ACNW Optimization Workshop

June 8, 2015

Northwestern's Kellogg Conference Center

340 E Superior St, Chicago, IL 60611

Wieboldt Hall, Room 250

Program Overview

9:30-10:00am	Breakfast available
10:00-11:00am	Brief welcome and Dimitris Bertsimas
11:00-11:30am	Stefan Wild
11:30-12:00pm	Michael Ferris
12:00-1:30pm	Lunch (Room 323)
1:30-2:00pm	Burhan Sandıkçı
2:00-2:30pm	Andreas Wächter
2:30-3:00pm	Break
3:00-3:30pm	Alberto Del Pia
3:30-4:00pm	Mihai Anitescu
4:00-5:30pm	Poster session
6:00pm	Dinner at Quartino's

All lectures are in Room 250 of Wieboldt Hall. The poster session takes place in Room 323.

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.

10:00am, Keynote Lecture

From predictive to prescriptive analytics

Dimitris Bertsimas, Operations Research Center and Sloan School of Management, MIT

Abstract: Operations Research and Management Science (OR/MS) as a field typically starts with models and aims to obtain decisions. Data by enlarge is rarely present. Machine learning (ML) as a field typically starts with data and aims to make predictions. Decisions are rarely addressed. In this work, we combine ideas from ML and OR/MS in developing a framework, along with specific methods, for using data to prescribe optimal decisions in OR/MS problems. In a departure from other work on data-driven optimization and reflecting our practical experience with the data available in applications of OR/MS, we consider data consisting, not only of observations of quantities with direct effect on costs/revenues, such as demand or returns, but predominantly of observations of associated auxiliary quantities. The main problem of interest is a conditional stochastic optimization problem, given imperfect observations, where the joint probability distributions that specify the problem are unknown.

We demonstrate that our proposed solution methods, which are inspired by ML methods such as local regression (LOESS), classification and regression trees (CART), and random forests (RF), are generally applicable to a wide range of decision problems. We prove that they are computationally tractable and asymptotically optimal under mild conditions even when data is not independent and identically distributed (iid) and even for censored observations. As an analogue to the coefficient of determination R^2 we develop a metric P termed the coefficient of prescriptiveness to measure the prescriptive content of data and the efficacy of a policy from an operations perspective.

To demonstrate the power of our approach in a real-world setting we study an inventory management problem faced by the distribution arm of an international media conglomerate, which ships an average of 1 billion units per year. We leverage both internal data and public online data harvested from IMDb, Rotten Tomatoes, and Google to prescribe operational decisions that outperform baseline measures. Specifically, the data we collect, leveraged by our methods, accounts for an 88% improvement as measured by our coefficient of prescriptiveness. Joint work with Nathan Kallus, MIT.

Bio: Dimitris Bertsimas is currently the Boeing Professor of Operations Research and the co-director of the Operations Research Center at MIT. He received his SM and PhD in Applied Mathematics and Operations Research from MIT in 1987 and 1988 respectively. He has been with the MIT faculty since 1988. His research interests include optimization, statistics and applied probability and their application in health care, finance, operations management and transportation. He has co-authored more than 160 scientific papers and three graduate level textbooks. His fourth book "The Analytics Edge" will be published this summer. He is former department editor in Optimization for *Management Science* and in Financial Engineering in *Operations Research*. He has supervised 53 doctoral students and he is currently supervising 18 others. He is a member of the National Academy of Engineering since 2005, and he has received numerous research awards including the Morse prize (2013), the Pierskalla award (2013), the best paper award in Transportation (2013), the Farkas prize (2008), the Erlang prize (1996), the SIAM prize in optimization (1996), the Bodossaki prize (1998) and the Presidential Young Investigator award (1991-1996). He has co-founded several companies in the areas of financial services, health care, aviation and publishing.

Lectures

11:00am

Optimization formulations and algorithms for diffractive imaging

Stefan Wild, Argonne National Laboratory

Abstract: We examine coherent X-ray diffraction imaging, which can be viewed as an underdetermined inverse problem involving experimentally measured data and physics-based constraints on the sampled object. We develop nonlinear optimization formulations for these kinds of structured phase retrieval problems. We discuss projection-based algorithms and alternating direction methods for solving these formulations, as well as heuristics for guarding against the multimodality of the associated objectives.

11:30am

Modeling, equilibria, power and risk

Michael Ferris, University of Wisconsin–Madison

Abstract: We look at models of competition and risk within the context of power system markets. We demonstrate when social optima exist, what properties on risk measures and contracts are needed, and how to solve these problems in large scale practical settings. We position these models within a general setting of Nash Games that include linking equilibrium constraints and situations where players solve multi-period stochastic programs.

1:30pm

A scalable approach to solving multi-stage stochastic integer programs

Burhan Sandıkçı, University of Chicago

Abstract: Multi-stage stochastic mixed-integer programs (SMIPs) are very appealing as a flexible modeling framework, but they are not widely adopted in practice, mostly due to their unbearable size for most practically relevant problems. We attempt to build on the powers of available optimization software and the parallel computing architectures to solve large-scale multi-stage SMIPs. In particular, we propose a bounding approach, which does not assume convexity but it rather relies on scenario decomposition and is inherently parallelizable with minimum communication overhead, and therefore, makes it amenable to solving truly large-scale instances. To demonstrate the power of our approach, we utilize problems from the literature with one modification: the sizes of the instances we use to test our approach are several orders of magnitude larger than their original versions. Our preliminary results demonstrate that the proposed method scales nicely with problem size and produces high quality solutions within a reasonable time.

2:00pm

A derivative-free method for the optimization of functions smoothed via Gaussian convolution using multiple importance sampling

Andreas Wächter, Northwestern University

Abstract: The optimization of an objective function that is computed by a complicated numerical procedure, such as the solution of differential equations, bears several challenges. For one, derivatives of the objective function, which are required by many efficient optimization methods, can often not be computed efficiently. To overcome this hurdle, a number of Derivative-Free Optimization (DFO) methods have been proposed in the past. The second challenge concerns the fact that the numerical procedure for the objective function computation can cause numerical noise, which manifests itself in discontinuities in the objective values. This noise can, for example, be the consequence of if-statements in the code that computes the objective function.

In this presentation we propose a new trust-region DFO method that addresses the presence of numerical noise by smoothing the objective with a Gaussian kernel. The trust-region model is computed by regression, using function values at randomly generated sample points around iterates and trial points. Multiple importance sampling makes it possible to reuse objective function evaluations across iterations. Numerical results on a set of test problems will be presented. This research is conducted in collaboration with Alvaro Maggiar, Irina Dolinskaya, and Jeremy Staum.

3:00pm

Minimizing polynomials over integers

Alberto Del Pia, University of Wisconsin–Madison

Abstract: Integer Polynomial Programming (IPP) is the natural extension of Integer Linear Programming obtained by allowing polynomial objective functions. More precisely, an IPP problem is the problem of minimizing a polynomial function over the integer points in a polyhedral region. Surprisingly, only recently the problem was shown to be NP-complete. We survey some recent results about IPP, including a complexity classification by degree of the minimization problem in dimension two.

3:30pm

Scalable Gaussian process analysis

Mihai Anitescu, Argonne National Laboratory

Abstract: We present a new scalable approach for Gaussian process fitting. The approach is based on a recently introduced unbiased stochastic approximation to the score equations for maximum likelihood calculation for Gaussian processes. Under certain conditions, including bounded condition number of the covariance matrix, the approach achieves $O(n)$ storage and nearly $O(n)$ computational effort per optimization step, where n is the number of data sites. Moreover, we prove that if the condition number of the covariance matrix is bounded, then the approximate score equations are nearly optimal in a well-defined sense. Our findings are validated by numerical experiments on simulated datasets of up to 1 million observations. We also report the performance and outcome of the approach to fit a space-time model to over 80,000 observations of total column ozone contained in the latitude band 40-50 degrees N during April 2012.

Poster Session

Tight second-stage formulations in two-stage stochastic mixed integer programs

Manish Bansal and Kuo-Ling Huang, Northwestern University

Abstract: We consider two-stage stochastic mixed integer programs (TSS-MIPs) with integer variables in the second stage and show that under suitable conditions, the second stage mixed integer programs can be convexified. We introduce a new sufficient condition under which a general polyhedra is integral. These theoretical results allow us to relax the integrality constraints in the second stage in certain situations. As special cases, we consider TSS-MIPs which have generalization of the well-known mixing (and continuous mixing) set in the second stage. Furthermore, we showcase how the convexification approach can also be utilized to solve general two-stage stochastic programs. We then use our results to develop valid parametric inequalities for four variants of the two-stage stochastic capacitated lot-sizing problem and present our computational results.

A multi-batch L-BFGS method for machine learning

Albert S. Berahas, Northwestern University

Abstract: We present a robust *multi-batch* L-BFGS implementation for large scale machine learning in which the samples used at each iteration change, but do so in a controlled way. Our proposed sampling strategy enforces that consecutive batches share a certain number of samples. Imposing such a condition allows for consistent curvature pair updates and thus reliable search directions, and by extension stable iterates. As a result, the proposed L-BFGS implementation, contingent on a reliable line search and sufficiently large batches, is robust with respect to the two adverse situations: disappearing nodes and changing batches, that often arise in such optimization implementations. The delicate issue of the line search is also investigated; we propose a backtracking line search based on the idea of confidence intervals and mention several other heuristics that can be employed. We present numerical results on several well known binary classification datasets such as rcv1, url, gisette as well as a Google speech dataset. The results suggest that our proposed *multi-batch L-BFGS* implementation shows much promise due to its high initial learning rate. On a final note, we provide a few remarks about several extensions of the implementation as well as ideas for parallelization. Joint work with Jorge Nocedal.

Cutting planes from extended LP formulations

Merve Bodur, University of Wisconsin–Madison

Abstract: We study extended formulations of polyhedral sets and the effect of applying split cuts to these sets. We observe that applying split cuts to extended formulations can be more effective than applying split cuts to the original formulation. Therefore, we first investigate structural properties of useful extended formulations to derive split cuts. For binary sets, we show that applying all elementary split cuts to a specific extended formulation yields the integer hull. We extend this idea to general mixed-integer sets and construct the best extended formulation for such sets in terms of split cuts. Lastly, we observe that Lovasz-Schrijver and Sherali-Adams hierarchies can be viewed as applying specific split cuts to appropriate extended formulations, and demonstrate how to strengthen these hierarchies using additional split cuts.

An inertia-free filter line-search algorithm

Naiyuan Chiang, Argonne National Laboratory

Abstract: We present a filter line-search algorithm for NLP, which does not require inertia information about the linear system. This inertia-free algorithm can enhance the modularity of the implementation, and it can solve the problems in which inertia information is not available. We also provide some numerical evidence to show the robustness of the new algorithm.

Branch-cut-and-price for the chance-constrained vehicle routing problem with correlated stochastic demands

Thai Dinh, University of Wisconsin–Madison

Abstract: We study a chance-constrained model for the vehicle routing problem with stochastic demands, in which a limit is imposed on the probability that each vehicle's capacity is exceeded. We first derive a valid edge-based formulation for the problem, using a lower bound on the minimum number of vehicles required to serve a subset of customers to adapt the well known rounded capacity cuts. We then present a branch-cut-and-price solution framework that requires mild assumptions on the random demands. In particular, the framework can solve problems, in which random demands are represented by a scenario model, where the scenarios could be obtained as a sample from any distribution. Columns are generated using a dynamic programming algorithm executed over small-cycle-free q-routes with a relaxed capacity constraint. Computational experiments will be presented to assess the effectiveness of the approach.

Interface between operator overloading and source transformation tools

Paul Hovland, Argonne National Laboratory

Abstract: Operator overloading algorithmic differentiation tools are appropriate for languages such as C++ because they exploit the features of the language to compute the adjoints. However, such tools may require extreme amounts of storage rendering AD of an application infeasible. Source transformation AD tools cannot handle complex features of C++ but can provide efficient derivatives for simple portions of an application. We have created an interface between the operator overloading tool ADOL-C and the source transformation tool ADIC2. ADOL-C is used to differentiate the portions of the code that exploit features of the code such as classes and templates. Portions of the code that involve computation written in a C-like manner are differentiated using ADIC2. We will present the design of interface and demonstrate it use on an application.

A modified DC algorithm for solving linear programs with equilibrium constraints

Francisco Jara-Moroni, Northwestern University

Abstract: We propose a method for finding local optima of linear programs with equilibrium constraints. The complementarity restriction is handled by a penalty term that can be expressed as the difference of convex functions. The reformulated problem is solved to optimality by the difference-of-convex functions algorithm with some variations exploiting the specific structure of the penalization.

A second-order method for convex ℓ_1-regularized optimization with active set prediction

Nitish Keskar, Northwestern University

Abstract: We describe an active-set method for the minimization of an objective function ϕ that is the sum of a smooth convex function and an ℓ_1-regularization term. A distinctive feature of the method is the way in which active-set identification and second-order subspace minimization steps are integrated to combine the predictive power of the two approaches. At every iteration, the algorithm selects a candidate set of free and fixed variables, performs an (inexact) subspace phase, and then assesses the quality of the new active set. If it is not judged to be acceptable, then the set of free variables is restricted and a new active-set prediction is made. We establish global convergence for our approach, and compare the new method against the state-of-the-art code LIBLINEAR.

A method for generating twice-continuously differentiable convex underestimators

Kamil Khan, Argonne National Laboratory

Abstract: Several deterministic methods for nonconvex optimization require lower bounding information that is obtained by minimizing convex underestimators of objective functions on appropriate subdomains. McCormick's relaxation scheme is an automatable and efficient method for generating appropriate convex underestimators for composite functions. However, the underestimators generated by this scheme may be nonsmooth, even if the original objective function is smooth. By modifying McCormick's treatment of products and compositions, we show that a variant of McCormick's scheme produces convex underestimators that are always twice-continuously differentiable, without sacrificing any of the useful properties of the traditional McCormick underestimators. Gradients may be computed efficiently for these differentiable underestimators using the reverse mode of automatic differentiation.

Algorithmic innovations in dual decomposition method for stochastic mixed-integer programs

Kibaek Kim, Argonne National Laboratory

Abstract: We develop algorithmic innovations for the dual decomposition method to address two-stage stochastic programs with mixed-integer recourse and provide a parallel software implementation that we call DSP . Our innovations include the derivation of valid inequalities that tighten Lagrangian subproblems and that guarantee the recovery of upper bounds for problems with (relatively) incomplete recourse. We also stabilize dual variables by solving the Lagrangian master problem with a primal-dual interior point method and provide termination criteria that guarantee finite termination of the algorithm. DSP can solve instances specified in C code, SMPS files (a standard format for stochastic programming), and StochJump (a Julia-based algebraic modeling language).

Finding multiple local minima for computationally expensive simulations

Jeff Larson, Argonne National Laboratory

Abstract: We present a multistart algorithm for finding high-quality local minima of computationally expensive simulations. Under relatively weak assumptions, the algorithm can be shown to almost surely find all local minima while only starting a finite number of local optimization runs. The algorithm is shown to be successful at finding multiple minima without degrading its ability to approximate the global minimum.

A trust-region-based approximate gradient sampling algorithm for unconstrained ℓ_1 minimization

Matt Menickelly, Argonne National Laboratory and Lehigh University

Abstract: We consider the minimization of a sum of absolute values of outputs of black box functions. Rather than consider a smoothing of these ℓ_1 functions, we develop a new algorithm using the model-based trust region framework of derivative-free optimization (DFO), but incorporate more recent ideas of approximate gradient sampling for nonsmooth optimization. In this algorithm, we iteratively solve smooth trust-region subproblems involving a "master model" defined using approximate gradient information. We prove that this algorithm converges to a Clarke stationary point almost surely and provide some preliminary numerical results.

Efficient training of structured SVMs via soft constraints

Ofer Meshi, Toyota Technological Institute at Chicago

Abstract: Structured output prediction is a powerful framework for jointly predicting interdependent output labels. Learning the parameters of structured predictors is a central task in machine learning applications. However, training the model from data often becomes computationally expensive. Several methods have been proposed to exploit the model structure, or decomposition, in order to obtain efficient training algorithms. In particular, methods based on linear programming relaxation, or dual decomposition, decompose the prediction task into multiple simpler prediction tasks and enforce agreement between overlapping predictions. In this work we observe that relaxing these agreement constraints and replacing them with soft constraints yields a much easier optimization problem. Based on this insight we propose an alternative training objective, analyze its theoretical properties, and derive an algorithm for its optimization. Our method, based on the Frank-Wolfe algorithm, achieves significant speedups over existing state-of-the- art methods without hurting prediction accuracy.

The Boolean quadric polytope for graphs with bounded treewidth

Carla Michini, University of Wisconsin–Madison

Abstract: In this work we study the problem of minimizing a quadratic function over binary vectors. The boolean quadric polytope arises from a standard linearization of the objective function and has been extensively studied in the literature. We focus on problems where the quadratic form yields a graph with special structure, and we exploit such structure to find tight relaxations. While in general the boolean quadric polytope is known to admit no compact linear extended formulation, for graphs with bounded treewidth we derive an extended formulation of polynomial size. This is a joint work with James Luedtke.

Reverse mode algorithmic differentiation for adjoints

Krishna Narayanan, Argonne National Laboratory

Abstract: Reverse mode algorithmic differentiation computes the adjoints of codes precisely and efficiently. The reverse mode employs checkpointing to store data between its forward and the reverse computational sweeps. When the code contains fixed point iterations, unnecessary checkpointing can result in excessive memory or disk usage. Reformulating the adjoint of the fixed point iteration, drastically reduces the usage. An implementation of the reformulation and its use in an ice sheet model will be presented.

A new Lagrangian approach for weakly coupled stochastic dynamic programs

Jagdish Ramakrishnan, University of Wisconsin–Madison

Abstract: Current approaches for solving multi-stage stochastic mixed-integer programs are limited in their ability to handle multiple sample paths per stage, due to an exponential growth in the number of sample paths. We consider a weakly coupled model, which is amenable to decomposition into small Markov Decision Problems via Lagrangian relaxation. We extend previous work on this approach by allowing the Lagrange multipliers to depend on the observation history. The approach will be illustrated numerically for the stochastic unit commitment problem.

A guaranteed, adaptive algorithm for univariate function minimization

Xin Tong, University of Illinois at Chicago

Abstract: We describe a new adaptive algorithm `funmin_g` for univariate function minimization on a bounded interval. A key feature of `funmin_g` is that it is guaranteed to return the global minimum and its locations within the user-specified error tolerances. The functions need not be convex. The guarantee requires the function being minimized to be smooth (twice differentiable) and lie inside a cone of functions with limited spikiness. Our new algorithm is illustrated by means of numerical examples and compared with MATLABs `fminbnd`. This is joint work with Fred Hickernell and Sou-Cheng Choi from the Department of Applied Mathematics, Illinois Institute of Technology.

Semi-supervised learning for discrete choice models

Jie Yang, Northwestern University

Abstract: We propose a new approach to estimate discrete choice model (DCM) when part of the data comes with choices and part of the data comes without choices. We start with two classic semi-supervised learning (SSL) algorithm, cluster-and-label and expectation maximization (EM) and then develop two new algorithms which determines the number of clusters automatically. A computational study has been conducted based on a public hotel data set and the performance of algorithms is presented and analyzed. We also report findings based on real world air travel shopping data from a major global distribution system provider. Joint work with Sergey Shebalov and Diego Klabjan.

Further Details

- There is no registration charge, and breakfast and lunch will be provided as part of the program: Thank you to the Department of Industrial Engineering & Management Sciences at Northwestern.
- We ask participants to cover the cost of dinner ($\approx$ \$40 without drinks).
- If you want a parking pass, we can arrange for one for the garage on the map (see page 12). Note: You do not need the pass to enter the garage, only when you depart, but we need to order the passes ahead of time.
- Please register for dinner and a parking pass by May 31st: http://doodle.com/t4smtiey84gamugu
- 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.

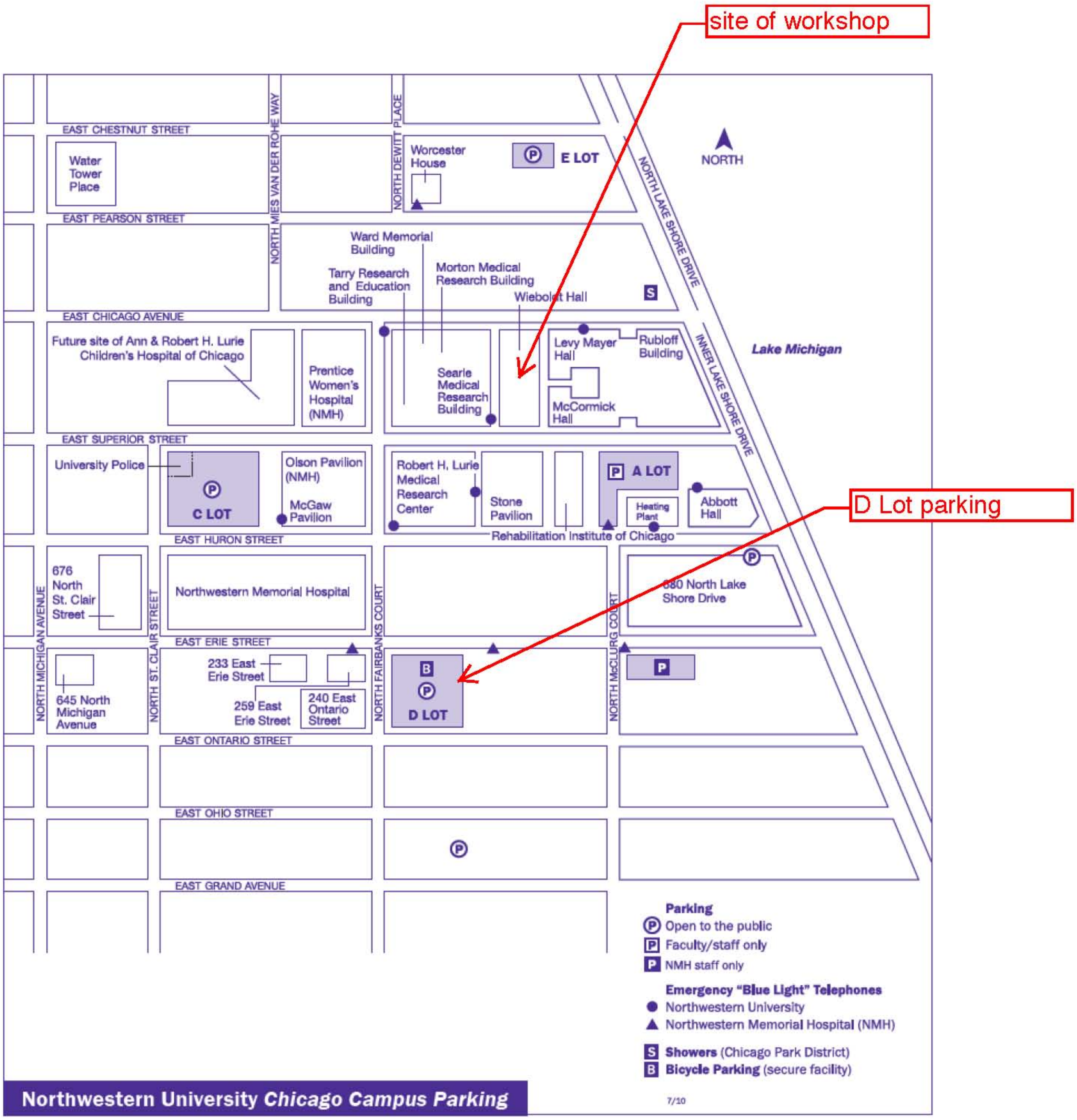
Program Committee

Jim Luedtke, University of Wisconsin–Madison

Dave Morton, Northwestern University

Burhan Sandıkçı, University of Chicago

Victor Zavala, Argonne National Laboratory



Please register for a reduced-rate parking pass by May 31st: http://doodle.com/t4smtiey84gamugu

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