
Optimization and Data Science
Optimisation et Science des Données

(Org: **Michael Friedlander** (University of British Columbia), **Abraham P Punnen** (Simon Fraser University) and/et
Mohamed Tawhid (Thompson Rivers University))

MONTAZ ALI, Witwatersrand University
Convex Formulation for Planted Quasi-Clique Recovery

In this talk, we consider the planted quasi-clique or γ -clique problem. This problem is an extension of the well known planted clique problem which is NP-hard. The maximum quasi-clique problem is applicable in community detection, information retrieval and biology. We propose a convex formulation using nuclear norm minimization for planted quasi-clique recovery. We carry out numerical experiments using our convex formulation and the existing mixed integer programming formulations. Results show that the convex formulation performs better than the mixed integer formulations when γ is greater than a particular threshold.

ALEKSANDR ARAVKIN, University of Washington (Institute for Health Metrics and Evaluation)
A Robust Risk Score for Evaluating Evidence in Global Health

How strong is the relationship between (red meat, alcohol) and heart disease? (Sugar sweetened beverages, BMI) and diabetes? (Smoking, air pollution) and lung cancer? Each of these risk-outcome pairs has been the subject of numerous large studies. With results in hand, can we rate the evidence, and compare these risks to guide policy, and initiate further studies?

We present a new methodology to answer these questions. The methodology comprises modeling between-study disagreement, capturing nonlinear dose-response relationships, detecting outliers, and accounting for nonlinear observation mechanisms inherent in how studies report results. It is now used to analyze more than 470 risk-outcome pairs in the Global Burden of Disease study, conducted by IHME and collaborators. We present the main model, highlight the role of optimization, and include recent results for select risk-outcome pairs of interest.

YANKAI CAO, University of British Columbia
A Global Optimization Algorithm for Clustering Problems

We present a reduced-space spatial branch and bound strategy for two-stage stochastic nonlinear programs. At each node, a lower bound is constructed by relaxing the non-anticipativity constraints, and an upper bound is constructed by fixing the first-stage variables. Both lower and upper bounds can be computed by solving individual scenario subproblems. Another key property is that we only need to perform branching on the first-stage variables to guarantee convergence.

We also extend this algorithm to address clustering problems (a class of unsupervised learning). Preliminary numerical results have demonstrated that our algorithm is able to solve problems with hundreds of samples to global optimality within a reasonable amount of time, while state-of-the-art global solvers can only deal with tens of samples. Moreover, global optimization can significantly improve performance on several datasets compared with local optimal algorithms, such as k-means.

MONICA GABRIELA COJOCARU, University of Guelph
On Minty-variational inequalities and evolutionary stable states of generalized stable games

In this work we emphasize the role played by Minty variational inequalities in evolutionary game theory for studying neutrally and evolutionary stable states of nonlinear population games. This connection allows deriving new results on the sets of neutrally stable and evolutionary stable states for generalized stable games as well as stability results for the replicator dynamics.

PAULA FERMÍN CUETO, University of Edinburgh

Machine learning and statistical methods for characterising and predicting capacity degradation of Li-ion cells

In automotive applications, Li-ion cells are typically considered to have reached their end-of-life when they are down to 80% of their initial capacity. However, the degradation of these cells typically displays a "knee" pattern: the capacity degrades at a slow rate up to a so-called "knee-point", after which it degrades very rapidly until its end-of-life. This knee-point therefore gives a more advanced warning of the cell's degradation than the end-of-life. Nevertheless, the industry does not have a standard definition or identification method for this crucial metric.

In this talk, we present robust statistical methods to identify two different knee-points in capacity degradation data of Li-ion cells. Following this identification step, we show how machine learning algorithms can be employed to successfully predict the occurrence of these knee-points from the first few discharge cycles of a cell's life. We rely on feature engineering to overcome the challenge of working with a very small, yet high-dimensional data set and we quantify the uncertainty of the predictions to build trust in our models.

GONÇALO DOS REIS, University of Edinburgh

State of Health for the capacity and internal resistance of Li-ion cells: A machine learning approach with knees and elbows

Degradation of lithium-ion cells with respect to increases of internal resistance (IR) has negative implications for rapid charging times and thermal management of cell in electric vehicles and energy storage applications. Despite this, IR and associated IR State of Health have received much less attention than the State of Health with respect to capacity degradation in Li-ion research. We address this by building on recent developments on "knee" identification for capacity degradation curves. We propose the concepts of "elbow-point" and "elbow-onset" for IR degradation curves, and create an identification algorithm for these variables.

We use machine learning Neural Network techniques to build independent capacity and IR predictor models achieving a MAPE of 0.4% and 1.6%, respectively. We then use the IR model to synthesize internal resistance data to complete the dataset from Attia et al 2020 for which no IR data was logged.

LUKASZ GOLAB, University of Waterloo

Explanation Tables

I will present a solution to the following data summarization problem: given a dataset with multiple categorical features and a binary outcome, construct a summary that offers an interpretable explanation of the factors affecting the outcome in terms of the feature value combinations. We refer to such a summary as an explanation table, which is a disjunction of overlapping patterns over the features, where each pattern specifies a conjunction of attribute=value conditions. We give an efficient and effective greedy algorithm to construct explanation tables that uses sampling to prune the space of candidate patterns. This is joint work with Parag Agrawal, Kareem El Gebaly, Guoyao Feng, Flip Korn, Divesh Srivastava and Michael Vollmer.

WARREN HARE, University of British Columbia

Imaginary Derivative Free Optimization

Consider the problem of minimizing an objective function that is provided by a blackbox. Suppose that, while the optimization problem seeks a real-valued solution, the blackbox is capable of accepting complex-valued input and returning complex-valued output. We explore using complex-variables in a model-based derivative-free optimization method. We begin by discussing how to construct such model and then present the results of some numerical experiments. Results suggest that the quality of a model-based DFO algorithm is (i) highly impacted by the number of function evaluations required to create the models and (ii) also impacted by the accuracy of the created models, but to a lesser degree.

THOMAS HUMPHRIES, University of Washington Bothell

Unrolled iterative algorithm for CT image reconstruction with learned penalty term

In the last three to four years there has been an explosion of interest in using deep learning techniques to address challenging problems in computed tomography (CT) image reconstruction, such as low-dose, sparse-view, and limited angle imaging. A wide variety of approaches have been proposed, including using deep neural networks (DNN) as pre- or post-processing steps, using neural networks to encode prior information within existing iterative reconstruction algorithms, or learning to solve the inverse problem altogether.

We present a CT reconstruction approach which unrolls a standard iterative algorithm and trains it end-to-end as a DNN. The DNN consists of fixed layers, corresponding to the basic iterative algorithm, as well as trainable layers, which have the effect of perturbing the solution between iterations. The trainable layers can be viewed as replacing the negative gradient of an unknown penalty function or regularizer, which can vary with the iteration number. In numerical experiments, we test the approach on sparse-view and limited-angle CT problems, and study the effect of network architecture on the effectiveness of the algorithm. The proposed method provides significant improvement over the basic iterative algorithm, as well as total variation minimization approach. Joint work with Yiran Jia, Noah McMichael, Pedro Mokalzel, and Dong Si (UW Bothell).

ABDELMONEM IBRHAIM, Mathematics Department, Faculty of Science, Al-Azhar University, Egypt
Binary whale optimization algorithm for feature selection

The principle of any approach for solving feature selection problem is to find a subset of the original features. Since finding a minimal subset of the features is an NP-hard problem, it is necessary to develop and propose practical and efficient heuristic algorithms. The whale optimization algorithm is a recently developed nature-inspired meta-heuristic optimization algorithm that imitates the hunting behavior of humpback whales to solve continuous optimization problems. In this paper, we propose a novel binary whale optimization algorithm (BWOA) to solve the feature selection problem. BWOA is especially desirable and appealing for feature selection problem whenever there is no heuristic information that can lead the search to the optimal minimal subset. Nonetheless, whales can find the best features as they hunt the prey. Rough set theory (RST) is one of the effective algorithms for feature selection. We use RST with BWOA as the first experiment, and in the second experiment, we use a wrapper approach with three different classifiers for feature selection. Also, we verify the performance and the effectiveness of the proposed algorithm by performing our experiments using 32 datasets from the UCI machine learning repository and comparing the proposed algorithm with some powerful existing algorithms in the literature. The results show that the proposed algorithm can provide an efficient tool to find a minimal subset of the original features.

ZHAOSONG LU, University of Minnesota
First-Order Augmented Lagrangian Methods for Convex Conic Programming

In this talk, we propose some first-order augmented Lagrangian (AL) methods for solving a class of convex conic programming with adaptive update on penalty parameters and inexactness associated with the AL subproblems. We establish their first-order oracle complexity for finding an approximate Karush–Kuhn–Tucker (KKT) point. To our best knowledge, our complexity is the lowest one among all existing first-order AL methods for finding an approximate KKT point.

IBRAHIM NUMANAGIĆ, University of Victoria
Optimization in Pharmacogenomics

High-throughput sequencing provides the means to determine the allelic decomposition— the exact sequence content of all gene copies present in the sample— for any gene of interest. When applied to pharmaceutical genes, such decomposition can be used to inform the drug treatment and dosage decisions. However, many clinically and functionally important genes are highly polymorphic and have undergone structural alterations, and as such present a significant challenge for the existing genotyping methods. Here we present a combinatorial optimization framework based on integer linear programming that is able to efficiently solve this problem for various pharmacogenes, including those with structural alterations. We also show how to adapt these linear programs for the emerging long-range barcoded sequencing datasets.

COURTNEY PAQUETTE, McGill University

Halting Time is Predictable for Large Models: A Universality Property and Average-case Analysis

Average-case analysis computes the complexity of an algorithm averaged over all possible inputs. Compared to worst-case analysis, it is more representative of the typical behavior of an algorithm, but remains largely unexplored in optimization. One difficulty is that the analysis can depend on the probability distribution of the inputs to the model. However, we show that almost all instances of high-dimensional data are indistinguishable to first-order algorithms. Particularly for a class of large-scale problems, which includes random least squares and one-hidden neural networks with random weights, the halting time is independent of the probability distribution. With this barrier for average-case analysis removed, we provide the first explicit average-case convergence rates showing a tighter complexity not captured by traditional worst-case analysis. Finally, numerical simulations suggest this universality property holds in greater generality. Joint work with Elliot Paquette, Fabian Pedregosa, and Bart van Merriënboer

MARK SCHMIDT, University of British Columbia

Faster Algorithms for Deep Learning?

The last 10 years have seen a revolution in stochastic gradient methods, with variance-reduced methods like SAG/SVRG provably achieving faster convergence rates than all previous methods. These methods give dramatic speedups in a variety of applications, but have had virtually no impact to the practice of training deep models. We hypothesize that this is due to the over-parameterized nature of modern deep learning models, where the models are so powerful that they could fit every training example with zero error (at least theoretically). Such over-parameterization nullifies the benefits of variance-reduced methods, because in some sense it leads to "easier" optimization problems. In this work, we present algorithms specifically designed for over-parameterized models. This leads to methods that provably achieve Nesterov acceleration, methods that automatically tune the step-size as they learn, and methods that achieve superlinear convergence with second-order information.

XIAOPING SHI, Thompson Rivers University

Graph-based change-point test

Modeling high-dimensional time series is necessary in many fields such as neuroscience, signal processing, network evolution, text analysis, and image analysis. Such a time series may contain unknown multiple change-points. For example, the time of cell divisions can be accessed using an automatic embryo monitoring system by a time-lapse observation. When a cell divides at some time point, the distribution of pixel values in the corresponding frame will change, and hence the detection of cell divisions can be formulated as a multiple change-point problem. In this talk, we introduce a powerful change-point test in terms of the shortest Hamiltonian path (SHP).

TAMON STEPHEN, Simon Fraser University

Minimal Cuts Set and Computing with Monotone Boolean Functions

We consider methods of generating Minimal Cut Sets in computational biology. This leads us to monotone Boolean functions, which can be used model phenomena in a variety of large data applications. Computing with monotone Boolean functions is an interesting and worthwhile mathematical challenge. We briefly introduce some of the techniques and issues involved.

JABED TOMAL AND JAN CIBOROWSKI, Thompson Rivers University

Detection of environmental thresholds by assessing discontinuities in slopes and variances via a Bayesian regression model

Co-author: Jan J.H. Ciborowski, Professor, Biological Sciences, University of Calgary

Abstract: An ecological threshold occurs at a point along an environmental stress gradient at which there is a discontinuous change in the conditional distribution of a biological response. Traditionally, ecological thresholds are estimated from the discontinuities in the central tendency (e.g., slope) using a piecewise linear regression model (PLRM). However, thresholds

can also be manifested as changes in the range of natural variation (e.g., conditional variance) for a given level of the environmental stress. In this paper, we defined a Bayesian PLRM by incorporating experts' knowledge about the relationships between the biological response relative to environmental stress represented via prior distributions. The posterior distributions of the thresholds are obtained by combining the information in the data with experts' prior knowledge, and optimized via Gibbs sampling and the Metropolis algorithm. We applied our method to two datasets relating an index of the health of marsh-nesting bird communities to habitat alteration (areal extent of land development adjacent to wetlands) within the Detroit River Area of Concern. Our preliminary analysis identified two potential thresholds - one manifested via the change in slope and the other observed from increased variance across the environmental stress gradient.

XUEKUI ZHANG, University of Victoria

The Optimal Design of Clinical Trials with Potential Biomarker Effects, A Novel Computational Approach

As a future trend of healthcare, personalized medicine tailors medical treatments to individual patients. It requires to identify a subset of patients with the best response to treatment. The subset can be defined by a biomarker (e.g. expression of a gene) and its cutoff value. Topics on subset identification have received massive attention. There are over 2 million hits by keyword searches on Google Scholar. However, designing clinical trials that utilize the discovered uncertain subsets/biomarkers is not trivial and rarely discussed in the literature. This leads to a gap between research results and real-world drug development.

To fill in this gap, we formulate the problem of clinical trial design into an optimization problem involving high-dimensional integration, and propose a novel computational solution based on Monte-Carlo and smoothing methods. Our method utilizes the modern techniques of General-Purpose computing on Graphics Processing Units for large-scale parallel computing. Compared to a published method in three-dimensional problems, our approach is more accurate and 133 times faster. This advantage increases when dimensionality increases. Our method is scalable to higher-dimensional problems since the precision bound of our estimated study power is a finite number not affected by dimensionality.