Colorado, Boulder*	Application of Adjoint Operators in Gradient Computations
2:35 - 2:55	Scott Walsh *University of Colorado, Denver*	Parameter Estimation Guided Experimental Design
3:00 - 3:20	Chad Waddington *Colorado State University*	Geolocation of EM Emitters Using Passive Bistatic Radar

Afternoon Session II - Room 4125
1:45 - 3:20

1:45 - 2:05	Ian Laga *University of Colorado, Boulder*	The Modified Matern Process
2:10 - 2:30	Mengjie Yao *University of Colorado, Denver*	Sparse Canonical Correlation Analysis
2:35 - 2:55	Lawrence Pelo & Sam Loos *University of Colorado, Denver*	A Similar Seasons Model for Predicting Baseball Performance
3:00 - 3:20	Michael Schmidt *Colorado School of Mines*	Novel Lagrangian Methods for Imperfectly Mixed Chemical Reactions

Afternoon Session III - Room 4119
1:45 - 3:45

1:45 - 2:05	Farhad Pourkamali-Anaraki *University of Colorado, Boulder*	Preconditioned Data Sparsification for Big Data with Applications to PCA and K-means
2:10 - 2:30	Paul Diaz *Colorado School of Mines*	Global Sensitivity Metrics from Active Subspaces
2:35 - 2:55	Derek Driggs *University of Colorado, Boulder*	Stable Principal Component Pursuit for the GPU
3:00 - 3:20	Zach Grey *Colorado School of Mines*	2D Viscous Laminar Flow Airfoil Design Applying Principal Component Analysis and Active Subspaces
3:25 - 3:45	Prosper Torsu *University of Wyoming*	Modeling of Uncertainty in Reservoir With a Stochastic Basis Method

Plenary Speaker

12:30 - 1:30, 2600

An Energetic Variational Approach for Ionic Fluids and Ion Channels

Dr. Chun Liu

Department of Mathematics, Pennsylvania State University

The interactions of ions flowing through biological systems have been a central topic in biology for more than 100 years. Flows of ions produce signaling in the nervous system, initiation of contraction in muscle, coordinating the pumping of the heart and regulating the flow of water through kidney and intestine. Ion concentrations inside cells are controlled by ion channel proteins through the lipid membranes. In this talk, a continuum model is derived from the energetic variational approach, which include the coupling between the electrostatic forces, the hydrodynamics, diffusion and crowding (due to the finite size effects). The model provides some basic understanding of some important properties of proteins, such as the ion selectivity and sensor mechanism. Transport of charged particles and ions in biological environments is by nature a multiscale and multiphysics problem. I will also discuss the roles of other important ingredients such as those of general diffusion and also the connection between kinetic description and continuum approaches.

About the Speaker

Dr. Chun Liu is a professor of mathematics at Pennsylvania State University where he has been a faculty member since 1998. Before arriving at Penn State, Dr. Liu received his Ph.D. from Courant Institute of Mathematical Sciences at New York University. He held positions as visiting assistant professor at Carnegie Mellon University and assistant professor at University of Georgia. Dr. Liu's research is focused on partial differential equations, calculus of variations and their applications in complex fluids. His research is highly cited and covers a variety of topics including energetic variational approaches, biology and physiology, liquid crystals, mixtures and interfaces, grain growth, evolution of surfaces, viscoelastic fluids and polymeric materials. Dr. Liu's contributions have included serving as a coordinator for the teaching training program for instructors and TAs in the Department of Mathematics at Penn State, and he was recently the 2014 recipient of the Teresa Cohen Service Award. Dr. Liu serves on several editorial boards including SIAM Journal on Mathematical Analysis.

MORNING SESSION I

A PEEK AT THE ABSORBING METHOD: WHEN A LITTLE LAZINESS HELPS

Luke Nelsen[1]
Advisor: Louis DeBiasio[2]
[1] ***University of Colorado, Denver***
[2] ***Miami University, Ohio***

A 2006 paper of Rödl, Ruciński and Szemerédi used a novel technique which has become known in extremal graph theory as the "absorbing method." The idea is to obtain a nearly spanning result (typically some desired subgraph) and then extend the result to be fully spanning by the use of an absorbing structure. Szemerédi refers to this approach as "laziness." Several papers have employed this laziness in the past decade, including one by the presenter. This talk will present a toy problem for the purpose of gaining insight into this technique. With whatever time remains we will highlight an application of this laziness in the submitted paper, "Monochromatic cycle partitions of graphs with large minimum degree," which can be accessed at http://arxiv.org/pdf/1409.1874.pdf.

FIXING CONGRESSIONAL DYSFUNCTION WITH POLYNOMIALS

Axel Brandt
University of Colorado, Denver

Although a great deal of finger-pointing can be done to cast blame for political gridlock in the US Congress, members of both parties are equally guilty of poor attendance records at congressional committee meetings. For the purposes of this talk, we will assume that the cause of these poor attendance records is due to two of their committees being scheduled to meet at the same time. After modeling committee scheduling as a graph coloring problem, we will discuss a way to ensure that every representative can attend meetings for every committee on which they serve. This approach embeds scheduling conflicts as zeros of a multivariate polynomial and uses the maximum degree of that polynomial to obtain constraints from which a nonzero solution to the polynomial exists; i.e. given sufficient availability, one can schedule committees without any conflicts. Graph coloring results using the same theoretical approach will also be presented.

COLOR BLIND INDEX OF GRAPHS

Lauren M. Nelsen[1]
Advisor: Paul Horn[1]
Collaborators: Jennifer Diemunsch[2]**, Nathan Graber**[2]**, Lucas Kramer**[3]**, Victor Larsen**[4]**, Luke Nelsen**[2]**, Devon Sigler**[2]**, Derrick Stolee**[5]**, and Charlie Suer**[6]
[1] ***University of Denver***
[2] ***University of Colorado, Denver***
[3] ***Bethel College***
[4] ***Kennesaw State University***
[5] ***Iowa State University***
[6] ***University of Louisville***

Let $c : E(G) \to [k]$ be an edge-coloring of a graph G, not necessarily proper. For each vertex v, let $\overline{c}(v) = (a_1, ..., a_k)$, where a_i is the number of edges incident to v with color i. Reorder $\overline{c}(v)$ for every v in G in nonincreasing order to obtain $c^\star(v)$, the color-blind partition of v. When $c^\star$ induces a proper vertex coloring, that is, $c^\star(u) \neq c^\star(v)$ for every edge uv in G, we say that c is color-blind distinguishing. The minimum k for which there exists a color-blind distinguishing edge coloring $c : E(G) \to [k]$ is the color-blind index of G, denoted $\mathrm{dal}(G)$. We introduce some previously known results about the color-blind index of certain graphs, and show that determining the color-blind index is more subtle than previously thought.

COLOR DEGREE SEQUENCES: A PRELIMINARY REPORT

Eric Sullivan
Collaborators: Michael Ferrara, Stephen Hartke, and Luke Nelsen
University of Colorado, Denver

Edge coloring and degree sequences are two fundamental areas of research in graph theory. Graph colorings have numerous applications in scheduling problems, such as register allocation in computer science, and degree sequences have applications in network science, such as reconstructing a social network form survey responses. Our initial research pursues one possible way of combining these two areas. In this talk, we will discuss preliminary results and avenues for continued study.

A RANDOMIZED ALGORITHM FOR ENUMERATING ZONOTOPE VERTICES

Kerrek Stinson[1]
Advisors: Paul Constantine[1] & David Gleich[2]
[1]***Colorado School of Mines***
[2]***Purdue University***

We propose a randomized algorithm for enumerating the vertices of a zonotope, which is a low-dimensional linear projection of a hypercube. The algorithm produces a pair of the zonotope's vertices by sampling a random linear combination of the zonotope generators, where the combination's weights are the signs of the product between the zonotope's generator matrix and random vectors with normally distributed entries. We study the probability of recovering particular vertices and relate it to the vertices' normal cones. This study shows that if we terminate the randomized algorithm before all vertices are recovered, then the convex hull of the resulting vertex set approximates the zonotope. In high dimensions, we expect the enumeration algorithm to be most appropriate as an approximation algorithm—particularly for cases when existing methods are not practical.

TRACELESS MATRICES THAT ARE NOT COMMUTATORS

Ikuko Saito
Advisors: Zachary Mesyan, Gene Abrams, & Sarbarish Chakravarty
University of Colorado, Colorado Springs

By a classical result, for any field F and a positive integer n, a matrix in $\mathbb{M}_n(F)$ is a commutator if only and if it has trace zero. This is no longer true if F is replaced with an arbitrary ring R. But the only known examples of matrices which have trace zero and are not commutators are of the size 2×2. The purpose of this talk is to construct an $n \times n$ matrix for any positive integer $n \geq 2$ which has trace zero but is not a commutator.

MORNING SESSION II

IMPROVING THE LAPACK TEST SUITE: TESTING, READABILITY, AND SCALABILITY.

Yi Meng
University of Colorado, Denver

We present a new design and prototyping for a software infrastructure for testing numerical linear algebra libraries. Our new algorithms uses Level 3 BLAS and therefore scales well to large matrix sizes. We study several variants for checking an LU factorization or a Cholesky factorization. Algorithms and numerical results will be presented.

PLASMA SIMULATION USING THE ONE-DIMENSIONAL PARTICLE-IN-CELL METHOD

Ben Sattelberg
Advisor: Stephen Pankavich
Colorado School of Mines

Plasmas, or high-temperature, low-density ionized gases, underlie a number of different applications, including pulsed power and nuclear fusion, in addition to describing a variety of natural phenomena, such as comet tails and solar wind. Simulation methods for plasmas that calculate individual particle dynamics are computationally intensive and typically produce more information than is necessary. Alternatively, Particle-in-cell (PIC) methods consider representative superparticles on a discrete grid to reduce computational time and smooth out unnecessary detail, while retaining the underlying structure of the plasma. The one-dimensional Vlasov-Poisson system, which serves as an effective model of plasma dynamics, can be solved quickly using PIC methods to obtain qualitatively accurate simulations. In this talk, we will discuss PIC methods for simulating this system of PDEs and, time permitting, their extension to global sensitivity methods for plasma studies.

KINEMATIC DYNAMOS IN LOW ROSSBY AND MAGNETIC PRANDTL NUMBER CONVECTION

Louie J. Long
University of Colorado, Boulder

The Earth's magnetic field is believed to be generated and maintained by the dynamo mechanism.

Because this mechanism is sustained by convectively-driven turbulence in an electrically conducting fluid within the liquid outer core, a vast range of spatio-temporal scales are involved that cannot be resolved using direct numerical simulations of the complete set of governing equations of fluid dynamics. In this study, the onset of dynamo action in a rotating, convecting layer of Oberbeck-Boussinesq fluid for the case of asymptotically low Rossby and magnetic Prandtl numbers is considered. It is believed that dynamo action is a result of compositional and thermal convection present in the liquid outer core due to light constituents in the inner core and thermal buoyancy, respectively. Thermal and compositional convection are investigated by considering Prandtl numbers of $Pr = 0.1, 1$, and 10 over a wide range of Rayleigh numbers, bracketing the cellular regime near the onset of convection to the strongly forced geostrophic turbulence regime. Simulations are utilized to collect the necessary statistics for locating the onset of dynamo action in parameter space. I will give primary focus to compositional convection at $Pr = 10$ to find that high Prandtl number fluids are significantly less efficient at driving low magnetic Prandtl number dynamos, probably due to the increased role of viscous dissipation for such fluids.

CONTROLLING WAVEBREAKING IN VISCOUS FLUID CONDUITS

Dalton Anderson
Advisors: Mark Hoefer & Michelle Maiden
University of Colorado, Boulder

This presentation will feature a new technique in the experimental investigation of dispersive hydrodynamics. Specifically, this technique allows for the detailed study of dispersive shock waves (DSWs) at the interface of high viscosity contrast fluid conduits. This work will feature experimental results, as well as theoretical derivations for predicting wavebreaking (gradient catastrophe). Mathematically, the central idea is to convert an initial value problem into an approximately equivalent boundary value problem. This is useful because, in practice, the only way to control a viscous fluid conduit is via a boundary condition.

The experimental system to which this technique is applied is a viscous fluid conduit resulting from high viscosity contrast between a viscous, buoyant interior fluid and a more dense and more viscous exterior fluid. The cross-sectional area of this conduit is modeled by a nonlocal, nonlinear, conservative, dispersive, third order partial differential equation (PDE). This equation is used in tandem with a single boundary condition to predict and control the breaking of DSWs in the conduit system.

The boundary condition is derived analytically by neglecting the dispersive term in the governing PDE, and solving the resulting equation backward in time via the method of characteristics. This result is used in experiment to compute an injection rate profile for a high precision piston pump, which generates the desired conduit shape. Utilizing this technique allows for good control of DSW formation as we demonstrate with direct measurements of actual versus predicted breaking locations. Controlling the breaking location of a DSW allows for the investigation of novel dynamics such as backflow, independent of boundary effects.

PARAMETERIZATION AND ANALYSIS OF VISCOUS FLUID CONDUIT EDGES FOR DISPERSIVE HYDRODYNAMICS

Will Farmer
Advisors: Mark Hoefer, Michelle Maiden & Peter Wills
University of Colorado, Boulder

Solitary waves, dispersive shock waves, and other coherent structures can be observed in the viscous fluid conduit system, a deformable, fluid-filled pipe that results from pumping a buoyant, viscous interior fluid into a dense, more viscous exterior fluid. Conduits are observed experimentally as a vertical column where the interior fluid is dyed black for contrast with the colorless exterior fluid. This work will resolve the problem of analyzing large-scale conduits that display curvature effects using a smoothing regression combined with machine learning techniques. Initial data are acquired from micro and macro scale high-resolution photo-graphs which are pipelined through an edge-detection algorithm, yielding the left and right edges of the conduit averaged to get the mean conduit value. This data can display a variety of behavior, from large-scale horizontal changes to small scale fluctuations. Noise is realized in several ways during this process, ranging from imperfections in the fluid pumping mechanism to the camera focus, and as a result a smoothing method must be applied in order to analyze the data properly. The smoothing method model uses overlapping Gaussian curves with varying heights to remove the noise fluctuations from the dataset and yields an analytic line. Then by using a nondimensionalized curvature value and the residuals from our smoothing process, we can classify the "straightness" of the conduit in question. From this point more novel machine learning classification algorithms can be utilized to determine the data quality of arbitrary points in our classification space. Parameterization of the Gaussian smoothing process focuses on repeatability and consistency between different datasets by finding the parameters that minimize the standard deviation of the residuals. This method is self improving and after more data is classified by hand the results will improve. Initial data for model and process testing was acquired from past experiments pending publication.

A PARTIAL DIFFERENTIAL EQUATION MODEL FOR FIRE RESOURCE MOVEMENT

Alex Masarie
Advisors: Yu Wei, Mike Bevers, Iuliana Oprea, and Matt Thompson
Colorado State University

Wildland fire is a naturally occurring phenomenon with which humans have been interacting for millennia. The relationship brings harmony (burning for fuel reduction, indigenous hunting practices, etc...) and discord (financial cost, safety of fire professionals, summer smoke, etc...), but both generate complex problems with ecological and social dimensions. Indeed, an entire scientific field of fire management has evolved over the past century to answer pressing subsets of such problems. Discrete time and space solution techniques are applied often because fire management situations are inherently discrete. Discrete models formulated as mathematical programs, stochastic processes, and cellular automata, among many others, have enhanced our understanding of fire, in terms of behavior and human interaction, while answering calls for actionable fire science. As data resolution increases, the feasibility of partial differential equations as continuous time-space domain models for complex fire improves. This study explores the archive (2011-2015) of national requests as collected by the Resource Ordering Status System (ROSS). For instance, if an ignition location and subsequent fire behavior warrant intensive suppression, the ROSS dataset tabulates this demand by noting which fire suppression resources (air tankers, helicopters, fire engines, hand crews, etc...) were requested and when. When the ROSS-managed supply pool is depleted, decision-makers face increased potential for an unfilled request leaving them without valuable personnel for their fire management effort. This presentation treats the scarce limit of fire resource sharing by exploring national sharing efficiency and local request priority using a basic partial differential heat equation model:

$$\frac{\partial u}{\partial t} + \vec{\nabla} \cdot \Big(k(\vec{x}, t) \vec{\nabla} u \Big) = f(\vec{x}, t),$$

where u characterizes resource density, $k(\vec{x}, t)$ captures resource mobility given current preparedness levels, and $f(\vec{x}, t)$ describes demand based on predicted fire behavior. We explore the possibility of finite difference techniques as tools to examine national sharing efficiency in a comparable manner to a dynamic math program. Similarly, finite element methods present an opportunity to study human-drawn decision-making borders such as Geographic Area Coordination Centers (GACCs) and the finer Predictive Service Areas (PSAs) input as suites of boundary conditions. Information about past, current, and future fire behavior can be parameterized as initial conditions completing the traditional initial boundary value problem (IBVP), which has various relaxations to industry standard network flow math programs. While it remains an open challenge to reconcile such mathematical complexity with actionable decision-making science, recent flourishes of partial differential equations in many domains of physical science have yielded innovative software and rapid algorithms as well as a trained mathematical workforce to implement them. Wildland fire problems require user interfaces with the flexibility and speed managers of the future will seek out not only to face danger and discord on the ground, but also achieve important ecological goals in harmony with fire.

MORNING SESSION III

MODELING REFUGEE MIGRATION

William Van Horn, Samuel Stanton, Zane Showalter-Castorena
Advisor: Gary Olson
University of Colorado, Denver

The 2016 COMAP MCM/ICM competition featured a problem on the subject of refugee migration. Teams were asked to develop a model and use it to support policy recommendations to mitigate the current refugee crisis that has developed in Europe and the Middle East. Solutions had to be created within the 96-hour competition time-frame. The model in this presentation was developed exclusively in those four days, with only inanimate resources for guidance. In order to model the dynamic nature of refugee migration, our team employed nonstationary Markov chains. Using an iteratively updated transition matrix, we effectively simulated the rapidly changing flows of migrants between host nations. We introduced stochastic parameters in order to investigate the effects of varying resources, unpredictable conditions, and rare high-impact events. Our research and analysis of the model led us to the conclusion that an optimal set of policies will prioritize the even distribution of refugees across safe haven nations through an efficient transportation network. Following the implementation of our proposed policies, our model predicts a 38.3% improvement in the distribution of economic burden on haven countries, and an 89% improvement in the number of critically overburdened haven nations overall.

A BIPARTITE 'PUSH-PERMIT' MODEL FOR REFUGEE FLOW

Scott Wurst, Nicolas Pinkowski, & Joseph Arehart
University of Colorado, Boulder

As hundreds of thousands of Syrians continue to flood into Western nations, effective policy and leadership is needed to manage the increasing challenge that this crisis presents. Such policies can either positively or negatively affect the burden placed on nations obligated to provide asylum for refugees under the UN Refugee Convention of 1951. Additionally, the same policies must also take great care to ensure that refugees fleeing one unstable environment are not cast into another. To facilitate the development of such policies, the objective of this report is to provide policies makers with the analysis of a model on refugee dynamics.

The model created in this study can be defined as a bipartite 'push-permit' model of refugee flow. Unlike a traditional 'push-pull' model, this model simulates the dynamics of refugee movement in response to only a pushing force and an analytically determined resistance value. The push is a intangible force that drives people to flee their native country and resistance represents a measure of the obstacles that inhibit
refugee movement (such as distance). However, once a refugee crosses the border of the EU, the movement of refugees is governed by an additional factor, policy, and how policy permits relocation. Part I models the push using a multi-node adaption of the Universal Mobility Model and Part II models the permitted movement within the EU as defined by current policies and proximately. Part II also simulates the corresponding burden on EU member states as determined by current EU economic policy, country GDP, population, and unemployment rate.

MODELING ACUTE STAGE HIV WITH HOMEOSTASIS

Nathan Neri
Advisor: Stephen Pankavich
Colorado School of Mines

The acute stage of in host HIV infection is marked by a rapid spike in viral load, followed by either a full clearance of HIV virions or a low level but persistent chronic infection. Clinical studies suggest that this critical period is influenced by initial concentrations of virions and T-cells, but this factor is unrepresented in the traditional three component model. We propose a new model which incorporates homeostasis with the immune system's response to infection. Local stability analysis of T-cell and virus populations leads to additional complexity in comparison to previous models. The system develops interesting nonlinear dynamics within biologically relevant parameter regimes, including bistable equilibria and Hopf bifurcation based on initial conditions, as desired.

CLASSIFICATION OF THE LINKED TWIST MAP

Rachel Neville
Advisor: Patrick Shipman
Colorado State University

In this project, we approach a classification problem with data arising from the linked twist map, a discrete dynamical system modeling fluid flow. The linked twist map has been used to model flows in DNA micro-arrays with a particular interest in understanding turbulent mixing. The linked twist map is a Poincaré section of eggbeater-type flow in continuous dynamical systems, capturing the behavior of the flow by viewing a particle's location at discrete time intervals. The orbits create complex island structures that vary significantly with the map parameter. As the patterns are forming, there are more significant variations in the patterns which causes classification to be difficult. Persistent homology provides a useful tool in this context. Persistent homology captures multiscale topological information about data; in this case providing a means to detect and compare the island structures. Persistence images allow one to choose the appropriate machine learning algorithm that has strengths appropriate the data under consideration and homological information from multiple dimensions may be leveraged simultaneously. Using these tools, it is possible to achieve good classification of orbit data based on the parameter.

DEVELOPMENT OF POLYGENIC RISK SCORES FOR VITILIGO FROM A GENOME-WIDE ASSOCIATION STUDY

Subrata Paul
Collaborator: Stephanie A. Santorico
University of Colorado, Denver

Generalized vitiligo (GV) is a common autoimmune disease in which white patches of skin and hair result from destruction of melanocytes. We have recently completed a Genome-Wide Association Study (GWAS) that examines genetic variation over the full genome to detect association with GV risk. The analysis included genotypes for 660,000 GWAS single nucleotide polymorphisms (SNPs) and 20,000,000 imputed SNPs for 2853 European-derived white (EUR) cases and 37,412 EUR controls, all having undergone quality control including ancestry matching. In such a large scale association study, not all markers individually achieve significance. An ensemble of such markers can be used to generate a Polygenic Risk Score (PRS) to summarize genetic effects on a particular trait. The markers, selected using a training sample, are used to construct a score by forming the weighted sum representing disease risk. An association of the score with the trait enables us to predict individual trait values. In this presentation, we use the results of our GWAS to select markers associated with generalized vitiligo, and then using PRSice, a software package to provide PRSs, we calculate polygenic risk scores. It is shown, using cross validation, that the PRS, based on the common variants, can explain a reasonable amount of variance of disease risk.

EXPECTED FRACTION OF BASES IN HUMAN CHROMOSOME 21 THAT MAY BE EDITED USING CRISPR-Cas9

Harrison Pielke-Lombardo
Advisor: Manuel Lladser
University of Colorado, Boulder

CRISPR-Cas9 is a novel genome editing technique that allows up to 20 DNA bases to be changed prior to a so called PAM-motif. This is a three nucleotide sequence of the form NRG, where N is any DNA base and R is a purine (i.e. an A or G base). Unfortunately, there are several regions in Chromosome 21 of a known length that are inaccessible by current DNA sequencing technologies and left as N-regions in the sequenced genome. In particular, the fraction of the chromosome that may be edited using CRISPR-Cas9 is unknown. We use Aho-Corasick automata to determine which bases are proximal to a PAM-motif and therefore subject to be changed by CRISPR-Cas9. Furthermore, by regarding N-regions as memoryless sequences, we estimate the expected number of PAM-proximal bases in Chromosome 21. This project has been partially funded by EXTREEMS-QED NSF grant #1407340.

AFTERNOON SESSION I

NOTIONS OF OPTIMALITY IN MULTICRITERIA OPTIMIZATION UNDER UNCERTAINTY

Devon Sigler
Advisor: Alex Engau
University of Colorado, Denver

This talk will present the general structure of a multicriteria optimization problem under uncertainty and introduce some notions of optimality for such problems. In general, deterministic multicriteria optimization problems do not yield a single optimal solution. Due to this fact, other notions of optimality such a Pareto optimality have been developed in order to help decision makers make meaningful comparisons between different solutions. However, in practice often one or many aspects of an optimization problem are uncertain and multicriteria optimization in no exception. As a result, concepts that allow decision makers to compare solutions to multicriteria optimization problems under uncertainty are needed. In this talk several notions of optimality for multicriteria optimization problems under uncertainty will be presented as well as some prelimary results regarding these notions.

APPLICATION OF ADJOINT OPERATORS IN GRADIENT COMPUTATIONS

James Folberth
Advisor: Stephen Becker
University of Colorado, Boulder

When using first-order optimization algorithms, it is often the case that the user must supply the gradient of the differentiable terms in the objective function. We consider two example problems that have a Euclidean error term involving a linear operation on the problem variables. The gradient of the Euclidean error term involves both the linear operator and its adjoint, which, in our examples, is not known in the literature. The first example is an image deblurring problem, where the linear operation is multi-stage wavelet synthesis. Our formulation of the adjoint holds for a variety of boundary conditions, which allows the formulation to generalize to a larger class of problems. The second example is a blind channel estimation problem taken from the convex optimization literature; the linear operation is convolution, but with a slight twist. In each example, we show how the adjoint operator can be applied efficiently.

PARAMETER ESTIMATION GUIDED EXPERIMENTAL DESIGN

Scott Walsh
Advisor: Troy Butler
University of Colorado, Denver

As stakeholders and policy makers increasingly rely upon quantitative predictions from advanced computational models, a problem of fundamental importance is the quantification and reduction of uncertainties in both model inputs and output data. The typical work- flow in the end-to-end quantification of uncertainties requires first formulating and solving stochastic inverse problems (SIPs) using output data on available quantities of interest (QoI). The solution to a SIP is often written in terms of a probability measure, or density, on the space of model inputs. Then, we can formulate and solve a stochastic forward problem (SFP) where the uncertainty on model inputs is propagated through the model to make quantitative predictions on either unobservable or future QoI data. As the fidelity of models increases to include behavior at a wide range of scales, the number of uncertain input parameters often increases significantly, while the number of additional experimental data, which are often costly to obtain, are typically limited. It is therefore becoming increasingly important that we optimally design data collection networks so that the data collected on available QoI leads to improved predictive capabilities of the computational model. In this talk, we use a measure-theoretic framework to formulate and solve SIPs. From this perspective, we discuss the geometric characteristics of using hypothetical sets of QoI that describe both the precision and accuracy in solutions to the SIP. This leads to a natural definition of the optimal experimental design, i.e., what the optimal configuration a finite set of sensors is in space-time. We motivate the solution of these problems with the ADvanced CIRCulation (ADCIRC) model using simulated data from Hurricane Gustav to determine an optimal placement of buoys in the Gulf of Mexico to capture high water marks of the storm surge in order to reduce uncertainties in bottom roughness characteristics.

GEOLOCATION OF EM EMITTERS USING PASSIVE BISTATIC RADAR

Chad Waddington
Advisors: Margaret Cheney[1] & James Given[2]
[1]*Colorado State University*
[2]*Naval Research Laboratory*

The geolocation of emitters of electromagnetic radiation such as radars, radio towers, cell towers, and other telecommunications equipment is of interest both in civilian and defense research. Location of such emitters has applications in locating and identifying low probability of intercept signals which are specifically designed to remain undetected. Geolocation may also have applications in passive image sensing. Current methods assume the locations and transmitted waveforms of such emitters are unknown.

In this talk we will examine the analytical foundations of passive sensing for a Bistatic system involving two receiving platforms. This analytic work will then be reinforced by examining the results obtained in specific numerical simulations involving one stationary receiver and one reciever with a known flight path for a variety of emitter distributions over a given scene.

AFTERNOON SESSION II

THE MODIFIED MATERN PROCESS

Ian Laga
Advisor: William Kleiber
University of Colorado, Boulder

The behavior of a stationary random field can be specified through either its covariance or spectrum. In spatial statistics, the Matern spectrum is the most popular choice due to separation of scale and smoothness effects. We propose a generalization of the Matern spectral density, generating random processes we term modified Matern processes. Our proposal allows for two additional parameters that can loosely be interpreted as arising from a continuous moving average process. Moreover, the Matern is a special case of our model. We illustrate the flexibility of the modified Matern through simulations and discuss parameter estimation via smoothed periodograms.

SPARSE CANONICAL CORRELATION ANALYSIS

Mengjie Yao
Advisor: Stephanie A. Santorico
University of Colorado, Denver

Canonical Correlation Analysis (CCA) is the primary method used to identify and quantify linear relationships between two sets of variables. The objective of CCA is to find the linear combination of each set of variables that maximally correlate with each other. When the original variables are highly correlated or when the number of variables exceeds the sample size, CCA cannot give a stable and unique solution. Also, the resulting linear combination of variables may be difficult to interpret, especially in the context of a large number of variables. Sparse Canonical Correlation Analysis (SCCA) Methods have been proposed to tackle these problems. Aimed at obtaining sparse loadings for the canonical variates, SCCA can be conducted through two different ways. The first being a sparse decomposition of the covariance matrices with the second being to recast CCA as a least squares problem, but with penalization added to the parameters. In this presentation, we will demonstrate the classical CCA method and two different approaches to achieve sparse CCA.

A SIMILAR SEASONS MODEL FOR PREDICTING BASEBALL PERFORMANCE

Lawrence Pelo & Sam Loos
University of Colorado, Denver

We developed a model to predict performance for Major League Baseball (MLB) players in terms of Wins Above Replacement (WAR). Our model calculates the distance between a given player-season (such as 2015 Nolan Arenado) and a selection of historical player-seasons dating to 1920. Once the most similar seasons are identified, the model calculates a weighted average of subsequent performances, resulting in a prediction of performance in terms of WAR for the given player. We tested our model on 2010 MLB data and found that the model performs comparably to established performance prediction systems. This model claimed second place in last October's Colorado Rockies Case Competition.

NOVEL LAGRANGIAN METHODS FOR IMPERFECTLY MIXED CHEMICAL REACTIONS

Michael Schmidt
Advisors: Stephen Pankavich & David Benson
Colorado School of Mines

Most traditional models for simulating chemical reactions are based on the principle of perfect mixing, which seldom occurs in natural settings. For example, Eulerian methods allow for non-homogeneity across grid points but still require perfect mixing within each grid element. In this talk, we present a novel Lagrangian method known as reactive particle tracking (RPT) that is able to simulate transport and reaction dynamics without the well-mixed assumption needed for other methods. One major advantage of RPT is that it allows for the formation of non-homogen- eous "islands" of reactant anywhere in the domain, without the necessity of small-scale discretization. For this reason, RPT is able to model the experimentally observed transition of a well-mixed system into one with imperfect mixing and the subsequent slowdown in reaction speed that follows. As well, the RPT model follows the analytic (well-mixed) PDE solution (with slope t^{-1}) before transitioning to the imperfectly mixed regime (slope $t^{-d/4}$, in d-dimensions). In this direction, we will describe the algorithm and present some preliminary simulation results.

AFTERNOON SESSION III

PRECONDITIONED DATA SPARSIFICATION FOR BIG DATA WITH APPLICATIONS TO PCA AND K-MEANS

Farhad Pourkamali-Anaraki
Advisor: Stephen Becker
University of Colorado, Boulder

We analyze a compression scheme for large data sets that randomly keeps a small percentage of the components of each data sample. The benefit is that the output is a sparse matrix and therefore subsequent processing, such as PCA or K-means, is significantly faster, especially in a distributed-data setting. Furthermore, the sampling is single-pass and applicable to streaming data. The sampling mechanism is a variant of previous methods proposed in the literature combined with a randomized preconditioning to smooth the data. We provide guarantees for PCA in terms of the covariance matrix, and guarantees for K-means in terms of the error in the center estimators at a given step. We present numerical evidence to show both that our bounds are nearly tight and that our algorithms provide a real benefit when applied to standard test data sets, as well as providing certain benefits over related sampling approaches.

GLOBAL SENSITIVITY METRICS FROM ACTIVE SUBSPACES

Paul Diaz
Colorado School of Mines

Active subspaces are an emerging set of mathematical tools that allow us to discover and exploit structures within a mathematical model. Given a model, sensitivity analysis aims to rank a model's input parameters with respect to their influence on a specified quantity of interested. We present novel global sensitivity metrics constructed from active subspaces and apply them to three different mathematical models. We also compare our proposed sensitivity metrics to conventional sensitivity metrics such as, Sobol's total sensitivity indices, and the coefficients of a least-squares-fit linear model approximation.

STABLE PRINCIPAL COMPONENT PURSUIT FOR THE GPU

Derek Driggs
University of Colorado, Boulder

Every analyst who is managing large sets of data has the same two goals: to remove unpredictable errors and to separate the meaningful patterns in the data from the background. Stable Principal Component Pursuit (SPCP) is a data-analytic algorithm that achieves both of these goals, and it does so with little computational effort.

One of the most widely used methods in contemporary data analysis, SPCP decomposes a matrix of noisy data into its underlying low rank and sparse components. In so doing, SPCP separates the sparsely distributed errors in the data-matrix from the patterned entries that contain relevant information. All of this can be formulated as a convex optimization problem, so the algorithm is computationally simple to execute. Due to its utility and feasibility, SPCP has found applications in a variety of fields, including machine vision, genome sequencing, and, the focus of this talk, psychometrics.

Although SPCP is powerful and simple to implement, it is extraordinarily slow. Applying SPCP to a typical data set from an MRI brain scan, for example, can often take several hours to run, even when using state-of-the-art solvers. On top of this, SPCP requires parameter tuning that is performed through cross-validation techniques, requiring several additional runs. To expand the applicability of SPCP, it must be accelerated.

In this talk, we will explore a method to accelerate SPCP through parallelization. Parallelized architectures trade an increase in communication costs for the ability to execute many tasks simultaneously, and they are ubiquitous in high-performance computing. A single Graphical Processing Unit (GPU), for example, contains hundreds of cores with thousands of threads each, all of them able to execute a simple task at the same time. Unfortunately, the traditional SPCP algorithm is communication-heavy, so the high cost of communication on a GPU outweighs the speedup due to parallelization.

The goal of this research is to decrease the communication burden within SPCP by parallelizing existing matrix-decomposition algorithms. By decreasing communication costs, SPCP is able to enjoy the computational speedup offered by the GPU, significantly decreasing runtimes in applications. We apply this parallelized SPCP to analyze MRI brain scans, and compare its performance to traditional SPCP algorithms.

2D VISCOUS LAMINAR FLOW AIRFOIL DESIGN APPLYING PRINCIPAL COMPONENT ANALYSIS AND ACTIVE SUBSPACES

Zach Grey
Advisor: Paul Constantine
Colorado School of Mines

The design and optimization of airfoils is a common engineering problem. For simplicity, only 2D viscous laminar flow solutions are consider to avoid long computational lead times for demonstrating the fundamentals of PCA (principal component analysis) and active subspaces. The fundamentals of the two solutions involve similar mathematical principals of eigenvalue decompositions. However, the problems which the motivate the use of the separate methods are entirely different and subsequently they are consider two different means of dimensionality reduction. These methods will be demonstrated in the context of 2D viscous laminar flow airfoil designs to highlight specific differences and applications were the methods are considered useful.

MODELING OF UNCERTAINTY IN RESREVOIR WITH A STOCHASTIC BASIS METHOD

Prosper Torsu
University of Wyoming

Quantifying uncertainty effects of coefficients that exhibit heterogeneity at multiple scales is among many out- standing challenges in subsurface flow models. Typically, the coefficients are modeled as functions of random variables governed by certain statistics. To quantify their uncertainty in the form of statistics (e.g., average fluid pressure or concentration) Monte-Carlo methods have been used. In a separate direction, multi-scale numerical methods have been developed to efficiently capture spatial heterogeneity that otherwise would be intractable with standard numerical techniques. Since heterogeneity of individual realizations can differ drastically, a direct use of multi-scale methods in Monte-Carlo simulations is problematic. Furthermore, Monte-Carlo methods are known to be very expensive as a lot of samples are required to adequately characterize the random component of the solution. In this study, we utilize a stochastic representation method that exploits the solution structure of the random process in order to construct a problem dependent stochastic basis. Using this stochastic basis representation a set of coupled yet deterministic equations is constructed. In the proposed method, enrichment of the solution space can be performed at multiple levels that offer a balance between computational cost, and accuracy of the approximate solution.