Northwestern | ENGINEERING

Center for Optimization and Statistical Learning

ACNTW Workshop

May 18, 2017

Northwestern's Kellogg Conference Center

340 E Superior St, Chicago, IL 60611

Wieboldt Hall, Room 147

Program Overview

9:30-10:00am	Breakfast available and brief welcome
10:00-10:30am	John Lafferty
10:30-11:00am	Aravindan Vijayaraghavan
11:00-11:30am	Srinadh Bhojanapalli
11:30-12:00pm	Rina Foygel Barber
12:00-1:30pm	Lunch: Room 440 (food) and 540 (seating)
1:30-2:00pm	Rob Nowak
2:00-2:30pm	Becca Willett
2:30-3:00pm	Break
3:00-3:30pm	Prasanna Balaprakash
3:30-4:00pm	Jorge Nocedal
4:00-5:30pm	Poster session
6:00pm	Dinner at Quartino's

All lectures are in Room 147 of Wieboldt Hall. The poster session is in Rooms 248 and 262.

Dinner

We have a private room at Quartino Ristorante & Wine Bar, 626 N. State St. Chicago, IL 60654:
https://www.quartinochicago.com/
The restaurant is walking distance (about 0.5 mile) from the conference site.

Lectures

10:00am

Superadaptive Optimization

John Lafferty, University of Chicago

Abstract: How difficult is it to minimize a specific convex function? This question is tricky to formalize–traditional complexity analysis is expressed in terms of the worst case over a large class of instances. We extend the classical minimax analysis of stochastic convex optimization by introducing a localized form of complexity for individual functions. The complexity measure is shown to be equivalent to a computational analogue of Fisher information and the modulus of continuity that are central to statistical estimation and inference. We also show that very simple algorithms can be superadaptive. Connections with other learning and testing problems are discussed. Joint work with Sabyasachi Chatterjee (Chicago), John Duchi (Stanford), and Yuancheng Zhu (Wharton).

10:30am

Learning Probabilistic Models for Graph Partitioning and Community Detection

Aravindan Vijayaraghavan, Northwestern University

Abstract: The Stochastic Block Model or the Planted Partition Model is the most widely used probabilistic model for community detection and clustering graphs in various fields, including machine learning, statistics, and social sciences. Many existing algorithms (e.g., spectral algorithms) successfully learn the communities or clusters when the data is drawn exactly according to the model. However, many of these guarantees do not hold in the presence of modeling errors, or when there is overlap between the different communities.

In this talk, I will address the following question: can we design robust efficient algorithms for learning probabilistic models for community detection that work in the presence of adversarial modeling errors?

I will describe different computationally efficient algorithms that provably recover communities or clusters (up to small recovery error). These algorithmic results will work for probabilistic models that are more general than the stochastic block model, or when there are different kinds of modeling errors or noise.

11:00am

Dropping Convexity for Scalable Machine Learning

Srinadh Bhojanapalli, Toyota Technological Institute at Chicago

Abstract: Many modern machine learning problems are solved using complex high dimensional models on large datasets and require scalable algorithms with low computation and memory complexity. This often leads to use of non-convex methods which do not necessarily come with computational/ statistical guarantees.

In this talk, I will first discuss how to leverage the special structure of rank constraint to design scalable algorithms for low rank recovery problems. Later, I will present our work on understanding the mystery of success of simple local search methods, such as stochastic gradient descent, in recovering the global optima of low rank recovery problems. I will also present a new framework to analyze the behavior of gradient descent methods for rank constrained problems and provide rigorous guarantees.

11:30am

A Projected Gradient Descent with Nonconvex Constraints

Rina Foygel Barber, University of Chicago

Abstract: Nonconvex optimization arises in many applications of high-dimensional statistics and data analysis, where data models and regularization terms can both often exhibit nonconvexity. While convex programs for structured signal recovery have been widely studied, comparatively little is known about the theoretical properties of nonconvex optimization methods. In this talk I will discuss the problem of projected gradient descent over nonconvex constraints, where the local geometry of the constraint set is closely tied to its convergence behavior. By measuring the local concavity of the constraint set, we can give concrete guarantees for convergence of projected gradient descent. Furthermore, by relaxing these geometric conditions, we can allow for approximate calculation of the projection step to speed up the algorithm.

1:30pm

Optimization for Metric Learning

Rob Nowak, University of Wisconsin–Madison

Abstract: Modeling human perceptions has many applications in cognitive, social, and educational science, as well as, in advertising and commerce. This talk discusses theory and optimization methods for learning metric representations of human perceptions. I will present recent our work ordinal embedding, also known as non-metric multidimensional scaling, which is the problem of representing items (e.g., images) as points in a low-dimensional Euclidean space given constraints of the form "item i is closer to item j than item k." In other words, the goal is to find a geometric representation of the items that is faithful to comparative similarity judgments. A variety of optimization algorithms exist for learning metric embeddings from comparison data, but the accuracy and performance of these methods were poorly understood. I will present a new theoretical framework that quantifies the accuracy of learned embeddings and indicates how many comparisons suffice in terms of the number of items and the dimension of the embedding. The theory also points to new optimization methods that outperform previously proposed methods.

2:00pm

Algebraic Variety Models for High-Rank Matrix Completion

Becca Willett, University of Wisconsin–Madison

Abstract: Work in the last decade on matrix completion has shown it is possible to leverage linear structure to interpolate missing values in a low-rank matrix. However, the assumption that the data is low-rank is not always met in practice, and it is of great interest to extend matrix completion theory and algorithms to other low-complexity nonlinear structures. In this work we study the problem of completing a matrix whose columns belong to an algebraic variety, i.e., the set of solutions to a system of polynomial equations. In this setting, the original matrix is possibly high-rank, but it becomes low-rank after mapping each column to a higher dimensional space of monomial features. Many well-studied extensions of linear models, including affine subspaces and their union, plus a rich class of nonlinear quadratic and higher degree curves and surfaces, are captured in a variety model. We propose an efficient matrix completion optimization algorithm that minimizes a convex or non-convex surrogate of the rank of the matrix of monomial features. Our algorithm uses the well-known kernel trick to avoid working directly with the high-dimensional monomial matrix. We show the proposed algorithm is able to recover synthetically generated data up to predicted sample complexity bounds. The proposed algorithm also outperforms standard low-rank matrix completion and subspace clustering techniques in experiments with real motion capture data.

This is joint work with Greg Ongie, Laura Balzano, and Robert Nowak.

3:00pm

Automatic Multi-Objective Modeling with Machine Learning

Prasanna Balaprakash, Argonne National Laboratory

Abstract: In recent years, automatic data-driven modeling with machine learning (ML) has received considerable attention as an alternative to analytical modeling for many modeling tasks in high performance computing. While ad hoc adoption of ML approaches has obtained success, the real potential for automation in data-driven modeling has yet to be achieved. In this talk, we will describe AutoMOMML, an end-to-end, ML-based framework to build predictive models for objectives such as performance, power, and energy of scientific applications. The framework adopts statistical and optimization approaches to reduce the modeling complexity and automatically identifies and configures the most suitable learning algorithm to model the required objectives based on hardware and application signatures.

3:30pm

Optimization Methods for Deep Neural Networks

Jorge Nocedal, Northwestern University

Abstract: We discuss the main challenges facing optimization methods for very large scale machine learning problems. Two case studies provide insights into the topology of loss functions associated with deep learning. We discuss the need to design algorithms that have both good generalization properties and the ability to parallelize. We present some intriguing observations about the optimization of deep neural networks.

Poster Session

Are Newton-Sketch and Subsampled Newton Methods Effective in Practice?

Albert Berahas, Northwestern University

Coauthors: Raghu Bollapragada and Jorge Nocedal (Northwestern University)

Abstract: The concepts of subsampling and sketching have recently received much attention by the optimization and statistics communities. In this work, we focus on their numerical performance, with the goal of providing new insights into their theoretical and computational properties. We pay particular attention to subsampled Newton and Newton-Sketch methods, as well as techniques for solving the Newton equations approximately. Our tests are performed on a collection of optimization problems arising in machine learning.

Efficient Computation of Derivatives for Optimization Problems in R and Python

Julie Bessac, Argonne National Laboratory

Coauthors: K. Kulshreshtha (Universitat Paderborn), S.H.K. Narayanan (Argonne National Laboratory), K. MacIntyre (Northwestern University)

Abstract: R and Python have become popular in fields such as machine learning, statistics, and data analysis. Many statistical and machine learning problems can be recast as optimization problems. Indeed they often involve the optimization of a cost function; for instance, maximizing the likelihood in the statistical model-oriented framework. However, the derivatives of the cost functions are numerically approximated or coded by hand. Algorithmic differentiation (AD) is a technique to compute derivatives of functions represented as evaluation procedures efficiently and accurately. Algorithmic differentiation tools have not been proposed for scripting languages like R or Python. ADOL-C is a C++ library that provides accurate derivatives. In this work, we create interfaces of ADOL-C for R and Python and we present statistical and machine learning optimization examples in R and Python. Especially, an application in R is introduced for a statistical model that was initially built for space-time prediction of surface wind speed.

Exact and Inexact Subsampled Newton Methods for Optimization

Raghu Bollapragada, Northwestern University

Coauthors: Jorge Nocedal (Northwestern University) and Richard Byrd (University of Colorado)

Abstract: We study the solution of stochastic optimization problems in which approximations to the gradient and Hessian are obtained through subsampling. We show how to coordinate the accuracy in the gradient and Hessian to yield a superlinear rate of convergence in expectation. We also consider inexact Newton methods and investigate what is the most effective linear solver in terms of computational complexity. We present numerical results on logistic regression problems.

Understanding Variation in the Nature of Stochastic Textured Surfaces with Manifold Learning

Anh Bui, Northwestern University

Coauthor: Daniel Apley (Northwestern University)

Abstract: This work aims to use manifold learning to discover the systematic variation in the nature of stochastic textured surfaces. Due to their stochastic character, it is challenging to obtain pairwise dissimilarities between stochastic textured surface samples, which are required input for manifold learning algorithms. We introduce a number of dissimilarity measures that involve implicitly estimating the joint distribution of the collection of pixels in each surface sample via an appropriate supervised learning model. We then use manifold learning on these dissimilarities to discover a low-dimensional parameterization of the variation in the nature of the stochastic textured surfaces. Visualizing how the surfaces change as the manifold parameters are varied helps to build an understanding of the physical nature of each systematic variation pattern. We demonstrate the approach with simulation and textile examples.

Multigrid Approach for Tomographic Reconstruction

Zichao (Wendy) Di, Argonne National Laboratory

Coauthors: Sven Leyffer and Stefan M. Wild (Argonne National Laboratory)

Abstract: Tomographic imaging refers to the reconstruction of a 3D object from its 2D projections by sectioning the object, through the use of any kind of penetrating wave, from many different directions. In particular, Fluorescence tomography can be used to reveal the internal elemental composition of a sample while transmission tomography can be used to obtain the spatial distribution of the absorption coefficient inside the sample. In this work, we integrate both modalities and formulate an optimization approach to simultaneously reconstruct the composition and absorption effect in the sample. However, the reconstruction performance still suffers from the curse of dimensionality. We apply a generally-applicable scalable approach multigrid-based optimization framework (MG/OPT) to enhance the reconstruction performance. We provide several numerical results on the performance of the joint inversion in terms of reconstruction quality, as well as significant speedup and improvement of accuracy further provided by MG/OPT.

Clustering in Euclidean Space under Additive Perturbation Stability

Abhratanu Dutta, Northwestern University

Coauthors: Aravindan Vijayaraghavan (Northwestern University) and Alex Wang (Northwestern University)

Abstract: Clustering is a well-studied and widely-utilized problem, both in theory and in practice. While most formulations of clustering (e.g. k-means clustering) are NP-hard in the worst-case, practitioners have had remarkable success with simple heuristics for these problems. Hence, it is compelling to consider paradigms that go beyond traditional worst-case analysis in order to design efficient algorithms with provable guarantees for this problem. We consider a natural new notion of instance stability that we believe interesting practical instances of k-means clustering satisfy. We consider instances whose optimal k-means clustering does not change when each point is moved by a small amount. Practical instances often have measurement errors; thus it is desirable for an instance with nice ground-truth clustering to have its optimal clustering remain the same in spite of such errors. We give efficient algorithms for k-means clustering that provably recover the optimal clustering under sufficient amount of additive perturbation stability.

Alternating Minimization and Alternating Descent for Nonconvex Optimization Problems

Wooseok Ha, University of Chicago

Coauthors: Rina Barber (University of Chicago)

Abstract: Many optimization problems in high-dimensional statistics and signal processing involve two decision variables to be minimized, where the variables often reflect different structures of the signals being considered. Alternating minimization is a widely used method for solving such optimization problems, but the general properties of alternating minimization has not yet been understood well in some settings. In this work, we study and analyze the performance of alternating minimization under the setting where the variables are constrained to nonconvex sets under standard assumptions on the loss function such as restricted strong convexity. Since in practice performing the exact alternating minimization might be intractable, we also approximate it with projected gradient descent steps and show that alternating descent approximates alternating minimization quickly, therefore obtaining fast convergence guarantees to the optimal. Our analysis depends strongly on the notion of local concavity coefficients, which have been recently proposed to measure and quantify the nonconvexity of a general nonconvex constraint set. We demonstrate our conditions on two important classes of the problems, low rank + sparse decomposition and multivariate regression problem, and provide some simulation results to see the empirical performance of the algorithms.

Activation Ensembles for Deep Neural Networks

Mark Harmon, Northwestern University

Coauthor: Diego Klabjan (Northwestern University)

Abstract: We propose a new methodology of designing activation functions within a neural network at each layer. We call this technique an "activation ensemble" because it allows the use of multiple activation functions at each layer. This is done by introducing additional variables, α, at each activation layer of a network to allow for multiple activation functions to be active at each neuron. By design, activations with larger values at a neuron is equivalent to having the largest magnitude. Hence, those higher magnitude activations are "chosen" by the network. We implement the activation ensembles on a variety of datasets using an array of Feed Forward and Convolutional Neural Networks. By using the activation ensemble, we achieve superior results compared to traditional techniques. In addition, because of the flexibility of this methodology, we more deeply explore activation functions and the features that they capture.

Model Enhancement via Annotations: Extension of EMP for Lagrangian Modifications

Olivier Huber, University of Wisconsin–Madison

Coauthor: Michael Ferris (Wisconsin-Madison)

Abstract: We describe an extension of the extended mathematical programming (EMP) approach for soft penalties and show how to model these using quadratic support functions. We extend problems to allow for soft constraints and constraints on risk measures such as conditional value at risk. Our approach allows straightforward specification of these constructs. Automatic program augmentation and transformations that understand the underlying structure create different models that are amenable to solution by existing solvers. Computational results on several different applications are shown, along with some comparisons of the efficiencies of various reformulation/solver combinations. The underlying technology is extensible to further model types and other modeling systems.

Logical Benders Decomposition for Quadratic Programs with Complementarity Constraints

Francisco Jara-Moroni, Northwestern University

Coauthors: Andreas Wächter (Northwestern University), Jong-Shi Pang (USC), and John Mitchell (RPI)

Abstract: We study a logical Benders decomposition approach to solving quadratic programs with linear complementarity constraints. It is based on a satisfiability master problem to which feasibility cuts are added, computed from primal and dual subproblems for chosen complementarity pieces. Interpreting the logical Benders decomposition approach from a branch-and-bound point of view, we propose several new methods for strengthening the feasibility cuts and guiding the master problem solution. Their efficiency is assessed by numerical experiments.

Optimal Clustering on a Graph

Gokce Kahvecioglu, Northwestern University

Coauthor: David Morton (Northwestern University)

Abstract: We study a clustering problem defined on an undirected graph with a weight function assigning nonegative weights to the edges, which denotes the importance of the connection between pairs of vertices. We remove a subset of edges to break the graph into a number of smaller pieces, i.e., clusters. Subject to a constraint on the total weight of edges that are removed, we maximize the number of clusters we form. We introduce a two-stage formulation of the clustering problem, and solve it by Benders decomposition algorithm. We also propose a bicriteria graph clustering problem, in which we maximize the number of clusters we form while minimizing the weight of deleting edges. Solving this bicriteria problem parametrically identifies the solutions that lie on the concave envelope of the efficient frontier, and the breakpoints on the concave envelope of the efficient frontier are nested.

Kernel L1-norm PCA

Cheolmin Kim, Northwestern University

Coauthor: Diego Klabjan (Northwestern University)

Abstract: In this work we develop an algorithm for the variance maximization version of kernel L1-norm PCA. Unlike kernel L2-norm PCA, the kernel version of L1-norm PCA is a hard problem in that it is not only non-convex but also non-smooth. However, through the reformulation, we make it a more geometrically interpretable problem where the goal is to minimize the L2-norm of a vector subject to a linear constraint involving the L1-norm terms. For this reformulated problem, we present a "fixed point" type algorithm that iteratively computes a -1,1 weight for each data point based on the kernel matrix and previous weights. We show that the kernel trick is applicable to this algorithm and provide a convergence analysis. We show the finite convergence of the algorithm as well as the linear convergence of objective values. Lastly, we computationally investigate the robustness of our algorithm and illustrate its usage for outlier detection.

Modeling Application I/O Performance Variability: A Probabilistic Graphical Model Approach

Sandeep Madireddy, Argonne National Laboratory

Coauthors: Prasanna Balaprakash, Philip Carns, Robert Latham, Robert Ross, Shane Snyder, and Stefan M. Wild (Mathematics & Computer Science Division, Argonne National Laboratory)

Abstract: I/O performance variability is a well-known problem for applications that run on leadership-class systems. However, this variability is not well understood and there is a lot speculation on how to detect and compensate for it. Often deterministic assumptions are made for the application I/O performance, which might introduce errors into the I/O performance models and lead to a suboptimal system configuration. We employ a probabilistic graphical model-based machine learning approach that capture the conditional dependence structure between the I/O performance of the applications and the state of the file systems to characterize the performance variability. The dependency structure is obtained by efficiently coalescing data-driven approaches and expert knowledge. Our approach allows one to also model latent variables and missing data.

Learning Rates for Network Parameters in High-dimensional Multi-variate Point Process Models

Ben Mark, University of Wisconsin–Madison

Coauthor: Rebecca Willett and Garvesh Raskutti (Wisconsin-Madison)

Abstract: Poisson autoregressive models are a common way of capturing self-exciting point processes, where events from nodes in a network either stimulate or inhibit events from other nodes. These models can be used to learn the structure of social or biological neural networks. However, ensuring stability in these models can require overly restrictive assumptions on the network structure. We will consider a generalization of the PAR model which allows for non-linear dampening effects. Such a model allows for self-excitation while avoiding exponential blow-up in event counts, but the introduction of non-linearities brings new challenges in the analysis.

Enumeration of Response Surface Designs Using Symmetry Reduction Techniques and High Throughput Computing

Jose Nunez Ares, UW Madison / KU Leuve

Coauthors: Jeffrey Linderoth (UW Madison) and Peter Goos (KU Leuven)

Abstract: Response Surface Designs are statistical designs widely used in product and process optimization. In this work, we enumerate 3-level Minimally Aliased Response Surface Designs (MARSDs), which are $\{-1, 0, 1\}$ matrices with certain desirable statistical properties. We implement a branch-and-prune algorithm in MINTO that exploits the inherent symmetry of the design matrices to efficiently enumerate all 3-level MARSDs of a given size and sparsity level. The enumeration is parallelized to run on a large-scale, heterogeneous, non-dedicated computing environment provided by HTCondor. The enumeration trees are extremely "skinny", so effective load balancing is a challenge. Our novel load-balancing scheme relies on phases of breadth-first search along with a statistical estimation of the size of the enumeration tree rooted at child nodes to achieve efficiency. We are able to find many new statistical designs of size significantly larger than previously known.

Fast First Order Methods for Escaping Saddle Points

Michael O'Neill, University of Wisconsin–Madison

Coauthor: Stephen J. Wright (Wisconsin-Madison)

Abstract: For smooth, convex optimization accelerated gradient methods obtain an optimal convergence rate of $O(1/k^2)$. When applied to nonconvex functions, variants of these methods globally converge to a first order stationary point at the same rate as gradient descent while maintaining this superior $O(1/k^2)$ rate in locally convex regions. In this work we show that an accelerated gradient method outperforms gradient descent in certain nonconvex regions as well. We demonstrate that this method escapes saddle points at a faster rate than gradient descent under certain smoothness conditions.

An Algorithm for Optimization Problems with Joint Chance Constraints using a Nonlinear Value-at-Risk Formulation

Alejandra Peña-Ordieres, Northwestern University

Coauthor: Andreas Wächter (Northwestern University)

Abstract: In this work we are presenting a nonparametric reformulation of optimizations problem with chance constraints. The probabilistic constraints are redefined in terms of Value-at-Risk (VaR) and a smoothing technique is applied on the historical empirical quantile via a Kernel. This method results in an approximation of the quantile function that reduces the nonconvexity of the feasible region derived from the empirical quantile approximation. Some computational results are included.

Improving the Expected Improvement Algorithm

Chao Qin, Northwestern University

Coauthor: Diego Klabjan and Daniel Russo (Northwestern)

Abstract: The expected improvement (EI) algorithm is a popular strategy for information collection in optimization under uncertainty. The algorithm is widely known to be too greedy, but nevertheless enjoys wide use due to its simplicity and ability to handle uncertainty and noise in a coherent decision theoretic framework. To provide rigorous insight into EI, we study its properties in a simple setting of Bayesian optimization where the domain consists of a finite grid of points. This is the so-called best-arm identification problem, where the goal is to allocate measurement effort wisely to confidently identify the best arm. In this framework, one can show formally that EI is far from optimal. To overcome this shortcoming, we introduce a simple modification of EI. Surprisingly, this simple change results in an algorithm that is asymptotically optimal for best-arm identification problems, and provably outperforms standard EI by an order of magnitude.

Complexity and Global Rates of Optimization Methods Based on Probabilistic Properties

Clément Royer, University of Wisconsin–Madison

Coauthors: Serge Gratton (University of Toulouse), Luis Nunes Vicente (University of Coimbra), and Zaikun Zhang (Hong Kong Polytechnic University)

Abstract: This poster presents a framework for analyzing the introduction of random aspects in otherwise deterministic optimization algorithms. In that setting, the classical properties that are instrumental for establishing convergence only hold with a given probability. Provided a suitable submartingale can be identified, the corresponding methods can be shown to converge with probability 1. Our analysis exploits this argument, as well as concentration inequalities, to provide complexity guarantees on these methods. We present a new complexity proof technique that enables us to derive global convergence rates and worst-case bounds, with overwhelmingly high probability. We derive the analysis in the context of trust-region methods based on probabilistic models, for both first and second order optimality measures. However, our technique is sufficiently general to apply to a wide range of optimization schemes with and without derivatives. We will provide examples of such methods, together with their probabilistic complexity results.

Optimization under Decision Dependent Uncertainty

Kartikey Sharma, Northwestern University

Coauthor: Omid Nohadani (Northwestern University)

Abstract: Robust optimization is increasingly used to solve optimization problems spread across multiple periods. In most such problems, the uncertainty sets are fixed. However in many cases, these sets can be influenced by decision variables. We present a two-period robust optimization approach in which future uncertainty sets can be affected by the decisions made in the first stage. We illustrate the advantages of this model on a shortest path problem with uncertain arc lengths.

Nested Multi-Instance Classification

Alexander Stec, Northwestern University

Coauthors: Diego Klabjan (Northwestern University) and Jean Utke (Data, Discovery & Decision Science, Allstate Insurance Company)

Abstract: There are classification tasks that take as inputs groups of images rather than single images. In order to address such situations, we introduce a nested multi-instance deep network. The approach is generic in that it is applicable to general data instances, not just images. The network has several convolutional neural networks grouped together at different stages. This primarily differs from other previous works in that we organize instances into relevant groups that are treated differently. We also introduce methods to replace instances that are missing and a manual dropout when a whole group of instances is missing. With specific pretraining, we find that the model works to great effect on our data. For our two cases with four and seven classes, we obtain 75.9% and 68.7% accuracy, respectively.

Bayes l_0 Regularized Least Squares

Lei Sun, University of Chicago

Coauthor: Nick Polson (Chicago Booth)

Abstract: Bayes l_0 regularized least squares is a variable selection technique for high dimensional inference problems. l_0-regularization is NP-hard to solve due to its non-convexity and the fact that you have to search a model space of all possible combinations. The gold standard for Bayesian sparse linear regression are spike-and-slab or Bernoulli-Gaussian (BG) priors. We explore the connection between regularized linear regression and sparsity inducing priors. We introduce a Single Best Replacement (SBR) algorithm that provides a fast sparse l_0 regularized least squares solution. We illustrate the computational efficiency that can be gained by using non-convex optimization methods rather than directly sampling the full posterior distribution with MCMC.

New Solution Approaches for the Maximum-reliability Stochastic Network Interdiction Problem

Eli Towle, University of Wisconsin–Madison

Coauthor: James Luedtke (Wisconsin-Madison)

Abstract: We investigate methods to solve the maximum-reliability stochastic network interdiction problem (SNIP). In this problem, a defender interdicts arcs on a directed graph to minimize an attacker's probability of undetected traversal through the network. The attacker's origin and destination are assumed to be random. SNIP can be formulated as a stochastic mixed-integer program via an extensive formulation. However, this formulation is not suited for solving large instances. We present two new approaches to solve SNIP. First, we introduce a significantly more compact extensive formulation. Second, we propose a path-based formulation of SNIP. This formulation is defined by exponentially many constraints, but the model can be solved via delayed constraint generation. We present valid inequalities for this new path-based formulation and propose a branch-and-cut (BC) algorithm to solve it. Computational results demonstrate that directly solving the compact formulation and the BC algorithm both provide an improvement over existing solution approaches.

An Attention-Based Deep Net for Learning to Rank

Baiyang Wang, Northwestern University

Coauthor: Diego Klabjan (Northwestern University)

Abstract: In information retrieval, learning to rank constructs a machine-based ranking model which given a query, sorts the search results by their degree of relevance or importance to the query. Neural networks have been successfully applied to this problem, and in this poster, we propose an attention-based deep neural network which better incorporates different embeddings of the queries and search results with an attention-based mechanism. This model also applies a decoder mechanism to learn the ranks of the search results in a listwise fashion. The embeddings are trained with convolutional neural networks or the word2vec model. We demonstrate the performance of this model with image retrieval and text querying data sets.

A Defender-attacker Model for Mitigating Cyber Risks in the Supply Chain

Kay Zheng, University of Wisconsin–Madison

Coauthor: Laura Albert McLay (Wisconsin-Madison)

Abstract: The globalization of supply chain introduces numerous risks to the Information and Communication Technology (ICT) infrastructure. Adversarial attacks, especially, provide a significant risk to the ICT supply chain. They are sophisticated and hard to detect, therefore preventive measures play a very important role in protection planning. In this paper, we study how to prioritize security investment in mitigations to protect ICT supply chain from adaptive adversaries. We propose new Stackelberg game models with the defender in the first stage deploys security mitigations that affect exploits of vulnerabilities in the system, and multiple attackers in the second stage aim to complete their attack goals through a critical path respectively. Additionally, we propose a stochastic model variant that address uncertainty regarding mitigation effectiveness. A Lagrangian heuristic is proposed that solves medium to large problem instances efficiently with a tight upper and lower bounds.

Further Details

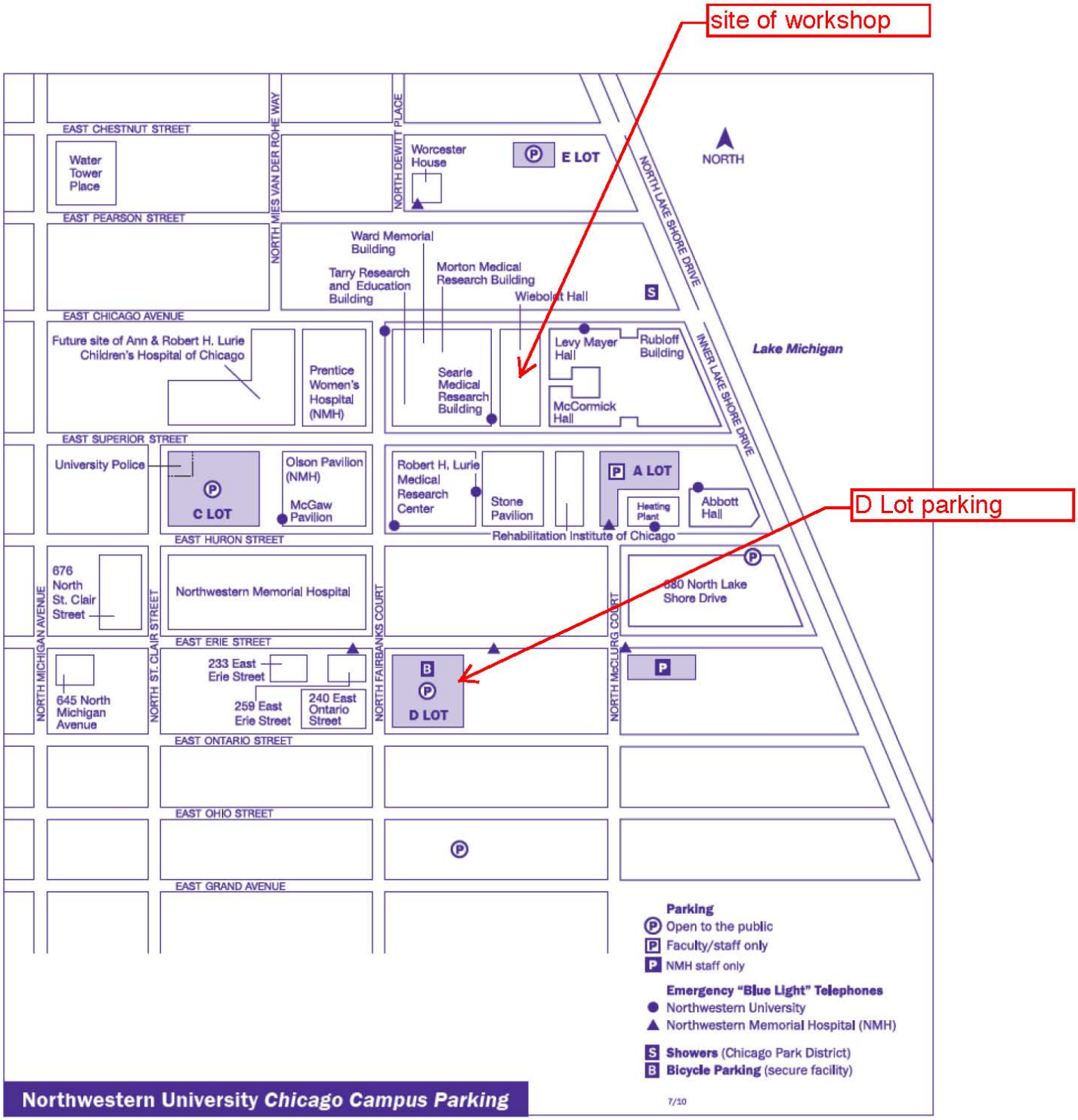
- There is no registration charge. Breakfast, lunch, and dinner will be provided as part of the program. Thank you to the Center for Optimization & Statistical Learning at Northwestern http://www.osl.northwestern.edu.
- If you want a parking pass, we can arrange for one for the garage on the map (see below). Note: You do not need the pass to enter the garage, only when you depart. However, we need to order the passes ahead of time, and if you didn't request one when registering, please email Tara Sadera tara.sadera@northwestern.edu.
- If you registered to present a poster, you will attach a paper poster to a 30" $\times$ 40" posterboard, which will be provided, as will pushpins. The posters will be presented on easels.

Program Committee

Jim Luedtke, University of Wisconsin–Madison

Mihai Anitescu, Argonne National Laboratory

Dave Morton, Northwestern University


site of workshop
D Lot parking
EAST CHESTNUT STREET
EAST PEARSON STREET
EAST CHICAGO AVENUE
EAST SUPERIOR STREET
EAST HURON STREET
EAST ERIE STREET
EAST ONTARIO STREET
EAST OHIO STREET
EAST GRAND AVENUE
NORTH MIES VAN DER ROHE WAY
NORTH DEWITT PLACE
NORTH LAKE SHORE DRIVE
INNER LAKE SHORE DRIVE
NORTH MICHIGAN AVENUE
NORTH ST. CLAIR STREET
NORTH FAIRBANKS COURT
NORTH McCLURG COURT
NORTH
Lake Michigan
Water Tower Place
Worcester House
E LOT
Ward Memorial Building
Tarry Research and Education Building
Morton Medical Research Building
Wiebolt Hall
Future site of Ann & Robert H. Lurie Children's Hospital of Chicago
Prentice Women's Hospital (NMH)
Searle Medical Research Building
Levy Mayer Hall
Rubloff Building
McCormick Hall
University Police
C LOT
Olson Pavilion (NMH)
McGaw Pavilion
Robert H. Lurie Medical Research Center
Stone Pavilion
A LOT
Heating Plant
Abbott Hall
Rehabilitation Institute of Chicago
676 North St. Clair Street
Northwestern Memorial Hospital
680 North Lake Shore Drive
645 North Michigan Avenue
233 East Erie Street
259 East Erie Street
240 East Ontario Street
D LOT
Parking
Open to the public
Faculty/staff only
NMH staff only
Emergency "Blue Light" Telephones
Northwestern University
Northwestern Memorial Hospital (NMH)
Showers (Chicago Park District)
Bicycle Parking (secure facility)
Northwestern University Chicago Campus Parking
7/10

FREE WIRELESS ACCESS

FOR NORTHWESTERN UNIVERSITY GUESTS

1. **CHOOSE** the "Guest-Northwestern" SSID from the list of available networks on your laptop, tablet, or smartphone.
2. **ENTER** your contact information and sponsoring individual/campus organization.
3. **READ** and **ACCEPT** the University's acceptable use policy.
4. **SELECT** "Register" to connect.

Please note: "Guest-Northwestern" is an unsecured connection to the Internet and should not be used to send or receive sensitive information.

www.it.northwestern.edu

Have Questions?

Contact the NUIT Support Center at 847-491-HELP (4357)

NORTHWESTERN UNIVERSITY

Please enter David Morton as "Sponsor Name." Note that eduroam is also available.

KELLOGG CONFERENCE CENTER

312-503-6100
kcc@kellogg.northwestern.edu

Steve Peavler
Assistant Director Wieboldt Hall
Kellogg School of Management
340 East Superior Street
Chicago, IL 60611
312-503-0021
312-503-6101 Fax

Theresa Kleptach
Director of Catering
Kellogg School of Management
340 East Superior Street
Chicago, IL 60611
312-503-1478
312-503-6101 Fax

The following hotels offer special discounts to Northwestern University guests. Kellogg staff recommends the MileNorth Hotel, a four-star business hotel located only one-half block from Wieboldt Hall.

Recommended:
MileNorth Hotel
166 East Superior
Chicago, Illinois 60611
Main Line - (312) 787-6000 - *Ask for Northwestern University guest rate.*

Distance from Midway: 26 minutes/13.71 miles
Distance from O'Hare: 29 minutes/18.36 miles
Distance from training venue: ½ block

Other hotels offering Northwestern University discounts:
The Allerton Hotel Chicago
701 North Michigan Avenue
Chicago, IL 60611
Main Line: 312-440-1500

Club Quarters
Chicago, Wacker at Michigan
75 East Wacker Drive
Chicago, Illinois 60601
Guest Services: 203-905-2100

DoubleTree Hotel Chicago-Magnificent Mile
300 East Ohio Street
Chicago, IL 60611
Reservations: 866-778-8536
Main Line: 312-787-6100

Hyatt Chicago Magnificent Mile (Formerly Wyndham Chicago Hotel)
633 North St. Clair
Chicago, IL 60611
312-274-4435

Omni Chicago Hotel
676 North Michigan Avenue
Chicago, IL 60611
312-944-6664

The Peninsula Chicago
108 East Superior Street
(at North Michigan Avenue)
Chicago, IL 60611
312-337-2888

KELLOGG CONFERENCE CENTER

312-503-6100
kcc@kellogg.northwestern.edu

Steve Peavler
Assistant Director Wieboldt Hall
Kellogg School of Management
340 East Superior Street
Chicago, IL 60611
312-503-0021
312-503-6101 Fax

Theresa Kleptach
Director of Catering
Kellogg School of Management
340 East Superior Street
Chicago, IL 60611
312-503-1478
312-503-6101 Fax

Other hotels nearby

Courtyard Chicago Downtown/Magnificent Mile
165 East Ontario Street
Chicago, IL 60611
Main Line (312) 329-2500

Public Chicago
1301 North State Parkway
Chicago, IL 60610
Main Line: 312-787-3700 or 888-506-3471

The Ritz-Carlton
160 East Pearson
At Water Tower Place
Chicago, Illinois 60611
312-266-1000

Sheraton Chicago Hotel and Towers
301 East North Water Street
Chicago, IL 60611
Main Line (312) 464-1000

The Talbott Hotel
20 East Delaware Place
Chicago, Illinois 60611
direct 312-397-3619 | hotel 312-944-4970
fax 312-475-5419

Westin Chicago River North
320 North Dearborn
Chicago, IL 60654
Main Line (312) 744-1900