2017 PKU Workshop on Computation and Big Data Analysis

June 20-21, 2017

Beijing, China

http://bicmr.pku.edu.cn/meeting/index?id=44

2017 PKU Workshop on Computation and Big Data Analysis

June 20-21, 2017

Beijing, China

http://bicmr.pku.edu.cn/meeting/index?id=44

Information for Participants

Conference Hotel For Invited Speakers

Hotel:	"Zhong Guan Xin Yuan" Global Village, Building 1 中关新园1号楼
Address:	No. 216 Zhongguancun North Road, Haidian District 北京市海淀区中关村北大街126号
Dates:	By default, the hotel room is reserved from June 19th (check in) to June 22th (check-out). Please let us know if you have a different arrival-departure schedule.
Arrival:	By air, please see this link By subway: line 4 to "east gate of Peking University"
Website:	www.pkugv.com
Tel:	+86-10-62752288

Conference Venue

Venue:	Lecture Hall, Jia Yi Bing Building 82 Jing Chun Yuan, BICMR 北京大学镜春园82号甲乙丙楼二层报告厅
Map:	PKU campus map

Meals

- Breakfasts will be complementary at the hotel.
- Lunches and dinners are provided by the workshop. Please let us know if you have any dietary restrictions or preferences.

Currency

Chinese currency is RMB. The current rate is about 6.83 RMB for 1 US dollar. The exchange of foreign currency can be done at the airport or the conference hotel. Please keep the receipt of the exchange so that you can change back to your own currency if you have RMB left before you leave China. Please notice that some additional processing fee will be charged if you exchange currency in China.

Parking at PKU Campus

If you plan to drive to PKU, please send us your license plate number; otherwise, your car cannot enter the PKU campus.

Contact Information

If you need any help, please feel free to contact

- Ms. Xiaoni Tan: xntan@bicmr.pku.edu.cn
- Jiang Hu: jianghu@pku.edu.cn

Sponsors

Tianyuan Fund for Mathematics, National Natural Science Foundation of China

National Natural Science Foundation of China

School of Mathematical Sciences, Peking University

Beijing International Center For Mathematical Research, Peking University

Committees

Organizing Committee

Weinan E, Peking University and Princeton University

Bin Dong, Peking University

Zaiwen Wen, Peking University

Yaxiang Yuan, Chinese Academy of Sciences

Pingwen Zhang, Peking University

Conference Schedule

Each talk is 45 minutes + 10 minutes for questions.

June 20, Tuesday

09:00-09:05 Opening Ceremony: Weinan E

09:05-10:00 Session 1
Chair: Pingwen Zhang

09:05-10:00 **Jianqing Fan**, Sparse Learning and Distributed PCA with Control of Statistical Errors and Computing Resources

10:00-10:10 Group Photo

10:10-10:20 Coffee Break

10:20-12:10 Session 2
Chair: Yinyu Ye

10:20-11:15 **Wotao Yin**, Asynchronous Parallel Algorithms for Large Scale Fixed-Point Problems and Optimization

11:15-12:10 **Guanghui Lan**, Decentralized stochastic gradient descent

12:10-14:00 Lunch

14:00-15:50 Session 3
Chair: Jianqing Fan

14:00-14:55 **Yaxiang Yuan**, Some progress on first-order and second-order optimization methods

14:55-15:50 **Shuzhong Zhang**, Block Optimization over Riemannian Manifolds with Linear Constraints

15:50-16:10 Coffee Break

16:10-18:00 Session 4
Chair: Yaxiang Yuan

16:10-17:05 **Zuowei Shen**, Image Restoration and Beyond

17:05-18:00 **Jieping Ye**, Big Data at Didi Chuxing

18:30 Dinner

June 21, Wednesday

09:00-09:55 Session 5
Chair: Shuzhong Zhang

09:00-09:55 **Yinyu Ye**, Graph Realization, Data Dimension Reduction and Sensor Network Localization by Convex Optimization

09:55-10:10 Coffee Break

10:10-12:00 Session 6
Chair: Zuowei Shen

10:10-11:05 **Jun Liu**, Sparse Slice Inverse Regression via LASSO

11:05-12:00 **Xiaohua Zhou**, Tree-based Ensemble Methods For Individualized Treatment Rules

12:10 Lunch

Abstracts

Sparse Learning and Distributed PCA with Control of Statistical Errors and Computing Resources

Jianqing Fan
Princeton University

High-dimensional sparse learning and analysis of Big Data data pose significant challenges on computation and communication. Scalable statistical procedures need to take into account both statistical errors and computing resource constraints. This talk illustrate this idea by using two important examples in statistical machine learning. The first one is to solve sparse learning via a computational framework named iterative local adaptive majorize-minimization (I-LAMM) to simultaneously control algorithmic complexity and statistical error when fitting high dimensional sparse models via a family of folded concave penalized quasi-likelihood. The algorithmic complexity and statistical errors are explicitly given and we show that the algorithm achieves the optimal statistical error rate under the weakest signal strength assumptions. The second problem is to study distributed PCA with communication constraints: each node machine computes the top eigenvectors and communicates to the central server; the central server then aggregates the information transmitted from the node machines and conducts another PCA based on the aggregated information. We investigate the bias and variance for such a distributed PCA. We derive the rate of convergence for distributed PCA, which depends explicitly on effective rank, eigen-gap, and the number of machines, and show that the distributed PCA performs as well as the whole sample PCA, even without full access of whole data.

Decentralized stochastic gradient descent

Guanghui Lan

Georgia Institute of Technology

Stochastic gradient descent (SGD) methods have recently found wide applications in large-scale data analysis, especially in machine learning. These methods are very attractive to process online streaming data as they only scan through the dataset only once but still generate solutions with acceptable accuracy. However, it is known that classical SGDs are ineffective in processing streaming data distributed over multi-agent network systems (e.g., sensor and social networks), mainly due to the high communication costs incurred by these methods.In this talk, we present a new class of SGDs, referred to as stochastic decentralized communication sliding methods, which can significantly reduce the aforementioned communication costs for decentralized stochastic optimization and machine learning. We show that these methods can skip inter-node communications while performing SGD iterations. As a result, these methods require a substantially smaller number of communication rounds than existing decentralized SGDs, while the total number of required stochastic (sub)gradient computations are comparable to those optimal bounds achieved by classical centralized SGD type methods.

Sparse Slice Inverse Regression via LASSO

Jun Liu

Harvard University

For multiple index models, it has recently been shown that the sliced inverse regression (SIR) is consistent for estimating the sufficient dimension reduction (SDR) subspace if and only if the dimension p and sample size n satisfies that $\rho = \lim \frac{p}{n} = 0$. Thus, when p is of the same or a higher order of n, additional assumptions such as sparsity have to be imposed in order to ensure consistency for SIR. By constructing artificial response variables made up from top eigenvectors of the estimated conditional covariance matrix, $\widehat{var}(\mathbb{E}[\boldsymbol{x}|y])$, we introduce a simple Lasso regression method to obtain an estimate of the SDR subspace. The resulting algorithm, Lasso-SIR, is shown to be consistent and achieve the optimal convergence rate under certain sparsity conditions when p is of order $o(n^2\lambda^2)$ where λ is the generalized signal noise ratio. We also demonstrate the superior performance of Lasso-SIR compared with existing approaches via extensive numerical studies and several real data examples.

Based on the joint work with Zhigen Zhao and Qian Lin.

Image Restoration and Beyond

Zuowei Shen

National University of Singapore

We are living in the era of big data. The discovery, interpretation and usage of the information, knowledge and resources hidden in all sorts of data to benefit human beings and to improve everyones day to day life is a challenge to all of us. The huge amount of data we collect nowadays is so complicated, and yet what we expect from it is so much. This provides many challenges and opportunities to many fields. As images are one of the most useful and commonly used types of data, in this talk, we start from reviewing the development of the wavelet frame (or more general redundant system) based approach for image restoration. We will observe that a good system for any data, including images, should be capable of effectively capturing both global patterns and local features. One of the examples of such system is the wavelet frame. We will then show how models and algorithms of wavelet frame based image restoration are developed via the generic knowledge of images. Then, the specific information of a given image can be used to further improve the models and algorithms. Through this process, we shall reveal some insights and understandings of the wavelet frame based approach for image restoration and its connections to other approaches, e.g. the partial differential equation based methods. Finally, we will also show, by many examples, that ideas given here can go beyond image restoration and can be used to many other applications in data science.

Graph Realization, Data Dimension Reduction and Sensor Network Localization by Convex Optimization

Yinyu Ye

Stanford University

We present recent progresses using convex optimization based model and method for the position estimation problem in Euclidean distance geometry such as graph realization, data dimension reduction, and sensor network localization. The optimization problem is set up so as to minimize the error in node positions to fit incomplete and noisy distance measures. We develop a convex optimization relaxation model and use the duality theory to derive necessary and/or sufficient conditions for whether a graph/network is "localizable" or not, when the distance measures are accurate. Observable gauges are developed to certify the quality of the position estimation of every sensor and to detect possible erroneous sensors. Furthermore, we develop regularization and gradient-based local search methods to round and improve the SDP solution when the distance measures are noisy Computations will be demonstrated to show the effectiveness of the method.

Big Data at Didi Chuxing

Jieping Ye
University of Michigan and Didi

Didi Chuxing is the largest ride-sharing company providing transportation services for over 400 million users in China. Every day, Didi's platform generates over 70TB worth of data, processes more than 9 billion routing requests, and produces over 13 billion location points. In this talk, I will show how big data and AI technologies have been applied to analyze such huge amount of transportation data to improve the travel experience for millions of people in China.

Some progress on first-order and second-order optimization methods

Yaxiang Yuan
Chinese Academy of Sciences

Optimization algorithms have been used ubiquitously in computational mathematics and data analysis. This talk will first present a few classic yet useful techniques. Then we will introduce a few progress on first-order and second-order optimization algorithms for nonlinear eigenvalue optimization, semi-definite programming and optimization with manifold constraints.

Asynchronous Parallel Algorithms for Large Scale Fixed-Point Problems and Optimization

Wotao Yin
University of California, Los Angeles

Since 2005, the single-threaded CPU speed has stopped improving significantly; it is the numbers of cores in each machine that continue to arise. On the other hand, most of our algorithms are still single-threaded, and because so, their running time will stay about the same in the future. To develop faster algorithms, especially for those large-scale problems, it is inevitable to consider parallel computing.

In parallel computing, multiple agents (e.g., CPU cores) collaboratively solve a problem by concurrently solving their simpler subproblems. For most, the subproblems depend on each other, so the agents must regularly exchange information. In asynchronous computing, each agent can compute with the information it has, even if the latest information from other agents has not arrived. Asynchronism is extremely important to the efficiency and resilience of parallel computing. Without asynchronism, all cores have to wait for the arrival of latest information, so the speed of parallel computing is dictated by the slowest core, the most difficult subproblem, and the longest communication delay. Without asynchronism, the entire parallel computing must stop when an agent (or a network link) fails and awaits a fix, and such failures will happen more often as the system gets larger. Today, most algorithms are still singled-threaded, and most of the already-parallelized algorithms are synchronous. In spite of both mathematical and coding challenges, we report recently established convergence and numerical results for a set of fixed point problems and optimization problems that arise in machine learning, image processing, portfolio optimization, second-order cone programming, and beyond

Block Optimization over Riemannian Manifolds with Linear Constraints

Shuzhong Zhang
University of Minnesota

In this talk we shall present some new results on non-convex block-optimization models over Riemannian manifolds, with binding linear constraints. We introduce some ADMM (Alternating Direction Method of Multipliers) style algorithms for a block optimization model where the objective is non-convex and each block variables are elements of some given manifolds. Moreover, there are also linear constraints linking all the variables. Such models arise naturally in tensor optimization with constraints, including approximative Tucker decomposition with constraints. Iteration complexity bounds for the iterates converging to a stationary solution are presented, together with preliminary numerical results.

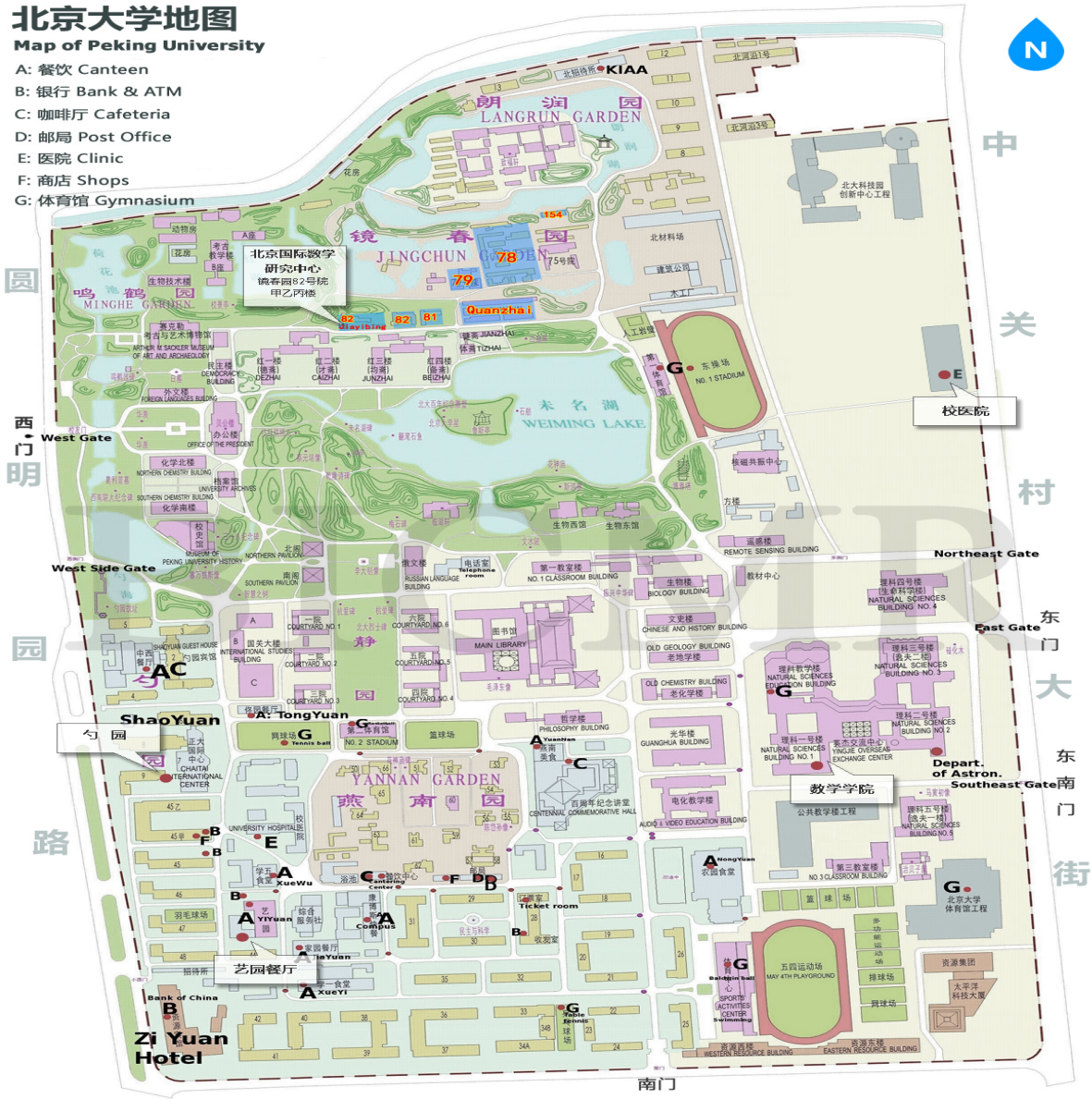
Tree-based Ensemble Methods For Individualized Treatment Rules

Xiaohua Zhou

Peking University and University of Washington

There is a growing interest in statistical methods for the personalized medicine or precision medicine, especially for deriving optimal individualized treatment rules (ITRs). An ITR recommends a patient to a treatment based on the patient's characteristics. The common parametric methods for deriving optimal ITR, which model the clinical endpoint as a function of the patient's characteristics in the first step, can have suboptimal performance when the conditional mean model is misspecified. Recent methodology development has cast the problem of deriving optimal ITR under a weighted classification framework. Under this weighted classification framework, we develop a weighted random forests (W-RF) algorithm that derives an optimal ITR nonparametrically. In addition, with the W-RF algorithm, we propose the variable importance measures for quantifying relative relevance of the patient's characteristics to treatment selection, and the out-of-bag estimator for the population average outcome under the estimated optimal ITR. Our proposed methods are evaluated through intensive simulation studies. We apply our methods to data from Clinical Antipsychotic Trials of Intervention Effectiveness Alzheimer ' s Disease Study as an illustration. This is a joint work with Kehao Zhu and Ying Huang.

Map of Peking University



The organizing committee wishes you a pleasant stay in BICMR!