2021 PKU Workshop on Optimization Theory and Methods

Dedicated to Prof. Yin Zhang's 70th birthday

January 30-31, 2021

Beijing, China

http://conference.bicmr.pku.edu.cn/meeting/index?id=92

2021 PKU Workshop on
Optimization Theory and Methods

Dedicated to Prof. Yin Zhang's 70th birthday

January 30-31, 2021

Beijing, China

http://conference.bicmr.pku.edu.cn/meeting/index?id=92

Information for Participants

Online Conference at zoom

zoom ID: 691 0515 0322

Password: 393901

Link: https://zoom.com.cn/j/69105150322?pwd=Tm4vT2dMempueWtGdU1WVzRTM1o3Zz09

The zoom conference will be broadcasted at TencentMeeting at the same time.

ID: 931 327 826

Password: 123456

Link: https://meeting.tencent.com/s/p9oNqgmlsFyx

Contact Information

If you need any help, please feel free to contact

- Prof. Zaiwen Wen: wenzw@pku.edu.cn
- Zhenyuan Zhu: zhenyuanzhu@pku.edu.cn

Sponsors

National Natural Science Foundation of China

Beijing International Center for Mathematical Research,
Peking University

Academy of Mathematics and Systems Science,
Chinese Academy of Sciences

School of Data Sciences,
The Chinese University of Hong Kong-Shenzhen

National Center for Applied Mathematics, Chongqing

Committees

Organizing Committee

Xin Liu, Chinese Academy of Sciences

Zaiwen Wen, Peking University

Junfeng Yang, Nanjing University

Wotao Yin, University of California, Los Angeles

Scientific Committee

Zhiquan Luo, The Chinese University of Hong Kong-Shenzhen

Xinmin Yang, Chongqing Normal University

Yaxiang Yuan, Chinese Academy of Sciences

Yin Zhang, The Chinese University of Hong Kong-Shenzhen

Conference Schedule (Beijing time)

Each talk is 30 minutes + 5 minutes for questions.

January 30, Saturday

09:00-09:10 Opening Ceremony
Chair: Zhiquan Luo/Yuhong Dai

09:10-10:20 Session S1
Chair: Shuzhong Zhang

- **09:10-09:45 Renato D.C. Monteiro**, Complexity of proximal augmented Lagrangian type methods for solving nonconvex composite optimization problems with nonlinear convex constraints
- **09:45-10:20 Sam Burer**, Exact SDPs for a class of (random and non-Random) diagonal QCQPs

10:20-10:30 Break

10:30-11:40 Session S2
Chair: Xinmin Yang

- **10:30-11:05 Shuzhong Zhang**, Investigations of the first-order methods for the variational inequality problem
- **11:05-11:40 Yinyu Ye**, Bandits with knapsacks via online linear programming: a (problem-dependent) logarithmic regret bound

11:40-14:00 Lunch

14:00-15:10 Session S3
Chair: Zaiwen Wen

- **14:00-14:35 Shiqian Ma**, A Riemannian block coordinate descent method for computing the projection robust Wasserstein distance
- **14:35-15:10 Yuan Shen**, Alternating direction method of multipliers for some nonnegative matrix factorization-based models

15:10-15:20 Break

15:20-16:30 Session S4
Chair: Liwei Xu

- **15:20-15:55 Lijun Xu**, An efficient ADMM-type algorithm for deep semi-nonnegative matrix factorization
- **15:55-16:30 Junfeng Yang**, A golden ratio primal-dual algorithm for structured convex optimization

16:30-16:40 Break

16:40-17:50 Session S5
Chair: Junfeng Yang

- **16:40-17:15 Bo Yu**, Results on Portfolio Optimization: Model, Theory and Algorithms
- **17:15-17:50 Liwei Xu**, Stochastic Gauss-Newton algorithms for online PCA

January 31, Sunday

09:00-10:10 Session V1
Chair: Yangyang Xu

09:00-09:35 **Zizhuo Wang**, Assign-to-seat: dynamic capacity control for selling train tickets

09:35-10:10 **Zhijun Wu**, Social distancing as a population game in networked social environments

10:10-10:20 Break

10:20-11:30 Session V2
Chair: Zizhuo Wang

10:20-10:55 **Yangyang Xu**, Near-optimal first-order methods for nonlinear programs with $O(1)$ functional constraints

10:55-11:30 **Xin Liu**, A distributed and secure algorithm for computing dominant SVD based on projection splitting

11:30-12:00 Special Session for Prof. Yin Zhang
Chair: Xin Liu/Wotao Yin

12:00-14:00 Lunch

14:00-15:10 Session V3
Chair: Shiqian Ma

14:00-14:35 **Wotao Yin**, Deep Learning based on Decentralized Optimization

14:35-15:10 **Yuhong Dai**, Accelerating the Barzilai-Borwein gradient method by imposing two-dimensional quadratic termination property

15:10-15:20 Break

15:20-16:30 Session V4
Chair: Yuhong Dai

15:20-15:55 **Zhiquan Luo**, Optimal Physical Cell Id (PCI) Assignment via Convex Relaxation

15:55-16:30 **Marc Teboulle**, Faster Lagrangian-based methods in convex optimization

16:30-16:40 Break

16:40-18:00 Session V5
Chair: Zhiquan Luo

16:40-17:15 **Gang Bao**, Reconstruction of a Random Periodic Structure via Inverse Scattering

17:15-18:00 **Yin Zhang**, Artificial Neural Nets: Depth vs Width

Abstracts

Reconstruction of a Random Periodic Structure via Inverse Scattering

Gang Bao
Zhejiang University
baog@zju.edu.cn

An efficient numerical method is presented for the inverse scattering problem of a time-harmonic plane wave incident on a perfectly reflecting random periodic structure. The method is based on a novel combination of the Monte Carlo technique for sampling the probability space, a continuation method with respect to the wavenumber, and the KL expansion of the random structure, which reconstructs key statistical properties of the profile for the unknown random periodic structure from boundary measurements of the scattered fields away from the structure. Numerical results are presented to demonstrate the reliability and efficiency of the proposed method. The talk is based on a recent joint work with Yiwen Lin and Xiang Xu.

Exact SDPs for a class of (random and non-Random) diagonal QCQPs

Sam Burer
University of Iowa
samuel-burer@uiowa.edu

We study a class of quadratically constrained quadratic programs (QCQPs), called *diagonal QCQPs*, which contain no off-diagonal terms $x_j x_k$ for $j \neq k$, and we provide a sufficient condition on the problem data guaranteeing that the basic Shor semidefinite relaxation is exact. Our condition complements and refines those already present in the literature and can be checked in polynomial time. These ideas are extended to show that a class of random diagonal QCQPs has exact semidefinite relaxations with high probability as the number of variables grows, while the number of constraints stays fixed. To the best of our knowledge, this is the first result establishing the exactness of the semidefinite relaxation for random diagonal QCQPs.

Accelerating the Barzilai-Borwein gradient method by imposing two-dimensional quadratic termination property

Yuhong Dai
Chinese Academy of Sciences
dyh@lsec.cc.ac.cn

Since its proposition in Cauchy (1847), one milestone work along the gradient method is the Barzilai-Borwein (nonmonotone) method (1988), while another significant work is the Yuan stepsize in (2006), which leads to the appearance of the efficient Dai-Yuan (monotone) gradient method (2005). In this talk, a new gradient stepsize will be delivered at the motivation of equiping the Barzilai-Borwein method with two-dimensional quadratic termination property. A remarkable feature of the new stepsize is that its computation only depends on the Barzilai-Borwein stepsizes in two previous iterations, without the need for exact line searches and Hessian, and hence it can easily be extended for nonlinear optimization. By adaptively taking long Barzilai-Borwein steps and some short steps associated with the new stepsize, we develop an efficient gradient method for unconstrained optimization. The proposed method is further extended for box-constrained constrained optimization and singly linearly box-constrained optimization by incorporating nonmonotone line searches and gradient projection techniques. Numerical experiments demonstrate that the proposed method outperforms the most successful gradient methods in the literature. This is a joint work with Yakui Huang and Xinwei Liu.

A distributed and secure algorithm for computing dominant SVD based on projection splitting

Xin Liu
Chinese Academy of Sciences
liuxin@lsec.cc.ac.cn

We propose and study a distributed and secure algorithm for computing dominant (or truncated) singular value decompositions (SVD) of large and distributed data matrices. We consider the scenario where each node privately holds a subset of columns and only exchanges "safe" information with other nodes in a collaborative effort to calculate a dominant SVD for the whole matrix. In the framework of alternating direction methods of multipliers (ADMM), we propose a novel formulation for building consensus by equalizing subspaces spanned by splitting variables instead of equalizing themselves. This technique greatly relaxes feasibility restrictions and accelerates convergence significantly, while at the same time yielding simple subproblems. We design several algorithmic features, including a low-rank multiplier formula and mechanisms for controlling subproblem solution accuracies, to increase the algorithm's computational efficiency and reduce its communication overhead. More importantly, unlike many existing distributed or parallelized algorithms, our algorithm preserves the privacy of locally-held data; that is, none of the nodes can recover the data stored in another node through information exchanged during communications. We present convergence analysis results, including a worst-case complexity estimate, and extensive experimental results indicating that the proposed algorithm, while safely guarding data privacy, has a strong potential to deliver a cutting-edge performance, especially when communication costs are high. (Joint work with Mr. Lei Wang and Dr. Yin Zhang)

Optimal Physical Cell Id (PCI) Assignment via Convex Relaxation

Zhiquan Luo
Chinese University of Hong Kong-Shenzhen
luozq@cuhk.edu.cn

In this work (jointly with Chengpiao Huang and Honghao Zhang), we consider the problem of assigning a PCI value to each network cell in a 5G wireless communication network. Given an interference matrix, the optimal PCI assignment is to assign the given PCI values to the cell nodes so that the interfering cells have different PCI values (modulo 3 or 30). The PCI value problem can be formulated as a quadratic assignment problem and is NP-hard. We propose an effective convex relaxation algorithm which can give a high quality practical solution in minutes for problems with thousands of cells.

A Riemannian block coordinate descent method for computing the projection robust Wasserstein distance

Shiqian Ma
University of California, Davis
sqma@ucdavis.edu

The Wasserstein distance has become increasingly important in machine learning and deep learning. Despite its popularity, the Wasserstein distance is hard to approximate because of the curse of dimensionality. A recently proposed approach to alleviate the curse of dimensionality is to project the sampled data from the high dimensional probability distribution onto a lower-dimensional subspace, and then compute the Wasserstein distance between the projected data. However, this approach requires to solve a max-min problem over the Stiefel manifold, which is very challenging in practice. The only existing work that solves this problem directly is the RGAS algorithm, which requires to solve an entropy-regularized optimal transport problem in each iteration, and thus can be costly for large-scale problems. In this talk, we propose a Riemannian block coordinate descent (RBCD) method to solve this problem, which is based on a novel reformulation of the regularized max-min problem over the Stiefel manifold. We analyze the complexity of arithmetic operations for RBCD to obtain an ϵ-stationary point, and show that it significantly improves the corresponding complexity of RGAS. Moreover, our RBCD has very low per-iteration complexity, and hence is suitable for large-scale problems. Numerical results on both synthetic and real datasets demonstrate that our method is more efficient than existing methods, especially when the number of sampled data is very large.

Complexity of proximal augmented Lagrangian type methods for solving nonconvex composite optimization problems with nonlinear convex constraints

Renato D.C. Monteiro
Georgia Institute of Technology
monteiro@isye.gatech.edu

This talks reviews augmented Lagrangian type methods for solving cone convex (and hence including linearly) constrained nonconvex composite optimization (CCC-NCO) problems and discusses their corresponding complexity results. It then describes a more recent inner accelerated proximal inexact augmented Lagrangian (NL-IAPIAL) method for solving the CCC-NCO problem which consists of: a) applying an inexact proximal point method to approximately solve a sequence of quadratic penalty subproblems for increasing values of the penalty parameter, and; b) solving all the generated prox subproblems by an accelerated composite gradient (ACG) method. It is shown that the method, started from any infeasible point x_0 generates a ρ-approximate stationary point in $\mathcal{O}(\rho^{-3})$ ACG iterations. This improves upon results obtained for previous algorithms in that: 1) it allows for the presence of a composite term in the objective function; 2) the feasible region does not have to be bounded; 3) the initial point does not have to be feasible; and 4) substantially reduces the previously known iteration complexity of $\mathcal{O}(\rho^{-6})$.

Alternating direction method of multipliers for some nonnegative matrix factorization-based models

Yuan Shen
Nanjing University Of Finance & Economics
ocsiban@126.com

We consider several non-negative matrix factorization (NMF)-based models, which could be equivalent to the classical k-means clustering problem under certain conditions. They are solved by the popular alternating direction method of multipliers (ADMM). As is well known that ADMM is not necessarily convergent for a nonconvex model such as NMF model, but it empirically works well for a wide range of nonconvex models. We establish the convergence result of ADMM for solving these models under mild conditions. Preliminary experimental results justified their performance by comparing them with the popular k-means heuristic approach in terms of clustering performance indices such as ARI or NMI.

Faster Lagrangian-based methods in convex optimization

Marc Teboulle
Tel Aviv University
teboulle@tauex.tau.ac.il

This talk presents a simple unifying framework to analyze and improve the convergence rate analysis of Lagrangian-based methods for convex optimization problems. Towards this goal we first introduce the notion of a nice primal algorithmic map, which plays a central role in the unification and in the simplification of the analysis of all Lagrangian-based methods. Equipped with a nice primal algorithmic map, we then define a versatile generic scheme, which allows for the design and analysis of Faster LAGrangian (FLAG) methods with new provably sublinear rate of convergence expressed in terms of functions values and feasibility violation of the original (non-ergodic) generated sequence. To demonstrate the power and versatility of our approach and results, we show that all well-known iconic Lagrangian-based schemes admit a nice primal algorithmic map, and hence share the new faster rate of convergence results within their corresponding FLAG.

Assign-to-seat: dynamic capacity control for selling train tickets

Zizhuo Wang
Chinese University of Hong Kong, Shenzhen
wangzizhuo@cuhk.edu.cn

We consider a revenue management problem arising from the sales of high-speed train tickets in China. Compared to traditional network revenue management problems, the new feature of our problem is that each request, if accepted, needs to be instantly assigned to a unique seat throughout the whole journey, and later adjustment is not allowed. Therefore, when making decisions, the seller needs to track not only the total capacity available but also the status of each seat. We build a modified network revenue management model for this problem. We first study a static problem where all requests are given. Despite that the problem is NP-Hard in general, we identify conditions of the capacity matrix for polynomial-time solvability, based on which we propose a booking limit control policy which achieves asymptotic optimality. We then introduce a bid-price control policy based on a novel maximal sequence idea, which allows for non-linearity in bid-prices and thus results in a more accurate approximation of the value function than the traditional bid-price control policy. Finally, combining the dynamic view of the maximal sequence and the static solution of the primal problem, we propose a re-solving a dynamic primal (RDP) policy which achieves uniformly bounded revenue loss under a mild assumption. Numerical experiments using both synthetic and real data show the advantage of our proposed policies in improving capacity allocation efficiency.

Social distancing as a population game in networked social environments

Zhijun Wu
Iowa State University
zhijun@iastate.edu

While social living is considered to be an indispensable part of human life in today's ever-connected world, social distancing has recently received much public attention on its importance since the outbreak of the coronavirus pandemic. In fact, social distancing has long been practiced in nature among solitary species, and been taken by human as an effective way of stopping or slowing down the spread of infectious diseases. Here we consider a social distancing problem for how a population, when in a world with a network of social sites, decides to visit or stay at some sites while avoiding or closing down some others so that the social contacts across the network can be minimized. We model this problem as a population game, where every individual tries to find some network sites to visit or stay so that he/she can minimize all his/her social contacts. In the end, an optimal strategy can be found for everyone, when the game reaches an equilibrium. We show that a large class of equilibrium strategies can be obtained by selecting a set of social sites that forms a so-called maximal r-regular subnetwork. The latter includes many well studied network types, which are easy to identify or construct, and can be completely disconnected (with r = 0) for the most-strict isolation, or allow certain degrees of connectivity (with r ¿ 0) for more flexible distancing. We derive the equilibrium conditions of these strategies, and analyze their rigidity and flexibility on different types of r-regular subnetworks. We also extend our model to weighted networks, when different contact values are assigned to different network sites.

An efficient ADMM-type algorithm for deep semi-nonnegative matrix factorization

Lijun Xu
Dalian Maritime University
lijun_xu@dlmu.edu.cn

In this work, we focus on deep semi-nonnegative matrix factorization (DSemiNMF) which has a wider application in the real world than traditional NMF. We propose an efficient algorithm based on the classic alternating direction method of multipliers (ADMM) for DSemiNMF. By utilizing structures in DSemiNMF, we derive an efficient updating rule for updating subproblems according to its KKT conditions. Numerical experiments are conducted to compare the proposed algorithm with state-of-the-art deep semi-NMF algorithm. Results show that our algorithm performs better in consuming time and the deep model can indeed obtain better clustering accuracy than the single-layer model.

Stochastic Gauss-Newton algorithms for online PCA

Liwei Xu
University of Electronic Science and Technology of China
xul@uestc.edu.cn

In this talk, we introduce a symmetric low rank product model for dominant eigenspace calculation associted with the online principal component analysis (PCA) problem. Two Gauss-Newton algorithms with dimilishing stepsize and adaptive stepsize are proposed, respectively. We construct the global convergence of the diminishing version and show its robustness. Moreover, the adaptive version illustrates its extraordinary numerical performance. Finally, we propose a variance reduced version of the Stochastic Gauss-Newton (SGN) which is of diminishing variance without full gradient calculation.

Near-optimal first-order methods for nonlinear programs with $O(1)$ functional constraints

Yangyang Xu
Rensselaer Polytechnic Institute
xuy21@rpi.edu

First-order methods (FOMs) have recently been applied and analyzed for solving problems with complicated functional constraints. Existing works show that FOMs for functional constrained problems have lower-order convergence rates than those for unconstrained problems. In particular, an FOM for a smooth strongly-convex problem can have linear convergence, while it can only converge sublinearly for a constrained problem if the projection onto the constraint set is prohibited. In this talk, I will show that the slower convergence is caused by the large number of functional constraints but not the constraints themselves. When there are only $O(1)$ functional constraints, I will show that an FOM can have almost the same convergence rate as that for solving an unconstrained problem, even without the projection onto the feasible set. In addition, given an $\varepsilon > 0$, I will show that a complexity result that is better than a lower bound can be obtained, if there are only $o(\varepsilon^{-\frac{1}{2}})$ functional constraints. The result is surprising but does not contradict to the existing lower complexity bound, because I focus on a specific subclass of problems. Experimental results on quadratically-constrained quadratic programs will be shown to demonstrate the theory.

A golden ratio primal-dual algorithm for structured convex optimization

Junfeng Yang
Nanjing University
jfyang@nju.edu.cn

Consider minimizing the sum of two closed proper convex functions, one of which involves a composition with a linear transform. We propose a golden ratio primal-dual algorithm (GRPDA), which is full splitting in the sense that the per iteration cost is dominated by the evaluation of the proximal point mappings of the two component functions and two matrix-vector multiplications. We show that GRPDA converges within a broader range of parameters than the classical primal-dual algorithm. An $O(1/N)$ ergodic convergence rate result is established based on the primal-dual gap function, where N denotes the number of iterations. When either the primal or the dual problem is strongly convex, an accelerated GRPDA is constructed to improve the ergodic convergence rate from $O(1/N)$ to $O(1/N^2)$. Numerical results on LASSO, nonnegative least-squares and minimax matrix game problems are given to demonstrate the efficiency of the proposed algorithms.

Joint with Xiaokai Chang from Lanzhou University of Science and Technology.

Bandits with knapsacks via online linear programming: a (problem-dependent) logarithmic regret bound

Yinyu Ye
Stanford University
yinyu-ye@stanford.edu

The Bandit problem is one of the fundamental topics in Reinforced Learning. In this talk, we consider the problem of bandits with knapsacks (BwK) from linear programming (LP) perspective and propose algorithms that achieve problem-dependent logarithmic regret bound. The problem dependence is characterized by the reduced-cost of each arm and the slackness of the resource constraints based on its underlying LP formulation, where the reduced-cost can be interpreted as an optimality/sub-optimality classification measure of each arm. We first discuss a case where all the reduced-costs are zero, i.e., all the arms are "optimal", in which case we show that the regret of any BwK algorithm is upper bounded by the remaining resources of the binding constraints at the horizon end. Inspired by the recent advances of the online LP analysis, we propose an adaptive sampling algorithm and derive a regret bound that has no dependence on the length of horizon T. Next, we consider the case that there exist non-optimal arms which possess non-zero reduced-cost values. Assuming the knowledge of the smallest among such values, we develop a first-explore-then-exploit algorithm that achieves a logarithmic regret bound in terms of T. The algorithm utilizes the knowledge for the design of the exploration phase, and an LP stability result ensures a high-probability identification of the optimal arms (who have zero reduced-costs). Finally, we relax the assumption on the prior knowledge of the critical value and propose a primal-dual analysis for the problem, which enables us to develop an elimination-based algorithm and achieve a logarithmic regret bound. Precisely, the regret bound is $O(\log T/\delta^2)$, where δ is the smallest nonzero reduced-cost value of both primal and dual LPs that pertain the problem. To the best our knowledge, this is the first problem-dependent logarithmic regret bound for solving the general BwK problem.

Joint work with Xiaocheng Li and Chunlin Sun.

Deep Learning based on Decentralized Optimization

Wotao Yin
University of California, Los Angeles
wotaoyin@math.ucla.edu

Communication is the bottleneck in distributed optimization, such as training large neural networks in a cluster. We propose to reduce communication costs by borrowing techniques from decentralized optimization, which is traditionally used in wireless communication/control networks. In a cluster, we restrict message passing to virtual subgraphs, which are sparse and may change over time. Using graph spectral analysis, we obtain improved communication and time complexities for distributed optimization. The results are implemented in BlueFog, a new open-source framework for both distributed traditional optimization and PyTorch-based neural network training. BlueFog is faster than state-of-the-art industrial packages such as Horovood.

This is joint work with Bicheng Ying (Google), Kun Yuan (Alibaba), Hanbin Hu (UCSB), and Ji Liu (Baidu).

Results on Portfolio Optimization: Model, Theory and Algorithms

Bo Yu
Dalian University of Technology
yubo@dlut.edu.cn

In this talk, I will give a brief introduction to our recent works on portfolio optimization. At first, a new distributionally robust modeling strategy based on kernel density estimation will be introduced, tractable reformulation of KDE distributionally robust portfolio optimization models, with CVaR, EVaR and HMCR as risk measures respectively, will be shown. Then some efficient algorithms for solving sparse portfolio optimization as well as their convergence will be introduced. Numerical results will also be given to show good performance of our new models and efficiency of our new algorithms.

Investigations of the first-order methods for the variational inequality problem

Shuzhong Zhang
University of Minnesota
zhangs@umn.edu

In this talk we discuss numerous first order algorithms for solving variational inequality (VI) problems, including the "heavy-ball" method, the extra-gradient algorithm, and the optimistic gradient method. We shall also discuss how these methods can be applied to solve a stochastic version of the VI problems, with applications from computing equilibrium solutions under a stochastic setting.

(joint work with Kevin Huang)

Artificial Neural Nets: Depth vs Width

Yin Zhang
The Chinese University of Hong Kong-Shenzhen
yinzhang@cuhk.edu.cn

I will share observations from some simple experiments and offer interpretations.