USC University of Southern California

The Daniel J. Epstein Department of Industrial and Systems Engineering

Conference on Nonconvex Statistical Learning

Vineyard Room, Davidson Conference Center
University of Southern California
Los Angeles, California

Friday May 26, 2017 and Saturday May 27, 2017

Sponsored by: The Division of Mathematical Sciences at the National Science Foundation; The Epstein Institute; and the Daniel J. Epstein Department of Industrial and Systems Engineering, Univeristy of Southern California

Conference Program

Friday May 26, 2017

07:15 – 08:00 AM	Registration and continental breakfast
08:00 – 08:15	Opening

The schedule of the talks is arranged alphabetically according to the last names of the presenters.

Chair:	**Jong-Shi Pang**
08:15 – 08:45 AM	Amir Ali Ahmad
08:45 – 09:15	Meisam Razaviyayn (substituting Andrea Bertozzi)
09:15 – 09:45	Hongbo Dong
09:45 – 10:15	Ethan Fang
10:15 – 10:45	**break**
Chair:	**Jack Xin**
10:45 – 11:15	Xiaodong He
11:15 – 11:45	Mingyi Hong
11:45 – 12:15 PM	Jason Lee
12:15 – 1:45 PM	lunch break
Chair:	**Phebe Vayanos**
01:45 – 02:15 PM	Po-Ling Loh
02:15 – 02:45	Yifei Lou
02:45 – 03:15	Shu Lu
03:15 – 03:45	Yingying Fan (substituting Zhi-Quan Luo)
03:45 – 04:15	**break**
Chair:	**Meisam Razaviyayn**
04:15 – 04:45	Jinchi Lv and Yingying Fan
04:45 – 05:15	Rahul Mazumder
05:15 – 05:45 PM	Andrea Montanari

06:45 – 09:30 PM **Dinner for speakers and invited guests only at University Club**

Conference Program (continued)

Saturday May 27, 2017

07:30 – 08:15 AM	Continental breakfast
Chair:	**Yufeng Liu**
08:15 – 08:45 AM	Gesualdo Scutari
08:45 – 09:15	Mahdi Soltanolkotabi
09:15 – 09:45	Defeng Sun
09:45 – 10:15	Qiang Sun
10:15 – 10:45	**break**
Chair:	**Jack Xin**
10:45 – 11:15	Akiko Takeda
11:15 – 11:45	Mengdi Wang
11:45 – 12:15 PM	Steve Wright
12:15 – 1:45	**lunch break**
Chair:	**Jong-Shi Pang**
1:45 – 2:15	Lingzhou Xue
2:15 – 2:45	Wotao Yin
2:45 – 3:15 PM	Yufeng Liu
3:15 – 3:30 PM	**closing**

A special issue of *Mathematical Programming, Series B* will be guest edited by Jong-Shi Pang, Yufeng Liu, and Jack Xin on the topics of this Conference. Speakers and participants are invited to submit their papers for consideration of publication in this volume. All papers will be rigorously refereed to conform with the high standard of the journal.

Titles/Abstracts/Short bios of presenters

Title: Nonnegative polynomials, nonconvex polynomial optimization, and applications to learning

Speaker: **Amir Ali Ahmadi**, Department of Operations Research and Financial Engineering Princeton University, Princeton, New Jersey 08544, U.S.A. Email: a_a_a@Princeton.edu.
Webpage: http://aaa.princeton.edu/

Abstract: The problem of recognizing nonnegativity of a multivariate polynomial has a celebrated history, tracing back to Hilberts 17th problem. In recent years, there has been much renewed interest in the topic because of a multitude of applications in applied and computational mathematics and the observation that one can optimize over an interesting subset of nonnegative polynomials using "sum of squares (SOS) optimization". In this talk, we give a brief overview of the developments in this field and show how they can be applied to two problems at the interface of machine learning and polynomial optimization. In part (i), we study the problem of learning a monotone polynomial from data. This is motivated by regression problems where the underlying function to be learned is monotone (consider, e.g., the price of a car as a function of its fuel efficiency). In part (ii), we study the problem of optimally decomposing a multivariate polynomials as the difference of two convex polynomials. This is motivated by certain majorization-minimization algorithms used in nonconvex optimization that require such a decomposition.

Short bio: Amir Ali Ahmadi is an Assistant Professor at the Department of Operations Research and Financial Engineering at Princeton University and an Associated Faculty member of the Department of Computer Science. Amir Ali received his PhD in EECS from MIT and was a Goldstine Fellow at the IBM Watson Research Center prior to joining Princeton. His research interests are in optimization theory, computational aspects of dynamics and control, and algorithms and complexity. Amir Ali's distinctions include the Sloan Fellowship in Computer Science, the NSF CAREER Award, the AFOSR Young Investigator Award, the DARPA Faculty Award, the Google Faculty Award, the Goldstine Fellowship of IBM Research, and the Oberwolfach Fellowship of the NSF. An undergraduate course of his at Princeton has received the 2017 Excellence in Teaching of Operations Research Award of the Institute for Industrial and Systems Engineers. Amir Ali is also the recipient of a number of best-paper awards, including the INFORMS Computing Society Prize (for best series of papers at the interface of operations research and computer science), the Best Conference Paper Award of the IEEE International Conference on Robotics and Automation, and the prize for one of two most outstanding papers published in the SIAM Journal on Control and Optimization in 2013–2015.

Title: Geometric graph-based methods for high-dimensional data

Speaker: **Andrea Bertozzi**, Department of Mathematics, University of California Los Angeles Los Angeles, California 90095-1555, U.S.A. (cancelled) Email: bertozzi@math.ucla.edu.
Webpage: http://www.math.ucla.edu/~bertozzi/

Abstract: This talk addresses methods for segmentation of large datasets with graph based structure. The method combines ideas from classical nonlinear PDE-based image segmentation with fast and accessible linear algebra methods for computing information about the spectrum of the graph Laplacian. The goal of the algorithms is to solve semi-supervised and unsupervised graph cut optimization problems. I will present results for image processing applications such as image labeling and hyperspectral video segmentation, and results from machine learning and community detection in social networks, including modularity optimization posed as a graph total variation minimization problem. I will also discuss uncertainty quantification methods for such data classification problems. The algorithms typically involve nonconvex functionals with local optimization strategies however there are connections to convex optimization problems and global optimization strategies.

Short bio: Andrea Bertozzi is an applied mathematician with expertise in nonlinear partial differential equations and fluid dynamics. She also works in the areas of geometric methods for image processing, crime modeling and analysis, and swarming/cooperative dynamics. Bertozzi completed all her degrees in Mathematics at Princeton. She was an L.E. Dickson Instructor and NSF Postdoctoral Fellow at the University of Chicago from 1991–1995, and the Maria Geoppert-Mayer Distinguished Scholar at Argonne National Laboratory from 1995–6. She was on the faculty at Duke University from 1995–2004 first as Associate Professor of Mathematics and then as Professor of Mathematics and Physics. She has served as the Director of the Center for Nonlinear and Complex Systems while at Duke. Bertozzi moved to UCLA in 2003 as a Professor of Mathematics. Since 2005 she has served as Director of Applied Mathematics, overseeing the graduate and undergraduate research training programs at UCLA. In 2012 she was appointed the Betsy Wood Knapp Chair for Innovation and Creativity. Bertozzi's honors include the Sloan Research Fellowship in 1995, the Presidential Early Career Award for Scientists and Engineers in 1996, and SIAM's Kovalevsky Prize in 2009. She was elected to the American Academy of Arts and Sciences in 2010 and to the Fellows of the Society of Industrial and Applied Mathematics (SIAM) in 2010. She became a Fellow of the American Mathematical Society in 2013 and a Fellow of the American Physical Society in 2016. She won a SIAM outstanding paper prize in 2014 with Arjuna Flenner, for her work on geometric graph-based algorithms for machine learning. Bertozzi is a Thomson-Reuters 'highly cited' Researcher in mathematics for both 2015 and 2016, one of about 100 worldwide in her field.

Bertozzi has served on the editorial boards of fourteen journals: SIAM Review, SIAM J. Math. Anal., SIAM's Multiscale Modeling and Simulation, Interfaces and Free Boundaries, Applied Mathematics Research Express (Oxford Press), Applied Mathematics Letters, Mathematical Models and Methods in the Applied Sciences (M3AS), Communications in Mathematical Sciences, Nonlinearity, and Advances in Differential Equations, Journal of Nonlinear Science, Journal of Statistical Physics, Nonlinear Analysis Real World Applications; and the J. of the American Mathematical Society. She served as Chair of the Science Board of the NSF Institute for Computational and Experimental Research in Mathematics at Brown University from 2010–2014 and previously on the board of the Banff International Research Station. She served on the Science Advisory Committee of the Mathematical Sciences Research Institute at Berkeley from 2012-2016. To date she has graduated 31 PhD students and has mentored 39 postdoctoral scholars.

Title: Structural properties of affine sparsity constraints

Speaker: Hongbo Dong, Department of Mathematics and Statistics, Washington State University, Pullman, Washington 99164-3113, U.S.A. Email: hongbo.dong@wsu.edu.
Webpage: http://www.math.wsu.edu/faculty/hdong/welcome.php

Abstract: We introduce a new constraint system for sparse variable selection in statistical learning. Such a system arises when there are logical conditions on the sparsity of certain unknown model parameters that need to be incorporated into their selection process. Formally, extending a cardinality constraint, an affine sparsity constraint (ASC) is defined by a linear inequality with two sets of variables: one set of continuous variables and the other set represented by their nonzero patterns. This paper aims to study an ASC system consisting of finitely many affine sparsity constraints. We investigate a number of fundamental structural properties of the solution set of such a non-standard system of inequalities, including its closedness and the description of its closure, continuous approximations and their set convergence, and characterizations of its tangent cones for use in optimization. Based on the obtained structural properties of an ASC system, we investigate the convergence of B(ouligand) stationary solutions when the ASC is approximated by surrogates of the step ℓ_0-function commonly employed in sparsity representation. Several examples with geometric intuitions will be discussed in this talk. Our study lays a solid mathematical foundation for solving optimization problems involving these affine sparsity constraints through their continuous approximations.

Short bio: Hongbo Dong is an Assistant Professor in the Department of Mathematics and Statistics at the Washington State University. He completed his Ph.D. at the University of Iowa in 2011 and worked as a postdoc in the University of Wisconsin-Madison until 2013. He previously worked in areas including copositive/conic programming and mixed-integer quadratic programming. He also collaborated actively with researchers in other disciplines in statistics and agronomy. I work in the area of mathematical optimization.

My current research concerns theory and algorithms for nonconvex optimization problems, especially those with structures such as the binary indicator (on/off) variables and constraints. I am also interested in applications of optimization in areas such as data analysis, sustainable agriculture, etc.

Title: Blessing of massive scale: Spatial graphical model estimation with a total cardinality constraint

Speaker: Ethan X. Fang, Department of Statistics, Pennsylvania State University, State College, Pennsylvania 18602, U.S.A. Email: ethanfangxy@gmail.com.
Webpage: http://stat.psu.edu/people/xxf13

Abstract: We consider the problem of estimating high dimensional spatial graphical models with a total cardinality constraint (i.e., the L0-constraint). Though this problem is highly nonconvex, we show that its primal-dual gap diminishes linearly with the dimensionality and provide a convex geometry justification of this blessing of massive scale phenomenon. Motivated by this result, we propose an efficient algorithm to solve the dual problem (which is concave) and prove that the solution achieves optimal statistical properties. Extensive numerical results are also provided. This is a joint work with Han Liu and Mengdi Wang.

Short bio: Ethan is an assistant professor at Penn State University. Before joining Penn State, he got his PhD from Princeton University in 2016 and his bachelor's degree from National University of Singapore in 2010. He works on different problems such as graphical model estimation, high-dimensional inference and adaptive trial design from both statistical and computational perspectives. He won numerous awards such as Best Paper Prize for Young Researchers in Continuous Optimization (jointly with Mengdi Wang and Han Liu), ENAR Distinguished Student Paper Prize and IMS Laha/Travel Award.

Title: Deep learning in vision and language intelligence

Speaker: Xiaodong He, Deep Learning Technology Center of Microsoft Research, Redmond, Washington 98052, U.S.A. Email: xiaohe@microsoft.com.
Webpage: https://www.microsoft.com/en-us/research/people/xiaohe/

Abstract: Deep learning, which exploits multiple levels of data representations that give rise to hierarchies of concept abstraction, has been the driving force in the recent resurgence of Artificial Intelligence (AI). In this talk, I will summarize rapid advances in cognitive AI, particularly including comprehension, reasoning, and generation across vision and natural language, and applications in vision-to-text captioning, text-to-image synthesis, and reasoning grounded on images for question answering and dialog. I will also discuss future AI breakthrough that will benefit from multimodal intelligence, which empowers the communication between humans and the real world and enables enormous scenarios such as universal chatbot and intelligent augmented reality.

Short bio: Xiaodong He is a Principal Researcher in the Deep Learning Technology Center of Microsoft Research. He is also an Affiliate Professor in the Department of Electrical Engineering at the University of Washington (Seattle). His research interests are mainly in artificial intelligence including deep learning, natural language processing, computer vision, speech, information retrieval, and knowledge representation. He has published more than 100 papers and one book in these areas. He received several awards including the Outstanding Paper Award at ACL 2015. He has led the development of Machine Translation Systems that won the No. 1 Place in the 2008 NIST MT Eval and the 2011 IWSLT Eval, respectively. More recently, he and colleagues developed the MSR image captioning system that won the first prize, tied with Google, at the COCO Captioning Challenge 2015. He is leading the image captioning effort now that is part of the Microsoft Cognitive Services and CaptionBot.ai. The work receives widely media coverage including Business Insider, Forbes, The Washington Post, CNN, BBC. He has held editorial positions on several IEEE Journals, served as an area chair for NAACL-HLT 2015, and served in the organizing committee/program committee of major speech and language processing conferences. He is an elected member of the IEEE SLTC for the term of 2015-2017. He is a senior member of IEEE and a member of ACL. He was the Chair of the IEEE Seattle Section in 2016.

Title: A proximal primal-dual algorithm for decomposing non-convex nonsmooth problems

Speaker: Mingyi Hong, Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, Iowa 50011, U.S.A. Email: `mingyi@iastate.edu`.
Webpage: `https://www.imse.iastate.edu/directory/faculty/mingyi-hong/`

Abstract: In this talk, we discuss a new method for decomposing non-convex nonsmooth optimization problems with linearly coupling constraints. The proposed method consists of one step of approximate primal gradient-type iterations followed by an approximate dual ascent step. Due to the special way that the primal and dual steps are designed, the proposed method can effectively decompose a number of challenging non-convex problems into simple subproblems (possibly with closed-form solutions). We analyze various properties of the proposed method, including convergence and convergence rate. Further, we discuss application of the proposed method in distributed eigenvalue decomposition problem, as well as solving certain generalized sparse principal subspace estimation problem.

Short bio: Mingyi Hong received his Ph.D. degree from University of Virginia in 2011. He is a Black & Veatch Faculty Fellow and an Assistant Professor with the Department of Industrial and Manufacturing Systems Engineering, Iowa State University. Since January 2017, he has been serving on the IEEE Signal Processing Society's Signal Processing for Communications and Networking (SPCOM) Technical Committee. His research interests are primarily in the fields of optimization theory and applications in signal processing and machine learning.

Title: Matrix completion, saddle points, and gradient descent

Speaker: Jason D. Lee, Marshall School of Business, University of Southern California, Los Angeles, California 90089, U.S.A. Email: `jasondlee88@gmail.com`.
Webpage: `http://www-bcf.usc.edu/~lee715/`

Abstract: Matrix completion is a fundamental machine learning problem with wide applications in collaborative filtering and recommender systems. Typically, matrix completion are solved by non-convex optimization procedures, which are empirically extremely successful. We prove that the symmetric matrix completion problem has no spurious local minima, meaning all local minima are also global. Thus the matrix completion objective has only saddle points and global minima. Next, we show that saddlepoints are easy to avoid for even Gradient Descent – arguably the simplest optimization procedure. We prove that with probability 1, randomly initialized Gradient Descent converges to a local minimizer. The same result holds for a large class of optimization algorithms including proximal point, mirror descent, and coordinate descent.

Short bio: Jason Lee is an assistant professor in Data Sciences and Operations at the University of Southern California. Prior to that, he was a postdoctoral researcher at UC Berkeley working with Michael Jordan. Jason received his PhD at Stanford University advised by Trevor Hastie and Jonathan Taylor. His research interests are in statistics, machine learning, and optimization. Lately, he has worked on high-dimensional statistical inference, analysis of non-convex optimization algorithms, and theory for deep learning.

Title: Sparse regression for block missing data without imputation

Speaker: Yufeng Liu, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, North Carolina 27599 Email: `yfliu@email.unc.edu`.
Webpage: `http://stat-or.unc.edu/people/faculty/yufeng-liu`

Abstract: Supervised learning techniques have been widely used in diverse scientific disciplines such as business, finance, biology and neuroscience. In this talk, I will present a new technique for flexible learning of data with complex block-missing structure. We focus on data with multiple modalities (sources or types). In practice, it is common to have block-missing structure for such multi-modality data. A new technique

effectively using all available data information without imputation will be discussed. The corresponding optimization problem can be implemented efficiently. Applications for the Alzheimer's Disease Neuroimaging Initiative (ADNI) data will be used to illustrate the performance of the proposed method.

Short bio: Listed in the section of organizers.

Title: High-dimensional robust regression

Speaker: Po-Ling Loh, Department of Electrical and Computer Engineering, University of Wisconsin, Madison, Wisconsin 53706, U.S.A. `Email: loh@ece.wisc.edu`.
Webpage: `https://wid.wisc.edu/profile/po-ling-loh/`

Abstract: We present results for high-dimensional linear regression using robust M-estimators with a regularization term. We show that when the derivative of the loss function is bounded, our estimators are robust with respect to heavy-tailed noise distributions and outliers in the response variables, with the usual order of $k \log p/n$ rates for high-dimensional statistical estimation. Our results continue a line of recent work concerning local optima of nonconvex M-estimators with possibly nonconvex penalties, where we adapt the theory to settings where the loss function only satisfies a form of restricted strong convexity within a local neighborhood. We also discuss second-order results concerning the asymptotic normality of our estimators, and provide a two-step M-estimation algorithm for obtaining statistically efficient solutions within the local region.

Short bio: Po-Ling Loh is an assistant professor in the ECE department at the UW-Madison, with a secondary appointment in statistics, and an affiliate of the Grainger Institute and Wisconsin Institute for Discovery. From 2014–2016, Po-Ling was an assistant professor in the statistics department at the Wharton School at the University of Pennsylvania. Po-Ling received an MS in computer science and a PhD in statistics from Berkeley in 2013 and 2014, and a BS in math with a minor in English from Caltech in 2009. She was the recipient of the 2014 Erich L. Lehmann Citation from the Berkeley statistics department for an outstanding PhD dissertation in theoretical statistics, and a best student paper award at the NIPS conference in 2012.

Title: Minimizing the difference of L1 and L2 norms with applications

Speaker: Yifei Lou, Mathematical Sciences Department, University of Texas Dallas, Dallas, Texas 75080 U.S.A. `Email: Yifei.Lou@utdallas.edu`.
Webpage: `https://sites.google.com/site/louyifei/Home`

Abstract: A fundamental problem in compressive sensing (CS) is to reconstruct a sparse signal under a few linear measurements far less than the physical dimension of the signal. Currently, CS favors incoherent systems, in which any two measurements are as little correlated as possible. In reality, however, many problems are coherent, in which case conventional methods, such as L1 minimization, do not work well. In this talk, I will present a novel non-convex approach, which is to minimize the difference of L1 and L2 norms (L1-L2) in order to promote sparsity. In addition to theoretical aspects of the L1-L2 approach, I will discuss two minimization algorithms. One is the difference of convex (DC) function methodology, and the other is based on a proximal operator, which makes some L1 algorithms (e.g. ADMM) applicable for L1-L2. Experiments demonstrate that L1-L2 improves L1 consistently and it outperforms Lp ($0 < p < 1$) for highly coherent matrices.

Short bio: Yifei Lou has been an Assistant Professor in the Mathematical Sciences Department, University of Texas Dallas, since 2014. She received her Ph.D. in Applied Mathematics from the University of California Los Angeles (UCLA) in 2010. After graduation, she was a postdoc in the School of Electrical and Computer Engineering, Georgia Institute of Technology, working on medical imaging applications. In 2012-2014, she was a postdoc at the Department of Mathematics, University of California Irvine. Her research interests include compressive sensing and its applications, image analysis (medical imaging, hyperspectral, imaging through turbulence), and (nonconvex) optimization algorithms.

Title: Confidence regions and intervals for sparse penalized regression using variational inequality techniques

Speaker: **Shu Lu**, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, North Carolina 27599-3260, U.S.A. Email: shulu@email.unc.edu
Webpage: http://www.unc.edu/~shulu/

Abstract: With the abundance of large data, sparse penalized regression techniques are commonly used in data analysis due to the advantage of simultaneous variable selection and prediction. In this talk, we discuss a framework to construct confidence intervals for sparse penalized regression with a wide range of penalties including the LASSO and the nonconvex penalties such as SCAD and MCP. We study the inference for two types of parameters: the parameters under the population version of the penalized regression and the parameters in the underlying linear model. We present convergence properties of the proposed methods as well as results for simulated and real data examples. This is based on joint work with **Yufeng Liu**, Liang Yin, Kai Zhang and Guan Yu.

Short bio: Shu Lu received her B.S. and M.S. in Civil Engineering from Tsinghua University, and her M.A. in Mathematics and Ph.D. in Industrial and Systems Engineering from the University of Wisconsin-Madison. She is currently Associate Professor at the Department of Statistics and Operations Research, University of North Carolina at Chapel Hill. Her research interests include variational inequalities and variational analysis, optimization under uncertainty, and their applications.

Title: A convex optimization approach to the automatic calibration of distributed sensors

Speaker: **Zhi-Quan Luo**, The Chinese University of Hong Kong, Shenzhen, China, and Department of Electrical and Computer Engineering, University of Minnesota (Twin Cities). (cancelled)
Email: luozq@cuhk.edu.cn. Webpage: http://people.ece.umn.edu/~luozq/

Abstract: In this work, we consider the problem of automatically calibrating the system biases in a group of distributed sensors using convex optimization. In practice, estimating sensors range and azimuth biases from their noisy asynchronous measurements is an important step in system calibration, and is generally very challenging due to nonlinear transformation between the sensors global and local coordinate systems as well as the timing inconsistencies between the measurements from different sensors. In this work, we formulate the problem as a novel nonlinear (nonconvex) least square problem by only assuming the existence of an object moving at an unknown constant velocity. We propose a block coordinate decent optimization algorithm, with a judicious initialization, for solving the problem. The proposed algorithm updates the range and azimuth biases by alternately solving linear least square problems and semidefinite programs. We prove that the proposed algorithm can find the globally optimal solution and the true system biases in the absence of observation noise. Simulations show the effectiveness and the efficiency of the proposed approach.

Short bio: Professor Luo received his BSc degree in Applied Mathematics from Peking University, China in 1984, and a Ph.D in Operations Research from the Massachusetts Institute of Technology in 1989. From 1989 to 2003, Professor Luo held a faculty position with the Department of Electrical and Computer Engineering, McMaster University, Canada, where he also served as the department head from 2000 to 2003, and held a tier-1 Canada Research Chair in Information Processing from 2001 to 2003. From 2003, he has been a full professor at the Department of Electrical and Computer Engineering, University of Minnesota (Twin Cities). Since 2014, Professor Luo has served as the Vice President (Academic) at the Chinese University of Hong Kong, Shenzhen, China.

Professor Luo's research mainly addresses mathematical issues in information sciences, with particular focus on the design, analysis and applications of optimization algorithms. He consults regularly with industry on topics related to signal processing and digital communication. Professor Luo is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and a Fellow of the Society for Industrial and Applied Mathematics (SIAM). He received the 2010 Farkas Prize from the INFORMS Optimization Society for outstanding contributions to the field of optimization. He also received four Best Paper Awards from the IEEE Signal Processing Society in 2004, 2009, 2011 and 2015 respectively, and a 2011 Best Paper Award from the EURASIP. In 2014, he was elected to the Royal Society of Canada.

Title: SOFAR: Large-scale association network learning

Speakers: Jinchi Lv and Yingying Fan, Marshall School of Business, University of Southern California, Los Angeles, California 90089, U.S.A.
Emails: jinchilv@marshall.usc.edu; fanyingy@marshall.usc.edu.
Webpages: https://www.marshall.usc.edu/personnel/1096;
https://www.marshall.usc.edu/personnel/536

Abstract: Many modern big data applications feature large scale in both numbers of responses and predictors. Better statistical efficiency and scientific insights can be enabled by understanding the large-scale response-predictor association network structures via layers of sparse latent factors ranked by importance. Yet sparsity and orthogonality have been two largely incompatible goals. To accommodate both features, in this paper we suggest the method of sparse orthogonal factor regression (SOFAR) via the sparse singular value decomposition with orthogonality constrained optimization to learn the underlying association networks, with broad applications to both unsupervised and supervised learning tasks such as bi-clustering with sparse singular value decomposition, sparse principal component analysis, sparse factor analysis, and spare vector auto regression analysis. Exploiting the framework of convexity-assisted nonconvex optimization, we derive nonasymptotic error bounds for the suggested procedure characterizing the theoretical advantages. The statistical guarantees are powered by an efficient SOFAR algorithm with convergence property. Both computational and theoretical advantages of our procedure are demonstrated with several simulation and real data examples.

Short bio: Yingying Fan is an Associate Professor in Data Sciences and Operations Department of the Marshall School of Business at the University of Southern California and an Associate Fellow of USC Dornsife Institute for New Economic Thinking (INET). She received her Ph.D. in Operations Research and Financial Engineering from Princeton University in 2007 under the supervision of Professor Jianqing Fan. Her research interests include deep learning, causal inference, personalized medicine and choices, scalable Bayesian inference, large-scale inference and false discovery rate control, networks, high-dimensional statistics, high-dimensional classification, big data problems, statistical machine learning, nonparametric statistics, business applications, and financial econometrics. Her papers have been published in journals in statistics, economics, and computer science. She serves as an associate editor of Journal of the American Statistical Association (2014-present), Journal of Econometrics (2015-present), The Econometrics Journal (2012-present), and Journal of Multivariate Analysis (2013-2016). She is the recipient of the Royal Statistical Society Guy Medal in Bronze (2017), the USC Marshall Inaugural Dr. Douglas Basil Award for Junior Business Faculty (2014), the American Statistical Association Noether Young Scholar Award (2013), NSF Faculty Early Career Development (CAREER) Award (2012), Zumberge Individual Award from USC's James H. Zumberge Faculty Research and Innovation Fund (2010), and USC Marshall Dean's Award for Research Excellence (2010), as well as a Plenary Speaker at the 2011 Institute of Mathematical Statistics Workshop on Finance, Probability, and Statistics held at Columbia University.

Jinchi Lv is McAlister Associate Professor in Business Administration in Data Sciences and Operations Department of the Marshall School of Business at the University of Southern California, Associate Professor in Department of Mathematics at USC, and an Associate Fellow of USC Dornsife Institute for New Economic Thinking (INET).

Title: Sparse multivariate statistics with discrete optimization

Speaker: Rahul Mazumder, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, U.S.A. Email: rahulmaz@mit.edu.
Webpage: http://www.mit.edu/~rahulmaz/

Abstract: Several statistical estimation tasks arising in modern multivariate statistics are naturally posed as discrete optimization problems. While continuous convex optimization methods have played a highly influential role in these tasks, the role of modern discrete optimization methods, namely, integer programming has been relatively less explored, despite the tremendous advances in the field over the past 10-15 years. In this talk I will describe how techniques in modern computational optimization: mixed integer optimization and first order methods in nonlinear optimization, provide a systematic algorithmic lens to address some key problems in sparse multivariate statistics. I will illustrate how this approach leads to estimators with very high quality statistical properties across the domains of sparse regression, robust statistical regression, nonparametric function estimation and factor analysis.

Short bio: Rahul Mazumder is an Assistant Professor in the Operations Research and Statistics Group at the MIT Sloan School of Management. He is also affiliated with the Operations Research Center and MITs Center of Statistics and Data Science. Prior to joining MIT, he was an Assistant Professor in the Department of Statistics, Columbia University from Fall 2013 through June 2015, and was also affiliated with the Data Science Institute, Columbia University. Rahul Mazumder completed his B.Stat. and M.Stat. from the Indian Statistical Institute, Kolkata in 2007. He received his PhD in statistics from Stanford University in 2012. Rahul Mazumder was a PostDoctoral Associate at MIT from 2012–2013. His research interests are in data science, statistical machine learning, large scale optimization, mathematical programming; and in particular, their interplay.

Title: The landscape of some statistical learning problems

Speaker: **Andrea Montanari**, Department of Electrical Engineering and Department of Statistics, Stanford University, Stanford, California 94305 U.S.A. `Email: montanari@stanford.edu`. Webpage: `http://www.stanford.edu/~montanar/index.html`

Abstract: Most high-dimensional estimation and prediction methods propose to minimize a cost function (empirical risk) that is written as a sum of losses associated to each data point (each example). Studying the landscape of the empirical risk is useful to understand the computational complexity of these statistical problems. I will discuss some generic features that can be used to prove that the global minimizer can be computed efficiently even if the loss in non-convex. A different mechanism arises in some rank-constrained semidefinite programming problems. In this case, optimization algorithms can only be guaranteed to produce an (approximate) local optimum, but all local optima are close in value to the global optimum. Finally I will contrast these with problems in which the effects of non-convexity are more dramatic. [Based on joint work with Yu Bai, Song Mei, Theodor Misiakiewicz and Roberto Oliveira]

Short bio: Andrea Montanari received a Laurea degree in Physics in 1997, and a Ph.D. in Theoretical Physics in 2001 (both from Scuola Normale Superiore in Pisa, Italy). He has been post-doctoral fellow at Laboratoire de Physique Thorique de l'Ecole Normale Suprieure (LPTENS), Paris, France, and the Mathematical Sciences Research Institute, Berkeley, USA. Since 2002 he is Chargé de Recherche (with Centre National de la Recherche Scientifique, CNRS) at LPTENS. In September 2006 he joined Stanford University as a faculty, and since 2015 he is Full Professor in the Departments of Electrical Engineering and Statistics. He was co-awarded the ACM SIGMETRICS best paper award in 2008. He received the CNRS bronze medal for theoretical physics in 2006, the National Science Foundation CAREER award in 2008, the Okawa Foundation Research Grant in 2013, and the Applied Probability Society Best Publication Award in 2015. He is an Information Theory Society distinguished lecturer for 2015-2016. In 2016 he received the James L. Massey Research & Teaching Award of the Information Theory Society for young scholars.

Title: Inference from low order marginals

Speaker: **Meisam Razaviyayn**, The Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California 90089, U.S.A. `Email: razaviya@usc.edu`. Webpage: `https://ise.usc.edu/directory/faculty/profile/?lname=Razaviyayn&fname=Meisam`

Abstract: In many modern inference problems, the task is to predict some target variable Y from some discrete feature vector $X = (X1, X2, \cdots, Xp)$. When the joint distribution of (X, Y) is known, this task can be done "optimally" by employing the Maximum A-posteriori Probability (MAP) decision rule. However, when only some low order marginals of the joint distribution of $(X, Y$ is known, this task does not have a simple solution. A fundamental question in this setting is as follows: among all probability distributions satisfying the estimated low order marginals, which one should be used for prediction? We formulate this problem as a robust optimization problem; and suggest to use the Hirschfeld-Gebelein-Renyi (HGR) correlation principle for finding an approximate solution. The approximate solution can be shown to lead to a classier with mis-classication rate no larger than twice the mis-classication rate of the optimal classier. Under a certain "separability" condition, an efficient algorithm is proposed for finding the proposed solution.

Short bio: Meisam Razaviyayn is an assistant professor at the department of Industrial and Systems Engineering at the University of Southern California. Prior to joining USC, he was a postdoctoral research fellow in the Electrical Engineering Department at Stanford University. He obtained his Ph.D. degree in Electrical Engineering with a minor in Computer Science from the University of Minnesota in 2014. He is the recipient of the Signal Processing Society Young Author Best Paper Award in 2015 and the University of Minnesota Doctoral Dissertation Fellowship in 2014. He was among the three finalists of the Best Paper Prize for Young Researcher in Continuous Optimization in ICCOPT 2013 and 2016, and the finalist for the best student paper award in SPAWC 2010. His research interests include the design and study of data analysis algorithms and tools which can efficiently scale to modern big data problems.

Title: In-network nonconvex large-scale optimization

Speaker: Gesualdo Scutari, School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47907-5400, U.S.A. `Email: gscutari@purdue.edu`.
Webpage: `https://engineering.purdue.edu/~gscutari/`

Abstract: Nowadays, large-scale systems are ubiquitous. Some examples/applications include wireless communication networks; electricity grid, sensor, and cloud networks; and machine learning and signal processing applications, just to name a few. In many of the above systems, i) data are distributively stored in the network (e.g., clouds, computers, sensors, robots), and ii) it is often impossible to run analytics on central fusion centers, owing to the volume of data, energy constraints, and/or privacy issues. Thus, distributed in-network processing with parallelized multi-processors is preferred. Moreover, many applications of interest lead to large-scale optimization problems with nonconvex, nonseparable objective functions. All this makes the analysis and design of distributed/parallel algorithms over networks a challenging task. In this talk we will present our ongoing work in this area. More specifically, we consider a large-scale network composed of agents aiming to distributively minimize a (nonconvex) smooth sum-utility function plus a nonsmooth (nonseparable), convex one. The latter is usually employed to enforce some structure in the solution, e.g., sparsity. The agents have access only to their local functions (data) but not the whole objective, and the network is modeled as a directed, time-varying, graph. We propose a distributed solution method for the above optimization wherein the agents in parallel minimize a convex surrogate of the original nonconvex objective while using a novel tacking mechanism and broadcast protocol to estimate locally the missing global information and distribute the computations over the network, respectively. We discuss several instances of the general algorithm framework tailored to specific (convex and nonconvex) applications while exploiting the trade-off between local computation, communication, and convergence rate. This is a joint work with Ying Sun (Purdue University).

Short bio: Gesualdo Scutari is an Associate Professor with the School of Industrial Engineering at Purdue University and is the Scientific Director for the area of Big-Data Analytics at the Cyber Center (Discovery Park) at Purdue University. His primary research interests focus on theoretical and algorithmic issues related to continuous (large-scale) optimization, big-data analytics, equilibrium programming, and their applications to signal processing, communications, and machine learning. He is an Associate Editor of the IEEE Transactions on Signal Processing and IEEE Transactions on Signal and Information Processing over Networks.

Title: Breaking sample complexity barriers via nonconvex optimization?

Speaker: Mahdi Soltanolkotabi, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089, U.S.A. `Email: soltanol@usc.edu`.
Webpage: `http://www-bcf.usc.edu/~soltanol/`

Abstract: In the past decade there has been significant progress in understanding when convex relaxations are effective for finding low complexity models from a near minimal number of data samples (e.g. sparse/low rank recovery from a few linear measurements). Despite such advances convex optimization techniques are often prohibitive in practice due to computational/memory constraints. Furthermore, in some cases convex programs are also suboptimal in terms of sample complexity and provably require significantly more data samples than what is required to uniquely specify the low complexity model of interest. In fact for many such problems certain sample complexity barriers have emerged so that there are no known computationally tractable algorithm that can beat the sample complexity achieved by such convex relaxations. Motivated by a problem in imaging, in this talk I will discuss my recent results towards breaking such barriers via natural nonconvex optimization techniques.

Short bio: Mahdi Soltanolkotabi is currently an assistant professor in the Ming Hsieh Department of Electrical Engineering at the University of Southern California. Prior to joining USC, he completed his PhD in electrical engineering at Stanford in 2014. He was a postdoctoral researcher in the EECS department at UC Berkeley during the 2014-2015 academic year. His research focuses on the design and mathematical understanding of computationally efficient algorithms for optimization, high dimensional statistics, machine learning, signal processing and computational imaging. Recently, a main focus of his research has been on developing and analyzing algorithms for non-convex optimization, with provable guarantees of convergence to the global optimum.

Title: Solving big optimization problems using the second-order sparsity

Speaker: Defeng Sun, Department of Mathematics, National University of Singapore, Republic of Singapore. `Email: matsundf@nus.edu.sg`.
Webpage: `http://www.math.nus.edu.sg/~matsundf/`

Abstract: Big optimization models are ubiquitous in machine learning, statistics, finance, signal processing, imaging science, geophysics and many other areas. Concerned with the huge computational burden of the interior-point methods (IPMs) for solving big-scale problems, convex or nonconvex, many researchers and practitioners tend to believe that the first-order methods such as the accelerated proximal gradient methods and the alternating direction methods of multipliers are the only options for the rescue. While these first-order methods have enjoyed successful stories in some interesting applications, they also encounter enormous numerical difficulties in dealing with many real data problems of big scales even only with a low or moderate solution quality. New ideas for solving these problems are highly sought both in practice and academic research. In this talk, we shall demonstrate how the second-order sparsity property exhibited in big sparse optimization models can be intelligently explored to overcome the mentioned difficulties either in IPMs or in the first order methods. One critical discovery is that the second-order sparsity allows one to solve sub-problems at costs even lower than several first order methods. For the purpose of illustration, we shall present highly efficient and robust semismooth Newton based augmented Lagrangian methods for solving various lasso and support vector machine models.

This talk is based on my joint works with **Kim-Chuan Toh** and our graduate students.

Short bio: Defeng Sun is Professor at Department of Mathematics, National University of Singapore. His main research interest lies in large scale matrix optimization and statistical learning. Currently he serves as associate editor to Mathematical Programming, both Series A and Series B, SIAM Journal on Optimization and others.

Kim-Chuan Toh is a Professor at the Department of Mathematics, National University of Singapore (NUS). He obtained his BSc degree in Mathematics from NUS in 1990 and the PhD degree in Applied Mathematics from Cornell University in 1996 under the direction of Nick Trefethen. He is currently an Area Editor for Mathematical Programming Computation, and an Associate Editor for the SIAM Journal on Optimization. He also serves as the secretary of the SIAM Activity Group on Optimization. He has been invited to speak at numerous conferences and workshops, including SIAM Annual Meeting in 2010 and ISMP in 2006. His current research focuses on designing efficient algorithms and software for convex programming, particularly large scale optimization problems arising from data science and large scale matrix optimization problems such as linear semidefinite programming (SDP) and convex quadratic semidefinite programming (QSDP).

Title: Statistical optimization: Solving nonconvex optimization from a statistician's perspective

Speaker: **Qiang Sun**, Department of Operations Research and Financial Engineering Princeton University, Princeton, New Jersey 08544, U.S.A. Email: `qsun.ustc@gmail.com`.

Abstract: Statistical optimization has received quite some interests recently. It refers to the case where hidden and local convexity can be discovered in most cases for nonconvex problems, making polynomial algorithms possible. It relies on careful analysis of the geometry near global optima. In this talk, I will explore this direction by focusing on sparse regression problems in high dimensions. A computational framework named iterative local adaptive majorize-minimization (I-LAMM) is proposed to simultaneously control algorithmic complexity and statistical error. I-LAMM effectively turns the nonconvex penalized regression problem into a series of convex programs by utilizing the locally strong convexity of the problem when restricting the solution set in an ℓ_1 cone. Computationally, we establish a phase transition phenomenon: it enjoys linear rate of convergence after a sub-linear burn-in. Statistically, it provides solutions with optimal statistical errors. Extensions to reduced-rank regression, matrix factorization and discrete models will be discussed.

Short bio: Qiang is currently a postdoc at Princeton University. He received his PhD in Biostats from UNC in 2014. He will join the University of Toronto as an assistant professor this fall. Qiang's research is mainly motivated by applications in imaging genetics and his research interest includes statistical optimization, non-asymptotic robustness and inference in high dimensions.

Title: Proximal DC algorithm for sparse optimization

Speaker: **Akido Takeda**, Department of Mathematical Analysis and Statistical Inference, The Institute of Statistical Mathematics, Tokyo 190-8562, Japan. Email: `atakeda@ism.ac.jp`.
Webpage: `http://www.ism.ac.jp/~atakeda/index-e.html`

Abstract: Many applications such as in signal processing, machine learning and operations research seek sparse solutions by adopting the cardinality constraint or rank constraint. We formulate such problems as DC (Difference of two Convex functions) optimization problems and apply DC Algorithm (DCA) to them. While the DCA has been widely used for this type of problems, it often requires a large computation time to solve a sequence of convex subproblems. Our algorithm, which we call Proximal DC Algorithm (PDCA), overcomes this issue of the ordinary DCA by employing a special DC decomposition of the objective function. In PDCA, closed-form solutions can be obtained for the convex subproblems, leading to efficient performance in numerical experiments. We also discuss the theoretical aspects: PDCA can be viewed as a nonconvex variant of the proximal gradient methods (PGM), which provides an insight on the relation between PGM and DCAs.

Short bio: Akiko Takeda is currently a professor at the Institute of Statistical Mathematics, Japan. She holds B.E. and M.E. degrees in Administration Engineering from Keio University and Dr.Sc. degree in Information Science from Tokyo Institute of Technology. Her research interests include solution methods for decision making problems under uncertainty and non-convex optimization problems, which appear in machine learning and energy systems.

Title: Randomized linear programming solves Markov decision problems much faster

Speaker: **Mengdi Wang**, Department of Operations Research and Financial Engineering Princeton University, Princeton, New Jersey 08544, U.S.A. Email: `mengdiw@Princeton.edu`.
Webpage: `http://www.princeton.edu/~mengdiw/`

Abstract: We propose a randomized linear programming algorithm for approximating the optimal policy of the discounted Markov decision problem. By leveraging the value-policy duality, the algorithm adaptively samples state transitions and makes exponentiated primal-dual updates. We show that it finds an ε-optimal policy using nearly-linear running time in the worst case. For Markov decision processes that are ergodic under every stationary policy, we show that the algorithm finds an ε-optimal policy using running time linear in the total number of state-action pairs, which is sublinear in the input size. These results provide new complexity benchmarks for solving stochastic dynamic programs. We will also show how this method can be applied in online reinforcement learning and obtain near-optimal regret.

Short bio: Mengdi Wang is interested in data-driven stochastic optimization and applications in machine and reinforcement learning. She received her PhD in Electrical Engineering and Computer Science from Massachusetts Institute of Technology in 2013. At MIT, Mengdi was affiliated with the Laboratory for Information and Decision Systems and was advised by Dimitri P. Bertsekas. Mengdi became an assistant professor at Princeton in 2014. She received the Young Researcher Prize in Continuous Optimization of the Mathematical Optimization Society in 2016 (awarded once every three years), the Princeton SEAS Innovation Award in 2016, and the NSF Career Award in 2017.

Title: Algorithmic tools for smooth nonconvex optimization

Speaker: **Stephen J. Wright**, Department of Computer Science, University of Wisconsin, Madison, Wisconsin 53706, U.S.A. Email: `swright@cs.wisc.edu`.
Webpage: `http://pages.cs.wisc.edu/~swright/`

Abstract: Unconstrained optimization of a smooth nonconvex objective over many variables is a classic problem in optimization. Several effective techniques have been proposed over the years, along with results about global and local convergence. There has been an upsurge of interest recently on techniques with good global complexity properties. (This interest is being driven largely by researchers in machine learning, who want to solve the nonconvex problems arising from neural network training and robust statistics, but it has roots in the optimization literature.) In this talk we describe the algorithmic tools that can be used to design methods with appealing practical behavior as well as provably good global convergence properties. These tools include the conjugate gradient and Lanczos algorithms, accelerated gradient, Newton's method, cubic regularization, and trust regions. We show how these elements can be assembled into a comprehensive method, and compare a number of proposals that have been made to date. We pay particular attention to the behavior of accelerated gradient methods in the neighborhood of saddle points.

Short bio: Stephen J. Wright holds the George B. Dantzig Professorship, the Sheldon Lubar Chair, and the Amar and Balinder Sohi Professorship of Computer Sciences at the University of Wisconsin-Madison. His research is in computational optimization and its applications to many areas of science and engineering. Prior to joining UW-Madison in 2001, Wright held positions at North Carolina State University (1986-90), Argonne National Laboratory (1990-2001), and the University of Chicago (2000-2001). He has served as Chair of the Mathematical Optimization Society and as a Trustee of SIAM. He is a Fellow of SIAM. In 2014, he won the W.R.G. Baker award from IEEE.

Wright is the author/coauthor of widely used text and reference books in optimization including *Primal Dual Interior-Point Methods* and *Numerical Optimization*. He has published widely on optimization theory, algorithms, software, and applications. Wright is current Editor-in-Chief of the *SIAM Journal on Optimization* and previously served as editor-in-chief or associate editor of Mathematical Programming (Series A), Mathematical Programming (Series B), SIAM Review, SIAM Journal on Scientific Computing, and several other journals and book series.

Title: Folded concave statistical learning of high-dimensional graphical models

Speaker: Lingzhou Xue, Department of Statistics, Pennsylvania State University, State College, Pennsylvania 18602, U.S.A. Email: `lingzhou@psu.edu`.
Webpage: `http://www.personal.psu.edu/lxx6/`

Abstract: Folded concave statistical learning has been shown to enjoy the (strong) oracle property, and it has received considerable attention in high-dimensional sparse estimation such as sparse (generalized) linear regression, sparse quantile regression, and sparse graphical models. In this talk, I will talk about our recent efforts on exploring nonconvex statistical learning of sparse graphical models. In the first part of this talk, I will discuss the folded concave penalized D-trace estimation procedure for learning sparse Gaussian (copula) graphical models, and explore its computational and statistical aspects. In the second part, I will present the folded concave penalized composite conditional likelihood approach for learning sparse Ising models with higher order interactions, where folded concave functions are used to approximate affine sparsity constraints. Simulation studies and real applications will be presented to demonstrate the power of our proposed methods.

Short bio: Lingzhou Xue is currently an Assistant Professor of Statistics at The Pennsylvania State University. He received his B.S. degree in Statistics from Peking University in 2008 and Ph.D. degree in Statistics from University of Minnesota in 2012. He was a postdoctoral research associate at Princeton University before joining Penn State. His research interests include high-dimensional statistical learning, large-scale inference, graphical/network models, and convex/nonconvex optimization.

Title: The convergence of ADMM and other first-order methods on nonconvex problems

Speaker: Wotao Yin, Department of Mathematics, University of California Los Angeles, Los Angeles, California 90095-1555, U.S.A. Email: Wotao Yin <wotaoyin@math.ucla.edu.
Webpage: `http://www.math.ucla.edu/~wotaoyin/`

Abstract: First-order methods have been surprising us with a lot of success at solving non-convex optimization problems in the literature. They include the minimization of quasinorms, matrix Schatten quasinorm, SCAD, bi-linear, and bi-convex functions, as well as those subject to orthogonality and sphere constraints. This talk will provide some insights toward when and why first-order methods such as ADMM and its multi-block extension converge to stationary points of non-convex problems. We provide some simple examples in which ADMM converges to the global solutions. We also present applications of non-convex ADMM with provable convergence to stationary points. This is joint work with Yu Wang (University of California at Berkeley) and Jinshan Zeng (Jiangxi Normal University).

Short bio: Wotao Yin is a professor in the Department of Mathematics of UCLA. His research interests lie in computational optimization and its applications in image processing, machine learning, and other inverse problems. He received his B.S. in mathematics from Nanjing University in 2001, and then M.S. and Ph.D. in operations research from Columbia University in 2003 and 2006, respectively. During 2006 - 2013, he was with Rice University. He won NSF CAREER award in 2008, Alfred P. Sloan Research Fellowship in 2009, and Morningside Medal in 2016.

Short bios of organizers, with the assistance of Miju Ahn

Miju Ahn (`Email: mijuahn@usc.edu`.) is a doctoral student in the Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, under the supervision of Professor Jong-Shi Pang. She received a B.A. in Applied Mathematics at the University of California, Berkeley. Her research interests include computational optimization and nonconvex statistical learning, especially in the area of sparse vector representation.

Yufeng Liu (`Email: yfliu@email.unc.edu`.
Webpage: `http://stat-or.unc.edu/people/faculty/yufeng-liu`) is a Professor in Department of Statistics and Operations Research, Department of Biostatistics, and Department of Genetics at University of North Carolina at Chapel Hill. His current research interests include statistical machine learning, high dimensional data analysis, and bioinformatics. He received the CAREER Award from National Science Foundation in 2008, Ruth and Phillip Hettleman Prize for Artistic and Scholarly Achievement in 2010, and the inaugural Leo Breiman Junior Award in 2017. He is an associate editor for Journal of Royal Statistical Society Series B, Journal of Multivariate Analysis, and previously served as an associate editor for Journal of the American Statistical Association, and Statistics Sinica. He was the chair for the Section on Statistical Learning and Data Science at The American Statistical Association in 2015, and is currently a fellow at American Statistical Association and an elected member of International Statistical Institute.

Jong-Shi Pang (`Email: jongship@usc.edu`; Webpage: `https://ise.usc.edu/directory/faculty/profile/?lname=Pang&fname=Jong-shi`) joined the University of Southern California as the Epstein Family Professor of Industrial and Systems Engineering in August 2013. Prior to this position, he was the Caterpillar Professor and Head of the Department of Industrial and Enterprise Systems Engineering at the University of Illinois at Urbana-Champaign for six years between 2007 and 2013. He held the position of the Margaret A. Darrin Distinguished Professor in Applied Mathematics in the Department of Mathematical Sciences and was a Professor of Decision Sciences and Engineering Systems at Rensselaer Polytechnic Institute from 2003 to 2007. He was a Professor in the Department of Mathematical Sciences at the Johns Hopkins University from 1987 to 2003, an Associate Professor and then Professor in the School of Management from 1982 to 1987 at the University of Texas at Dallas, and an Assistant and then an Associate Professor in the Graduate School of Industrial Administration at Carnegie-Mellon University from 1977 to 1982. During 1999 and 2001 (full time) and 2002 (part-time), he was a Program Director in the Division of Mathematical Sciences at the National Science Foundation.

Professor Pang was a winner of the 2003 George B. Dantzig Prize awarded jointly by the Mathematical Programming Society and the Society for Industrial and Applied Mathematics for his work on finite-dimensional variational inequalities, and a co-winner of the 1994 Frederick W. Lanchester Prize awarded by the Institute for Operations Research and Management Science. Several of his publications have received best paper awards in different engineering fields: signal processing, energy and natural resources, computational management science, and robotics and automation. He is an ISI Highly Cited Researcher in the Mathematics Category between 1980–1999; he has published 3 widely cited monographs and more than 100 scholarly journals in top peer reviewed journals. Dr. Pang is a member in the inaugural 2009 class of Fellows of the Society for Industrial and Applied Mathematics. Professor Pang's general research interest is in the mathematical modeling and analysis of a wide range of complex engineering and economics systems with focus in operations research, (single and multi-agent) optimization, equilibrium programming, constrained dynamical systems, and most recently, the interface between optimization, statistics, and data science and engineering.

Meisam Razaviyayn (`Email: razaviya@usc.edu`.
Webpage: `https://ise.usc.edu/directory/faculty/profile/?lname=Razaviyayn&fname=Meisam`) is an assistant professor at the department of Industrial and Systems Engineering at the University of Southern California. Prior to joining USC, he was a postdoctoral research fellow in the Electrical Engineering Department at Stanford University. He obtained his Ph.D. degree in Electrical Engineering with a minor in Computer Science from the University of Minnesota in 2014. He is the recipient of the Signal Processing

Society Young Author Best Paper Award in 2015 and the University of Minnesota Doctoral Dissertation Fellowship in 2014. He was among the three finalists of the Best Paper Prize for Young Researcher in Continuous Optimization in ICCOPT 2013 and 2016, and the finalist for the best student paper award in SPAWC 2010. His research interests include the design and study of data analysis algorithms and tools which can efficiently scale to modern big data problems.

Phebe Vayanos (Email: phebe.vayanos@usc.edu. Webpage: https://ise.usc.edu/directory/faculty/profile/?lname=Vayanos&fname=Phebe) is an Assistant Professor of Industrial and Systems Engineering and an Associate Director of the Center for Artificial Intelligence in Society at the University of Southern California. Her research interests include optimization under uncertainty, data-driven optimization and analytics, with applications in healthcare, energy, security, and education. Prior to joining USC, she was lecturer in the Operations Research and Statistics Group at the MIT Sloan School of Management, and a postdoctoral research associate in the Operations Research Center at MIT. She holds a PhD degree in Operations Research and an MEng degree in Electrical & Electronic Engineering, both from Imperial College London.

Jack Xin (Email: jxin@math.uci.edu. Webpage: https://www.math.uci.edu/~jxin/) has been Professor of Mathematics at UC Irvine since 2005. He received his Ph.D. in applied mathematics at Courant Institute, New York University in 1990. He was a postdoctoral fellow at Berkeley and Princeton in 1991 and 1992. He was assistant and associate professor of mathematics at the University of Arizona from 1991 to 1999. He was professor of mathematics from 1999 to 2005 at the University of Texas at Austin. His research interests include applied analysis, computational methods and their applications in multi-scale problems, sparse optimization, and data science. He authored over one hundred journal papers and two Springer books, and became an ISI highly cited researcher in mathematics in 2002. He is a fellow of the John S. Guggenheim Foundation, and an inaugural fellow of the American Mathematical Society (AMS) in 2012. He is Editor-in-Chief of Society of Industrial and Applied Mathematics (SIAM) Interdisciplinary Journal Multi-scale Modeling & Simulation (MMS).