2018 Workshop on New Computing-Driven Opportunities for Optimization

August 12-18, 2018

Wuyi, Fujian, China

http://bicmr.pku.edu.cn/~wenzw/wuyishan.html

Scan the QR code at the right side to download the most recent program

2018 Workshop on New Computing-Driven Opportunities for Optimization

August 12-18, 2018

Wuyi, Fujian, China

http://bicmr.pku.edu.cn/~wenzw/wuyishan.html

Information for Participants

Conference Hotel

Hotel:	Youchen Hotel
Address:	Bai Hua Read, Wuyishan City 武夷山市囿臣酒店，武夷山市百花路
Dates:	By default, the hotel room is reserved from August 12th (check in) to August 18th (check-out). Please let us know if you have a different arrival-departure schedule.
Arrival:	The city of Wuyishan is connected to all major entry cities of China (Beijing, Shanghai, Guangzhou, etc.) via both direct flights and high-speed rail trains. Volunteers will pick attendees up from the airport and three train stations of Wuyishan. They will hold papers with the name of the workshop.

Meals

- Breakfasts, 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 6.81 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.

Contact Information

If you need any help, please feel free to contact

- Prof. Zaiwen Wen: wenzw@pku.edu.cn
- Prof. Zheng Peng: pzheng@fzu.edu.cn, +86-18659168012
- Jiang Hu: jianghu@pku.edu.cn

Sponsors

National Natural Science Foundation of China

Tianyuan Foundation, National Natural Science Foundation of China

The Mathematical Society of China

Fuzhou University

Wuyi University

Committees

Organizers

Thorsten Koch, Zuse Institute Berlin, Technische Universitaet Berlin, Germany

Zaiwen Wen, Peking University

Wotao Yin, University of California, Los Angeles

Yaxiang Yuan, Chinese Academy of Sciences

Local Organizers

An Chang, Fuzhou University

Genghua Fan, Fuzhou University

Yonglin Liu, Wuyi University

Zheng Peng, Fuzhou University

Wenxing Zhu, Fuzhou University

Organization Format

- This workshop does not organize talks in the traditional fashion. To sparkle new directions and collaborations, there will be
  - 8 overview sessions (each has a 30-minute presentation and a 15-minute discussion by one speaker)
  - 11 one-hour topics sessions (each has a 30-minute presentation and a 30-minute discussion by one to three discussion leads)
  - 6 free discussion sessions (each has 30 minutes by one to three leads)
- An open poster session. Any poster file sent to us by August 1 will be printed at Wuyishan for the authors, free of charge.
- Arrangement for all participants:
  - Participants can submit a few slides to the sessions (even if they are not designated as speakers or discussion leads. Coordination information will be sent around May 15).
  - The participants are supposed to join all sessions that they possibly can contribute to. They are expected to take part in the discussion actively. We should stress that the discussion is interactive and the success depends a lot on the participation of all participants.
  - For each session, a participant will be asked to take notes during the discussion. She/he will also work with the discussion leads on a summary of the main points raised in the discussion.
- For discussion efficency, we ask the lead(s) of each one-hour discussion to:
  - prepare (or organize a team to prepare) a few slides to introduce the topic, the background and state of the art for those not working in the topic field;
  - prepare (or organize a team to prepare) a slide with a few topics and a few questions;
  - give a presentation in 15-30 minutes (with slides and a speech) and leave time for discussions;
  - moderating the discussion and get the audience to participate in the discussion.
- A report will be released after the workshop. For the talk in each session, we will ask a brief abstract, a list of co-authors and references if available. For the discussion in each session, we will ask an abstract which briefly summarizes the main points raised in the discussion.

Workshop Schedule

August 13, Monday

09:15-10:30 **Introduction**
Leads: Thorsten Koch/Wotao Yin/Zaiwen Wen/Yaxiang Yuan

10:30-10:40 **Group Photo**

10:40-11:00 **Break**

11:00-12:00 **Topic: Integer programming, MIP**
Leads: Thorsten Koch, Noter: Cheng Lu

12:00-13:45 **Lunch**

13:45-14:30 **Overview: Randomization in Optimization Algorithm Design – How Much is Too Much**
Leads: Yinyu Ye, Noter: Jinyan Fan

14:30-15:00 **Break**

15:00-16:00 **Topic: Building (stochastic) First-order Methods from Monotone Operators**
Leads: Wotao Yin, Noter: Andre Milzarek

16:00-16:30 **Break**

16:30-17:30 **Topic: First-order Methods for Large-scale Optimization**
Leads: Zhaosong Lu, Noter: Zi Xu

17:30-18:00 **Topic: Cross Domain Example Problems**
Leads: Shuzhong Zhang/Zhiquan Luo, Noter: Yafeng Liu

18:00 **Dinner**

August 14, Tuesday

09:15-10:00 **Overview: MINLP**
Leads: Jon Lee, Noter: Yong Xia

10:00-10:30 **Break**

10:30-11:30 **Topic: Non-Convex Optimization for Machine Learning, Signal Processing, and Statistics**
Leads: Anthony So, Noter: Ke Wei

11:30-12:00 **Connections between our fields I**
Leads: Jie Sun/Wotao Yin, Noter: Cong Sun

12:00-13:45 **Lunch**

13:45-14:30 **Overview: Parallel MILP**
Leads: Ted Ralphs, Noter: Hailin Sun

14:30-15:00 **Break**

15:00-16:00 **Topic: Optimization with orthogonality constraints**
Leads: Yin Zhang/Xin Liu, Noter: Bo Jiang (Nanjing)

16:00-16:30 **Break**

16:30-17:30 **Topic: ADMM**
Leads: Xiaoming Yuan/Deren Han, Noter: Peng Zheng

17:30-18:00 **Preparation for posters**

18:00 **Dinner**

19:00-21:00 **Poster Session**

August 15, Wednesday

06:00-07:00 **Breakfast**

06:00-12:00 **Excursion**

12:00-13:45 **Lunch**

15:15-16:00 **Non-problems in Optimization for Statistics**
Leads: Jongshi Pang, Noter: Caihua Chen

16:00-16:30 **Break**

16:30-17:30 **Topic: Advances in Globalization Techniques**
Leads: Serge Gratton, Noter: Zaikun Zhang

17:30-18:00 **Open and Challenging Academic Problems**
Leads: Yinyu Ye/Defeng Sun, Noter: Hui Zhang

18:00 **Dinner**

August 16, Thursday

09:15-10:00 Overview: Theory and Algorithms for Two-stage Stochastic Variational Inequalities
Leads: Xiaojun Chen, Noter: Junfeng Yang

10:00-10:30 Break

10:30-11:30 Topic: Traditional Nonlinear Programming
Leads: Yuhong Dai/Hongchao Zhang, Noter: Cong Sun

11:30-12:00 Teaching, Students, Publication
Leads: Deren Han/Zhenyue Zhang, Noter: Wei Bian

12:00-13:45 Lunch

13:45-14:30 Overview: Tensor and Polynomial Optimization
Leads: Shuzhong Zhang, Noter: Bo Jiang (SHUFE)

14:30-15:00 Break

15:00-16:00 Topic: Some Recent Development of Deep Modeling in Medical Imaging
Leads: Bin Dong/Zhihua Zhang, Noter: Ming Yan

16:00-16:30 Break

16:30-17:30 Topic: Recent Advances on Optimization for Machine Learning
Leads: Zhouchen Lin/Shiqian Ma, Noter: Lingfeng Niu

17:30-18:00 Funding Opportunities, Outreach Other Fields
Leads: Yuhong Dai/Xinmin Yang, Noter: Zi Xu

18:00 Dinner

19:00-21:00 Banquet

August 17, Friday

09:15-10:00 **Overview: Semi-smooth Newton method**
Leads: Defeng Sun, Noter: Chao Ding

10:00-10:30 **Break**

10:30-11:30 **Topic: Optimization and Signal Processing**
Leads: Zhiquan Luo/Mingyi Hong, Noter: Yafeng Liu

11:30-12:00 **Connections between our fields II**
Leads: Zhouchen Lin/Bin Dong, Noter: Junfeng Yang

12:00-13:45 **Lunch**

13:45-14:30 **Overview: Algorithms for SDP and Conic Solvers**
Leads: Kim-Chuan Toh, Noter: Ying Cui

14:30-15:00 **Break**

15:00-15:30 **Next steps & Follow-ups**
Leads: Thorsten Koch/Wotao Yin/Zaiwen Wen/Yaxiang Yuan

18:00 **Dinner**

Abstracts

Overview and Topic Sessions

Integer Programming, MIP

Koch Thorsten

Zuse Institute Berlin, Technische Universitäet Berlin

We will take a tour through computational integer programming. Starting with an introduction into the basic techniques, we will progress to an overview on the current state-of-the-art and end up with the challenges ahead of us. The later includes connecting to other areas as NLP and CP.

Randomization in Optimization Algorithm Design – How Much is Too Much

Yinyu Ye
Stanford University

Randomization technique has been widely used in designing optimization algorithms to achieve properties that deterministic algorithms cannot do, such as randomized SGD, BCD, ADMM, Reinforced Learning etc. In this talk, we use sampling-based Value-Iteration for Markov decision process (MDP) and randomly assembled DMM to illustrate that randomness needs to be carefully managed to take advantage of the technique for large scale optimization.

Building (stochastic) First-order Methods from Monotone Operators

Wotao Yin
University of California, Los Angeles

We will overview a wide range of first-order optimization methods that can be built from a few concepts and tricks in the monotone operator theory. They include gradient, proximal-gradient, (proximal) method of multipliers, alternating minimization, PDHG, Chambolle Pock, Condat-Vu, (standard, proximal, and linearized) ADMM, and PD3O. Finite-sum and block-coordinate-friendly properties are used to develop parallel and asynchronous methods. However, we will leave out topics such as line search, Nesterov/heavy-ball accelerations, conditional gradients, and second-order methods.

Co-author: Ernest K. Ryu.

First-order Methods for Large-scale Optimization

Zhaosong Lu
Simon Fraser University

In this talk we overview some development of first-order methods for solving large-scale optimization problems. In particular, we review subgradient methods, proximal gradient methods, stochastic gradient methods, incremental gradient methods, block-coordinate proximal gradient methods, cutting plane method, first-order penalty and augmented Lagrangian methods. Also, we discuss some future research in first-order methods.

MINLP

Jon Lee
University of Michigan

MINLP (Mixed-integer nonlinear programming) refers to global optimization in the context of nonlinear functions and a combination of continuous and discrete variables. Looking at the entire class of such problems is absurd. But thinking about the big picture leads to tractable subclasses (from both theoretical and practical viewpoints) as well as hardness results (again from both theoretical and practical viewpoints). The different viewpoints inform one another, helping us build effective algorithms for meaningful subclasses. I will give a bit of a sermon about this way of thinking, as well as describe the major algorithmic paradigms and where they are headed. In particular, we have algorithms (notably NLP-based branch-and-bound, outer approximation, and hybrids) aimed at models that are convex once the discrete variables are relaxed, and we have algorithms (notably spatial branch-and-bound) aimed at "factorable" non-convex formulations. Big issues that are being grappled with include: (i) the tradeoff between relaxation simplicity and tightness (important at both the modeling and algorithm/software levels), (ii) methods for non-factorable (and at the extreme, black-box) models, (iii) integration of extreme convex-optimization (i.e., conic algorithms / disciplined convex programming) with other techniques, (iv) identification and exploitation of convexity, and of course (v) robustness and scaling.

Non-Convex Optimization for Machine Learning, Signal Processing, and Statistics

Anthony So

The Chinese University of Hong Kong

Optimization is now widely reckoned to be an indispensable tool in machine learning, signal processing, and statistics. Indeed, many of the advances in these fields rely crucially on the formulation of suitable optimization models and deployment of efficient numerical optimization algorithms. Over the past decades, convex optimization has emerged as a powerful and by far the most extensively used paradigm for tackling signal processing and machine learning applications. This is largely due to the fact that convex optimization problems often possess favorable theoretical and computational properties, and that many problems of practical interest have been shown to admit convex formulations or good convex approximations. By contrast, even though many applications in these fields admit natural non-convex formulations, efforts to develop techniques to handle them directly have been more scattered. This is perhaps not very surprising, as it is a common belief that non-convex optimization problems are intractable and lack nice theoretical properties. While such belief is justified when general non-convex optimization problems are concerned, existing results on a number of non-convex formulations of practical problems suggest quite the opposite. First, many applications from machine learning, signal processing, and statistics give rise to non-convex formulations that are well structured. In particular, they often exhibit properties akin to those of convex optimization problems and can be solved to optimality more efficiently than their convex reformulations or approximations. For instance, it has been shown that various non-convex formulations of the low-rank matrix factorization problem have no spurious local minima (i.e., all local minima are global) [3, 8, 20, 21, 12], and that simple first-order methods (such as alternating minimization or gradient descent) for solving such formulations will converge linearly to a globally optimal solution [10, 34, 33, 27]. Similar results have been established for non-convex formulations of a number of applications, including blind deconvolution [32, 11, 13], deep neural networks [28, 29], dictionary recovery over the sphere [25, 26], MIMO detection [17], phase retrieval [6, 19, 24, 31], phase synchronization [5, 16], source localization [14], and tensor decomposition [1, 9]. Second, it is becoming clear that the use of convex reformulation or approximation techniques may not always be adequate or even desirable for dealing with contemporary signal processing and machine learning applications. This necessitates the development of new algorithmic approaches to tackle structured non-convex optimization problems directly. One example is the sparse linear regression problem. A popular approach to tackling it is to approximate the non-convex sparsity-inducing ℓ_0-(pseudo)norm by the convex ℓ_1-norm, thus leading to a convex optimization problem known as the LASSO. Despite its computational tractability, the LASSO has several shortcomings as a variable selection and prediction procedure when noise and correlated variables are present. Many different non-convex regularizers have been proposed to address these shortcomings [30, 35], and these have motivated the development of, e.g., non-convex non-Lipschitz optimization [4] and mixed-integer optimization [2, 15, 18] methods to solve the resulting non-convex formulations. Another example is resource allocation designs in next-generation communication networks (such as sum-rate maximization and max-min-fair optimization in MIMO networks), which usually involve large-scale non-convex optimization problems. This has led to the development of parallel and distributed methods for solving such problems [7, 22, 23].

From the above discussion, we see that non-convex optimization techniques could have a significant impact on both our theoretical understanding of and algorithmic capability to handle contemporary machine learning, signal processing, and statistics applications. In this session, we aim to provide insights into how structures of the non-convex formulations of various practical problems can be exploited in algorithm design, showcases some notable successes in this line of study, and identifies important research issues that are motivated by existing or emerging applications.

Bibliography

[1] A. Anandkumar, R. Ge, and M. Janzamin. Analyzing Tensor Power Method Dynamics in Overcomplete Regime. *Journal of Machine Learning Research*, 18(22):1–40, 2017.

[2] D. Bertsimas, A. King, and R. Mazumder. Best Subset Selection via a Modern Optimization Lens. *The Annals of Statistics*, 44(2):813–852, 2016.

[3] S. Bhojanapalli, B. Neyshabur, and N. Srebro. Global Optimality of Local Search for Low Rank Matrix Recovery. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29: Proceedings of the 2016 Conference*, pages 3873–3881, 2016.

[4] W. Bian and X. Chen. Worst–Case Complexity of Smoothing Quadratic Regularization Methods for Non–Lipschitzian Optimization. *SIAM Journal on Optimization*, 23(3):1718–1741, 2013.

[5] N. Boumal. Nonconvex Phase Synchronization. *SIAM Journal on Optimization*, 26(4):2355–2377, 2016.

[6] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase Retrieval via Wirtinger Flow: Theory and Algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

[7] P. Di Lorenzo and G. Scutari. NEXT: In–Network Nonconvex Optimization. *IEEE Transactions on Signal and Information Processing over Networks*, 2(2):120–136, 2016.

[8] R. Ge, J. D. Lee, and T. Ma. Matrix Completion has No Spurious Local Minima. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett,

editors, *Advances in Neural Information Processing Systems 29: Proceedings of the 2016 Conference*, pages 2973–2981, 2016.

[9] R. Ge and T. Ma. On the Optimization Landscape of Tensor Decompositions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30: Proceedings of the 2017 Conference*, pages 3653–3663, 2017.

[10] P. Jain, P. Netrapalli, and S. Sanghavi. Low–Rank Matrix Completion Using Alternating Minimization. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC 2013)*, pages 665–674, 2013.

[11] P. Jung, F. Krahmer, and D. Stger. Blind Demixing and Deconvolution at Near–Optimal Rate. *IEEE Transactions on Information Theory*, 64(2):704–727, 2018.

[12] Q. Li, Z. Zhu, and G. Tang. The Non–Convex Geometry of Low–Rank Matrix Optimization. Accepted for publication in *Information and Inference: A Journal of the IMA*, 2018.

[13] X. Li, S. Ling, T. Strohmer, and K. Wei. Rapid, Robust, and Reliable Blind Deconvolution via Nonconvex Optimization. Accepted for publication in *Applied and Computational Harmonic Analysis*, 2018.

[14] H. Liu, Y.-M. Pun, and A. M.-C. So. Local Strong Convexity of Maximum–Likelihood TDOA–Based Source Localization and Its Algorithmic Implications. In *Proceedings of the IEEE 7th International Workshop on Computational Advances in Multi–Sensor Adaptive Processing (CAMSAP 2017)*, pages 1–5, 2017.

[15] H. Liu, T. Yao, and R. Li. Global Solutions to Folded Concave Penalized Nonconvex Learning. *The Annals of Statistics*, 44(2):629–659, 2016.

[16] H. Liu, M.-C. Yue, and A. M.-C. So. On the Estimation Performance and Convergence Rate of the Generalized Power Method for Phase Synchronization. *SIAM Journal on Optimization*, 27(4):2426–2446, 2017.

[17] H. Liu, M.-C. Yue, A. M.-C. So, and W.-K. Ma. A Discrete First-Order Method for Large-Scale MIMO Detection with Provable Guarantees. In *Proceedings of the 18th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2017)*, pages 669–673, 2017.

[18] R. Mazumder and P. Radchenko. The *Discrete Dantzig Selector*: Estimating Sparse Linear Models via Mixed Integer Linear Optimization. *IEEE Transactions on Information Theory*, 63(5):3053–3075, 2017.

[19] P. Netrapalli, P. Jain, and S. Sanghavi. Phase Retrieval Using Alternating Minimization. *IEEE Transactions on Signal Processing*, 63(18):4814–4826, 2015.

[20] D. Park, A. Kyrillidis, C. Caramanis, and S. Sanghavi. Finding Low–Rank Solutions to Smooth Convex Problems via the Burer–Monteiro Approach. In *Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing*, pages 439–446, 2016.

[21] D. Park, A. Kyrillidis, C. Caramanis, and S. Sanghavi. Non–Square Matrix Sensing without Spurious Local Minima via the Burer–Monteiro Approach. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017)*, pages 65–74, 2017.

[22] G. Scutari, F. Facchinei, and L. Lampariello. Parallel and Distributed Methods for Constrained Nonconvex Optimization—Part I: Theory. *IEEE Transactions on Signal Processing*, 65(8):1929–1944, 2017.

[23] G. Scutari, F. Facchinei, L. Lampariello, S. Sardellitti, and P. Song. Parallel and Distributed Methods for Constrained Nonconvex Optimization—Part II: Applications in Communications and Machine Learning. *IEEE Transactions on Signal Processing*, 65(8):1945–1960, 2017.

[24] J. Sun, Q. Qu, and J. Wright. A Geometric Analysis of Phase Retrieval. Accepted for publication in *Foundations of Computational Mathematics*, 2017.

[25] J. Sun, Q. Qu, and J. Wright. Complete Dictionary Recovery Over the Sphere I: Overview and the Geometric Picture. *IEEE Transactions on Information Theory*, 63(2):853–884, 2017.

[26] J. Sun, Q. Qu, and J. Wright. Complete Dictionary Recovery Over the Sphere II: Recovery by Riemannian Trust–Region Method. *IEEE Transactions on Information Theory*, 63(2):885–914, 2017.

[27] R. Sun and Z.-Q. Luo. Guaranteed Matrix Completion via Non–convex Factorization. *IEEE Transactions on Information Theory*, 62(11):6535–6579, 2016.

[28] L. Venturi, A. Bandeira, and J. Bruna. Neural Networks with Finite Intrinsic Dimension Have No Spurious Valleys. Manuscript, available at https://arxiv.org/abs/1802.06384, 2018.

[29] C. Yun, S. Sra, and A. Jadbabaie. A Critical View of Global Optimality in Deep Learning. Manuscript, available at https://arxiv.org/abs/1802.03487, 2018.

[30] C.-H. Zhang and T. Zhang. A General Theory of Concave Regularizations for High Dimensional Sparse Estimation Problems. *Statistical Science*, 27(4):576–593, 2012.

[31] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi. A Nonconvex Approach for Phase Retrieval: Reshaped Wirtinger Flow and Incremental Algorithms. *Journal of Machine Learning Research*, 18(141):1–35, 2017.

[32] Y. Zhang, Y. Lau, H.-W. Kuo, S. Cheung, A. Pasupathy, and J. Wright. On the Global Geometry of Sphere–Constrained Sparse Blind Deconvolution. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, pages 4381–4389, 2017.

[33] T. Zhao, Z. Wang, and H. Liu. A Nonconvex Optimization Framework for Low Rank Matrix Estimation. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28: Proceedings of the 2015 Conference*, pages 559–567, 2015.

[34] Q. Zheng and J. Lafferty. A Convergent Gradient Descent Algorithm for Rank Minimization and Semidefinite Programming from Random Linear Measurements. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and

R. Garnett, editors, *Advances in Neural Information Processing Systems 28: Proceedings of the 2015 Conference*, pages 109–117, 2015.

[35] Z. Zheng, Y. Fan, and J. Lv. High Dimensional Thresholded Regression and Shrinkage Effect. *Journal of the Royal Statistical Society, Series B*, 76(3):627–649, 2014.

Overview of Parallel MILP

Ted Ralphs
Lehigh University

In this talk, we will survey the current state of research in parallel methods for solving mixed integer linear optimization problems (MILPs). Much effort has gone into developing methods for parallelizing the solution of MILPs over the past several decades and a large number of mature parallel MILP solvers (both commercial and open source) now exist. Surveying the landscape, it appears that the progress has been meager overall. We discuss the challenges in parallelizing current algorithms, the current state of the art, why we have struggled, and what the future may hold.

Co-authors:

- Laszlo Ladanyi (CPLEX)
- Matthew Saltzman (Clemson University)
- Yan Xu (SAS Institute)
- Thorsten Koch (ZIB)
- Yuji Shinano (ZIB)
- Steve Maher (Lancaster University)

Optimization with Orthogonality Constraints

Yin Zhang
The Chinese University of Hong Kong, Shenzhen

Xin Liu
Chinese Academy of Sciences

Zaiwen Wen
Peking University

This session covers minimization with respect to a matrix X subject to orthogonality constraints $X^\top X = I$. The review is split into four parts: 1) a few interesting applications from combinatorial optimization, eigenvalue problems, the total energy minimization in electronic structure calculation, etc. These problems are generally difficult because the constraints are not only non-convex but also numerically expensive to preserve during iterations. 2) a few recent advance for solving these problems including gradient type methods and second-order type methods. 3) a few specific analysis on achieving global optimality. 4) a few challenging issues for future investigation.

References

1. P.-A. Absil, C. G. Baker, and K. A. Gallivan, Trust-region methods on Riemannian manifolds, Foundations of Computational Mathematics, 7 (2007), pp. 303–330.

2. P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, Princeton, NJ, 2008.

3. Afonso S. Bandeira, Nicolas Boumal, and Vladislav Voroninski, On the low-rank approach for semidefinite programs arising in synchronization and community detection, (2016).

4. Jiang Hu, Zaiwen Wen, and Yaxiang Yuan, Adaptive regularized Newton method for riemannian optimization, SIAM J. Matrix Anal. Appl., 2018

5. Xin Liu, Zaiwen Wen, Xiao Wang, Michael Ulbrich, and Yaxiang Yuan, On the analysis of the discretized Kohn-Sham density functional theory, SIAM J. Numer. Anal., 53 (2015), pp. 1758–1785.

6. Xin Liu, Zaiwen Wen, and Yin Zhang, Limited memory block Krylov subspace optimization for computing dominant singular value decompositions, SIAM J. Sci. Comput., 35 (2013), pp. A1641–A1668.

7. Xin Liu, Zaiwen Wen, and Yin Zhang, An efficient Gauss-Newton algorithm for symmetric low-rank product matrix approximations, SIAM J. Optim., 25 (2015), pp. 1571–1608.

8. Bart Vandereycken, Low-rank matrix completion by riemannian optimization, SIAM J. Optim., 23 (2013), pp. 1214–1236.

9. Zaiwen Wen, Andre Milzarek, Michael Ulbrich, and Hongchao Zhang, Adaptive regularized self-consistent field iteration with exact Hessian for electronic structure calculation, SIAM J. Sci. Comput., 35 (2013), pp. A1299–A1324.

10. Zaiwen Wen and Wotao Yin, A feasible method for optimization with orthogonality constraints, Math. Program., 142 (2013), pp. 397–434.

11. Zaiwen Wen and Yin Zhang, Accelerating Convergence by Augmented Rayleigh–Ritz Projections For Large-Scale Eigenpair Computation, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 273–296.

12. Honglin Yuan, Xiaoyi Gu, Rongjie Lai, and Zaiwen Wen, Global optimization with orthogonality constraints via stochastic diffusion, tech. report, Peking University, 2017.

ADMM

Xiaoming Yuan
Hong Kong University

Deren Han
Beihang University and Nanjing Normal University

The alternating direction method of multipliers (ADMM) has been well studied, particularly in the past decade, mainly because of its easy implementability, competitive numerical performance and applicability in a wide spectrum of areas. Its pros and cons have been well noticed in the literature; and it has been significantly extended in various contexts such as multi-block-function convex optimization problems, big-data scenarios, matrix optimization problems, control problems, and nonconvex optimization problems. It is good timing to review its state-of-the-art, to open discussions on unsolved problems in this topic and to foresee its future.

Non-problems in Optimization for Statistics

Jongshi Pang
University of Southern California

The phrase "non-problems" in the title refers to a collection of adjectives that start with "non"; these include "non-convex", "non-differentiable", "non-traditional", "non-trivial", and "non-stochastic gradient" (as a counter to a topical research theme), all in the context of optimization for statistics. Outside a stand-alone optimization problem, the phrase could include "non-cooperative" game theory as this is an area where there are significant research opportunities when combined with the other "non-subjects". I will present a variety of these non-problems, give a brief summary of research on them, and suggest some opportunities to expand these non-subjects.

Much of our research is performed jointly with Dr. Ying Cui at the University of Southern California (USC). Our study on the statistics problems is in collaboration with Bodhisattva Sen at Columbia University and Yufeng Liu (and his student Zhengling Qi) at the University of North Carolina. Our work on stochastic games is joint with Suvrajeet Sen and his student Junyi Liu at USC and Uday Shanbhag and his postdoc Jinlong Lai at Penn State.

Advances in globalization techniques

Serge Gratton
CERFACS

Globalization techniques for nonconvex optimization have undergone a series of improvement in recent years. In 2008 [5] a complexity theory was proposed for trust-region algorithms solving unconstrained problems. The theory has been extended in many directions [4]. It has been generalized to the case of constraints. It has stimulated the development of algorithms with good complexity such as the adaptive cubic regularization [2]. It has also been extended to optimization without derivatives [6,9], which has motivated the study of globalization techniques based on probabilistic information [7,8,3] following an initial work on trust region [1]. The advent of large-scale nonconvex possibly nonsmooth problems, including those arising from machine learning, calls for further exploration of the globalization techniques. Parallelization, randomization, and exploitation of inaccurate information will be instrumental to tackle the new challenges. This session will address recent advances in globalization techniques and discuss perspectives for further investigations.

References

1. A. S. Bandeira, K. Scheinberg, and L. N. Vicente, Convergence of trust-region methods based on probabilistic models, SIAM Journal on Optimization, 24 (2014) 1238-1264 ion methods based on probabilistic models, SIAM Journal on Optimization, 24 (2014) 1238-1264

2. C. Cartis, N. Gould, and Ph. L. Toint, Adaptive cubic overestimation methods for unconstrained optimization. Part I: motivation, convergence and numerical results, Mathematical Programming 127 (2011), 245-295

3. C. Cartis, K. Scheinberg, Global convergence rate analysis of unconstrained optimization methods based on probabilistic models, Mathematical Programming, 169 (2018), 337-375

4. G. N. Grapiglia, J. Yuan, and Y. Yuan, On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization, Mathematical Programming 152 (2014), 491-520

5. S. Gratton, A. Sartenaer, and Ph. L. Toint, Recursive Trust-Region Methods for Multiscale Nonlinear Optimization, SIAM Journal on Optimization, 19 (2008) 414-444

6. S. Gratton, C. W. Royer, and L. N. Vicente, A second-order globally convergent direct-search method and its worst-case complexity, Optimization, 65 (2016) 1105-1128

7. S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang, Complexity and global rates of trust-region methods based on probabilistic models, IMA Journal of Numerical Analysis, 38 (2018) 1579-1597

8. S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang, Direct search based on probabilistic descent, SIAM Journal on Optimization, 25 (2015) 1515-1541

9. L. N. Vicente, Worst case complexity of direct search, EURO Journal on Computational Optimization, 1 (2013) 143-153

Theory and algorithms for two-stage stochastic variational inequalities

Xiaojun Chen
The Hong Kong Polytechnic University

The two-stage stochastic variational inequality (SVI) provides a powerful modeling paradigm for many important applications in which uncertainties and equilibrium are present. The two-stage SVI is to find a pair: here-and-now solution and wait-and-see solution. The here-and-now solution represents now-decisions, while the wait-and-see solution depends on future events described by random variables. This talk reviews new developments in optimization theory and algorithms for two-stage SVI.

Co-authors: Ting Kei Pong, Alexander Shapiro, Defeng Sun, Hailin Sun, Roger Wets, Huifu Xu, Junfeng Yang

References

1. X. Chen and R.J-B. Wets, Stochastic Equilibrium and Variational Inequalities, a special issue of Math. Program., 165(2017).

2. X. Chen, T.K. Pong and R. J-B Wets, Two-stage stochastic variational inequalities: an ERM-solution procedure, Math. Program., 165(2017), 71-111.

3. R.T. Rockafellar and R. J.-B. Wets, Stochastic variational inequalities: single-stage to multistage, Math. Program., 165(2017), 331-360.

4. R.T. Rockafellar and J. Sun, Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging, Math. Program., (2018)

5. X. Chen, H. Sun and H. Xu, Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems, Math. Program., (2018)

6. X. Chen, A. Shapiro and H. Sun, Convergence analysis of sample average approximation of two-stage stochastic generalized equations, SIAM J. Optim. (minor revision 2018).

Traditional Nonlinear Programming

Yu-Hong Dai
Chinese Academy of Sciences

Hongchao Zhang
Louisiana State University

In this talk we overview some development of numerical methods for nonlinear programming. In particular, we review unconstrained optimization, constrained optimization, derivative-free optimiation and global optimization. New approximation models, algorithmic complexity, ..., attract more and more attentions of the community. We also discuss some future research in nonlinear programming.

Tensor and Polynomial Optimization

Shuzhong Zhang

Institute of Data and Decision Analytics, The Chinese University of Hong Kong (Shenzhen) and Department of Industrial and Systems Engineering, University of Minnesota

In this talk we shall survey approximation results for optimization models where the objective and constraints are tensor and/or polynomial functions. In particular, we shall discuss approximation algorithms for nonconvex quadratic optimization. Recent developments regarding the cones of nonnegative (high-order) polynomials will be introduced as well.

Part I: Nonconvex Quadratic Optimization

Co-authors:

- Simai He
- Tom Luo
- Jiawang Nie
- Nikos Sidiroupolous
- Paul Tseng

References

1. Z.Q. Luo, N.D. Sidiropoulos, P. Tseng, and S. Zhang, Approximation Bounds for Quadratic Optimization with Homogeneous Quadratic Constraints, SIAM Journal on Optimization, 18, 1 28, 2007.

2. S. He, Z.Q. Luo, J. Nie, S. Zhang, Semidefnite Relaxation Bounds for Indefinite Homogeneous Quadratic Optimization, SIAM Journal on Optimization, 19, 503 523, 2008.

Part II: Tensor and Polynomials Optimization

Co-authors:

- Simai He
- Bo Jiang
- Zhening Li

References

1. S. He, Z. Li, and S. Zhang, Approximation Algorithms for Homogeneous Polynomial Optimization with Quadratic Constraints, Mathematical Programming, 125, 353 383, 2010.
2. S. He, B. Jiang, Z. Li, and S. Zhang, Probability Bounds for Polynomial Functions in Random Variables, Mathematics of Operations Research, 39 (3), 889 907, 2014.
3. Z. Li, S. He and S. Zhang, Approximation Methods for Polynomial Optimization, Springer, 2012.

Part III: Cones of Nonnegative Polynomials
Co-authors:

- Bilian Chen
- Simai He
- Bo Jiang
- Zhening Li

References

1. B. Jiang, Z. Li, and S. Zhang, On Cones of Nonnegative Quartic Forms, Foundations of Computational Mathematics, 17, 161 197, 2017.
2. B. Chen, S. He, Z. Li, and S. Zhang, On New Classes of Nonnegative Symmetric Tensors, SIAM Journal on Optimization, 27 (1), 292 318, 2017.

Some Recent Development of Deep Modeling in Medical Imaging

Bin Dong
Peking University

Zhihua Zhang
Peking University

In this talk, we will review some recent work on deep learning in medical imaging. we will mostly focus on image reconstructions such as CT and MRI. We will also mention some other applications of deep learning in image segmentation and synthesis. In the end, we will present the remaining challenges of deep learning in medical imaging from my own perspective.

Recent Advances on Optimization for Machine Learning

Zhouchen Lin
Peking University

Shiqian Ma
UC Davis

This talk will discuss recent advances on optimization for machine learning. We will cover stochastic and distributed algorithms for both convex and nonconvex machine learning problems, and theory of nonconvex optimization problems arising in deep learning.

Sparse Semismooth Newton Methods and Big Data Composite Optimization

Defeng Sun
The Hong Kong Polytechnic University

Big data optimization problems provide many challenges as well as plenty of opportunities for algorithm developers. Concerned with inherent huge computational burdens of the interior point methods (IPMs) for solving optimization problems of large scales, many researchers and practitioners tend to the first order methods (FOMs) such as the accelerated proximal gradient methods and the alternating direction methods of multipliers for the rescue. While these FOMs have indeed been very successful in a number of interesting applications, they also encounter enormous numerical difficulties in dealing with many real data optimization problems of big scales even with a low or moderate solution quality. New ideas for solving these problems are highly sought both in practice and in academic research. In this talk, we shall explain why sparse semismooth Newton methods can fully exploit the second order sparsity/low rank information exhibited in big composite optimization models. Consequently, scalable and efficient algorithms can often be designed to overcome the mentioned numerical difficulties either in IPMs or in FOMs. A highly efficient software called LassoNAL for solving the well-known Lasso problem will be used to demonstrate that solving big data optimization models even with higher accuracy are becoming realistic.

Optimization and Signal Processing

Zhiquan Luo

University of Minnesota and The Chinese University of Hong Kong, Shenzhen

Mingyi Hong

University of Minnesota

In this session, we will provide an overview of optimization challenges that arise in a number of new research areas in signal/information processing and wireless communication. In the first part of this session, we will present research problems mainly from two areas: 1) resource allocation for wireless communication, and 2) distributed information processing. Specific topics to be covered include:

- Efficient network virtualization for multiple on-demand services over future communication networks;
- Optimal coordination of unmanned Ariel Vehicles (UAVs) for efficient wireless access;
- Enhancing wireless resource allocation by combining modern machine learning and optimization methods;
- Efficient distributed optimization for emerging signal/information processing problems.

In the second part of this session, we will invite the audience to have an open discussion on those problems and beyond.

Algorithms for SDP and Conic Solvers

Kim-Chuan Toh
National University of Singapore

In the first part, we present an overview of primal-dual path-following interior-point solvers and discuss the numerical challenges in the implementations. In the second part, we present an overview of augmented Lagrangian based solvers (in particular semismooth Newton-CG ALM) for large-scale SDP problems with polyhedral bound constraints. We also discuss the design of ALM solvers for more general conic programming such as QSDP.

The talk is based on joint work with Defeng Sun, Xudong Li, Liuqin Yang, Xinyuan Zhao, Micheal Todd, Reha Tutuncu.

Poster Session

Exact Separation for Unsplittable Network Design Problem

Liang Chen
Chinese Academy of Sciences

The multicommodity capacitated network design problem with unsplittable flow arc set with multi-facility is an $\mathcal{NP}$-hard problem. In this work, we propose an exact cutting plane algorithm. Its core is an exact separation for the unsplittable flow arc set with multi-facility including an initialization of partial LP and a normalization of coefficients in row generation. We prove our exact separation cuts are facet-defining inequality for the set in reduced dimension space. Furthermore, a series of subproblems in sequential lifting are solved by a pseudo-polynomial complexity algorithm modified from BOUKNAP that solves bounded knapsack problem. Numerical experiments show much efficiency of the exact separation procedure.

Parallelizable Algorithms for Optimization Problems with Orthogonality Constraints

Bin Gao
Chinese Academy of Sciences

To construct a parallel approach for solving optimization problems with orthogonality constraints is usually regarded as an extremely difficult mission, due to the low scalability of the orthogonalization procedure. However, such demand is particularly huge in some application domains such as material computation. In this paper, we propose two infeasible algorithms, based on augmented Lagrangian penalty function, for solving optimization problems with orthogonality constraints. Unlike the classic augmented Lagrangian method, in our algorithms, the prime variables are updated by a gradient step or a column-wise normalized gradient step, meanwhile the dual variables are updated by a closed-form expression which holds at any first-order stationary point. The orthogonalization procedure is only invoked once as the last step of the above mentioned two algorithms. Consequently, the main parts of the proposed algorithms can be parallelized naturally. We establish global subsequence convergence results for our proposed algorithms. Worst-case complexity and local convergence rate are also studied under some mild assumptions. Numerical experiments, including tests under parallel environment, illustrate that our new algorithms attain good performances and a high scalability in solving discretized Kohn-Sham total energy minimization problems.

Structured Quasi-Newton Methods for Optimization with Orthogonality Constraints

Jiang Hu
Peking University

In this paper, we study quasi-Newton methods for optimization problems with orthogonality constraints. Note that the Riemannian Hessian of the objective function is a composition of the Euclidean Hessian and some operations on the Euclidean gradient. Due to the computational costs of two parts, we construct an approximation to the Euclidean Hessian while keep the part related to Euclidean gradient. More specifically, our idea is to keep these parts of lower computational costs but substitute those parts of higher computational costs by the limited-memory quasi-Newton update. The initial quasi-Newton matrix is further constructed from a limited-memory Nyström approximation to the expensive parts. Consequently, our subproblems approximate the original objective function in the Euclidean space and preserve the orthogonality constraints without performing the so-called vector transports. When the subproblems are solved to certain accuracy, both global and local q-superlinear convergence can be established under certain mild conditions. Preliminary numerical experiments on the linear eigenvalue problem and the electronic structure calculation show the effectiveness of our method compared with the state-of-art algorithms.

Vector Transport-Free SVRG with General Retraction for Riemannian Optimization: Complexity Analysis and Practical Implementation

Bo Jiang (Nanjing)
Nanjing Normal University

In this paper, we propose a vector transport-free stochastic variance reduced gradient (SVRG) method with general retraction for empirical risk minimization over Riemannian manifold. Existing SVRG methods on manifold usually consider a specific retraction operation, and involve additional computational costs such as parallel transport or vector transport. The vector transport-free SVRG with general retraction we propose in this paper handles general retraction operations, and do not need additional computational costs mentioned above. As a result, we name our algorithm S-SVRG, where the first "S" means simple. We analyze the iteration complexity of S-SVRG for obtaining an ϵ-stationary point and its local linear convergence by assuming the Łojasiewicz inequality, which naturally holds for PCA and holds with high probability for matrix completion problem. We also incorporate the Barzilai-Borwein step size and design a very practical S-SVRG-BB method. Numerical results on PCA and matrix completion problems are reported to demonstrate the efficiency of our methods.

On Adaptive Cubic Regularized Newton's Methods for Convex Optimization via Random Sampling

Bo Jiang (SHUFE)
Shanghai University of Finance and Economics

In this paper, we consider an unconstrained optimization model where the objective is a sum of a large number of possibly nonconvex functions, though overall the objective is assumed to be smooth and convex. Our bid to solving such model uses the framework of cubic regularization of Newton's method. As well known, the crux in cubic regularization is its utilization of the Hessian information, which may be computationally expensive for large-scale problems. To tackle this, we resort to approximating the Hessian matrix via sub-sampling. In particular, we propose to compute an approximated Hessian matrix by either uniformly or non-uniformly sub-sampling the components of the objective. Based upon such sampling strategy, we develop both standard and accelerated adaptive cubic regularization approaches and provide theoretical guarantees on global iteration complexity. We show that the standard and accelerated sub-sampled cubic regularization methods achieve iteration complexity in the order of $O(\epsilon^{-1/2})$ and $O(\epsilon^{-1/3})$ respectively, which match those of the original standard and accelerated cubic regularization methods using the full Hessian information. The performances of the proposed methods on regularized logistic regression problems show a clear effect of acceleration in terms of epochs on several real data sets.

Coauthor: Xi Chen, Tianyi Lin, Shuzhong Zhang.

A Semi-smooth Newton Method for Solving Semidefinite Programs in Electronic Structure Calculations

Yongfeng Li
Peking University

The well-known interior point method for semidefinite progams (SDPs) can only be used to tackle problems of relatively small scales. First-order methods such as the the alternating direction method of multipliers (ADMM) have much lower computational cost per iteration. However, their convergence can be slow, especially for obtaining highly accurate approximations. In this paper, we present a practical and efficient second-order semi-smooth Newton type method based on solving a fixed-point mapping derived from an equivalent form of the ADMM. We discuss a number of techniques that can be used to improve the computational efficiency of the method and achieve global convergence. Then we further consider the application in electronic structure calculations. The ground state energy of a many-electron system can be approximated by an variational approach in which the total energy of the system is minimized with respect to one and two-body reduced density matrices (RDM) instead of many-electron wavefunctions. This problem can be formulated as a semidefinite programming problem. Extensive numerical experiments show that our approach is competitive to the state-of-the-art methods in terms of both accuracy and speed.

A Sparse Completely Positive Relaxation of the Modularity Maximization for Community Detection

Haoyang Liu
Peking University

Consider the community detection problem under either the stochastic block model (SBM) assumption or the degree-correlated stochastic block model (DCSBM) assumption. The modularity maximization formulation for the community detection problem is NP-hard in general. We propose a sparse and low-rank completely positive relaxation for the modularity maximization problem, we then develop an efficient row-by-row (RBR) type block coordinate descent (BCD) algorithm to solve the relaxation and prove an $\mathcal{O}(1/\sqrt{N})$ convergence rate to a stationary point where N is the number of iterations. A fast rounding scheme is constructed to retrieve the community structure from a solution to the above relaxation. Non-asymptotic high probability bounds on the misclassification rate are established to justify our approach. We further develop an asynchronous parallel RBR algorithm to speed up the convergence. Extensive numerical experiments on both synthetic and real world networks show that the proposed approach enjoys advantages in both clustering accuracy and numerical efficiency. Our numerical results indicate that the newly proposed method is a quite competitive alternative for the community detection problem on sparse networks with over 50 million nodes.

Tightness of an enhanced semidefinite relaxation for MIMO detection

Cheng Lu
North China Electric Power University

We consider a fundamental problem in modern digital communications known as multiple-input multiple-output (MIMO) detection, which can be formulated as a complex quadratic programming problem subject to unit-modulus and discrete argument constraints. Various semidefinite relaxation (SDR) based algorithms have been proposed to solve the problem in the literature. In this paper, we first show that the conventional SDR is generically not tight for the problem. Then, we propose a new and enhanced SDR and show its tightness under an easily checkable condition, which essentially requires the level of the noise to be below a certain threshold. The above results have answered an open question posed by A.-M.C. So. Numerical simulation results show that our proposed SDR significantly outperforms the conventional SDR in terms of the relaxation gap.

A Stochastic Semismooth Newton Method for Nonconvex Nonsmooth Optimization

Andre Milzarek
Peking University

We present a globalized stochastic semismooth Newton method for solving optimization problems involving smooth nonconvex and nonsmooth convex terms in the objective function. The class of problems that can be solved within our algorithmic framework comprises a large variety of important applications such as expected and empirical risk minimization, structured dictionary learning, and other problems arising in machine learning, statistics, or image processing. We assume that only noisy gradient and Hessian information of the smooth part of the objective function is available via calling stochastic first- and second-order oracles. Our approach utilizes approximate second order information and stochastic semismooth Newton steps for a prox-type fixed-point equation to accelerate the basic stochastic proximal gradient method for convex composite programming. Inexact growth conditions are incorporated to monitor the quality and acceptance of the Newton steps. We prove that the proposed algorithm converges globally to stationary points in expectation and almost surely. Moreover, under standard assumptions, the method can be shown to locally turn into a pure semismooth Newton method and fast local convergence can be established with high probability. Finally, we provide numerical experiments demonstrating the efficiency of the stochastic semismooth Newton method.

Augmented Lagrangian Methods of Multipliers for VLSI Global Placement

Zheng Peng
Fuzhou University

We proposed several augmented Lagrangian methods of multipliers for solving the large scale constrained nonlinear programming raised in the flied of the very large-scale integration (VLSI) placement problem, and established some convergence properties of the proposed methods. Experimental results on the International Symposium on Physical Design (ISPD) benchmarks, compared with some state-of-the-art methods, show that the global placement methods resulted by the proposed methods are very efficient.

A Branch-and-Cut Algorithm for Mixed Integer Bilevel Linear Optimization Problems and Its Implementation

Ted Ralphs
Lehigh University

We describe a comprehensive algorithmic framework for solving mixed integer bilevel linear optimization problems (MIBLPs) using a generalized branch-and-cut approach. The framework presented merges features from existing algorithms (for both traditional mixed integer linear optimization and MIBLPs) with new techniques to produce a flexible and robust framework capable of solving a wide range of bilevel optimization problems. The framework has been fully implemented in the open-source solver MibS. The paper describes the algorithmic options offered by MibS and presents computational results evaluating the effectiveness of the various options for the solution of a number of classes of bilevel optimization problems from the literature.

New stepsizes for the gradient method

Cong Sun

Beijing University of Posts and Telecommunications

Gradient method is famous for its simplicity and low complexity, which attracts more and more attentions for large scale optimization problems. A good stepsize plays an important role to construct an efficient gradient method. This paper proposes a new framework to generate stepsizes for gradient methods. By adopting different criterions, we propose four new gradient methods. For 2-dimensional unconstrained problems with convex quadratic objective functions, we prove that the new methods either terminate in finite iterations or converge R-superlinearly; for n-dimensional problems, we prove that all the new methods converge R-linearly. Numerical experiments show that the new methods enjoy lower complexity and outperform the existing gradient methods.

Convergence Analysis of Sample Average Approximation of Two-stage Stochastic Generalized Equations

Hailin Sun
Nanjing University of Science and Technology

A solution of two-stage stochastic generalized equations is a pair: a first stage solution which is independent of realization of the random data and a second stage solution which is a function of random variables. This paper studies convergence of the sample average approximation of two-stage stochastic nonlinear generalized equations. In particular, an exponential rate of the convergence is shown by using the perturbed partial linearization of functions. Moreover, sufficient conditions for the existence, uniqueness, continuity, and regularity of solutions of two-stage stochastic generalized equations are presented under an assumption of monotonicity of the involved functions. These theoretical results are given without assuming relatively complete recourse and are illustrated by two-stage stochastic non-cooperative games of two players.

Stochastic Trust Region Methods with Trust Region Radius Depended on Probabilistic Models

Xiaoyu Wang
Chinese Academy of Sciences

In this work, we present a general stochastic trust region scheme in which the trust region radius formula is directly associated with the random model. The proposed scheme is analyzed based on random models and estimates with certain probability as long as some assumptions are satisfied. We also show a concrete algorithm STRME in which the trust region radius is linearly correlated with the 2-norm of the stochastic gradient. Moreover, the convergence and complexity results of STRME method in nonconvex, convex and strongly convex settings have all been analyzed. In the end, we present some numerical experiments to reveal the benefits of the proposed STRME methods compared to the STORM and other related approaches based on stochastic gradients.

Spectral Compressed Sensing via Projected Gradient Descent

Ke Wei
Fudan University

Let x be a spectrally sparse signal consisting of r complex sinusoids with or without damping. We consider the spectral compressed sensing problem, which is about reconstructing x from its partial revealed entries. By utilizing the low rank structure of the Hankel matrix corresponding to x, we develop a computationally efficient algorithm for this problem. The algorithm starts from an initial guess computed via one-step hard thresholding followed by projection, and then proceeds by applying projected gradient descent iterations to a non-convex functional. Based on the sampling with replacement model, we prove that $O(r^2 \log n)$ observed entries are sufficient for our algorithm to achieve the successful recovery of a spectrally sparse signal. Moreover, extensive empirical performance comparisons show that our algorithm is competitive with other state-of-the-art spectral compressed sensing algorithms in terms of phase transitions and overall computational time.

An Incremental Gauss-Newton Method for l_2 Regularized Finite Sum Minimization

Yuchen Wu
Chinese Academy of Sciences

Machine learning is a hot topic in which optimization is a fundamental tool. We consider large scale finite sum problems in machine learning, where the problem size is large, the problem dimension is not very large and the residue is very small. We take advantage of nonnegativity of objective function and transform the original problem into a nonlinear least square problem. We propose a framework modified incremental Gauss-Newton(mIGN) for solving this problem. We also propose a minibatch version of mIGN, so we can reduce the number of matrix updation. Also, our method can be parallelized. Then, we prove the local convergence rate and global convergence of mIGN under mild assumptions. Also, our theoretical demonstration indicates that it will converge fast when residue is small. Finally, our numerical experiments show that our method is powerful in solving finite sum problems. We also show that we can use less time if we use the minibatch mIGN.

Approximating the weighted maximin dispersion problem over an Lp-ball: SDP relaxation is misleading

Yong Xia
Beihang University

Consider the problem of finding a point in a unit n-dimensional Lp-ball ($p \geq 2$) such that the minimum of the weighted Euclidean distance from given m points is maximized. We show in this paper that the recent SDP-relaxation-based approximation algorithm (Haines et al., SIAM J Optim 23(4):22642294, 2013) will not only provide the first theoretical approximation bound of $(1 - O(\sqrt{\ln(m)/n}))/2$, but also perform much better in practice, if the SDP relaxation is removed and the optimal solution of the SDP relaxation is replaced by a simple scalar matrix.

Global Convergence Guaranteed Approach for Eigenspace Calculation

Nachuan Xiao
Chinese Academy of Sciences

Eigenspace calculation can be modeled as Rayleigh-Ritz minimization with orthogonality constraints. When the dimension of required eigenspace increases, the traditional feasible algorithms lack of parallel scalability. [Eigpen] [SLRP] introduced two useful unconstrained models for eigenspace calculation, and provided a gradient algorithm and a Gauss-Newtom algorithm, respectively, which enjoy good numerical performance including efficiency and scalability. However, the authors of these papers can only prove the global convergence of these two algorithms to first-order stationary points, and then global convergence to global optimizer of such algorithms is lack. In this paper, we put forward a new unconstrained model and propose an alternating minimization algorithm (AMA) to solve this model. Since each step of AMA is nothing but a shifted power method, we can prove its global convergence to a global optimizer with high probability. Preliminary numerical results also stand for the efficiency of the new proposed algorithms.

Faster, but not more, iterations. Accelerated Asynchronous Parallel Algorithm: A2BCD

Wotao Yin
University of California, Los Angeles

A2BCD is the first and currently fastest asynchronous algorithm because it achieves an optimal rate for minimizing differentiable and strongly convex functions. A2BCD "asynchronizes" accelerated non-uniform BCD, or NU_ACDM, to eliminate CPU idle time, reduce communication bottleneck, and tolerate stragglers. Although the computation in A2BCD may use delayed information, it retains the accelerated convergence rate of NU_ACDM, which can be sqrt(n) times better than Nesterov's accelerated random BCD. We also modify Lan and Zhou's example to show A2BCD has the optimal rate. Most async algorithms are currently analyzed by treating delays as noise, but we don't. This is the key.

Co-author: Robert Hannah.

Theoretical Linear Convergence of Unfolded ISTA and Its Practical Weights and Thresholds

Wotao Yin
University of California, Los Angeles

Unfolding iterative algorithms as neural networks has been shown an empirical success in solving sparse recovery problems. However, its theoretical understanding is still immature, which prevents us from fully utilizing the power of neural networks. In this work, we study unfolded ISTA for sparse signal recovery. We introduce a weight structure that is necessary for asymptotic convergence to the true sparse signal. With this structure, unfolded ISTA can attain a linear convergence, which is better than the sublinear convergence of ISTA/FISTA in general cases. Furthermore, we propose to incorporate thresholding in the network to perform support selection, which is easy to implement and able to boost the convergence rate both theoretically and empirically. Extensive simulations, including sparse vector recovery and a compressive sensing experiment on real image data, corroborate our theoretical results and demonstrate their practical usefulness.

Co-authors: Xiaohan Chen, Jialin Liu, Zhangyang Wang

Global Complexity Analysis of Inexact Successive Quadratic Approximation method for Regularized Optimization under Mild Assumptions

Hui Zhang
National University of Defense Technology

The successive quadratic approximation (SQA) method for regularized optimization(also called proximal quasi Newton method, or variable metric forward-backward splitting method) is numerically efficient for minimizing the sum of a smooth function and a convex function. In this paper, we try to investigate the global complexity of inexact SQA with line searches under mild assumptions. First, we show that the algorithm is well defined in the real Hilbert space even without the Lipschitz assumption. Second, under the quadratic growth condition, we prove that the objective value sequence linearly converges to the minimum; with local Lipschitz assumption, the iterate sequence is also linearly convergent. Finally, for convex cases, it converges sublinearly in terms of the function values; if the iterates converge to the optimal set, then the convergence rate could be improved from $O(1/k)$ to $o(1/k)$.

List of Participants

- Wei Bian, Harbin Institute of Technology
- Xiaojun Chen, The Hong Kong Polytechnic University
- Caihua Chen, Nanjing University
- Yuhong Dai, Chinese Academy of Sciences
- Chao Ding, Chinese Academy of Sciences
- Bin Dong, Peking University
- Jinyan Fan, Shanghai Jiaotong University
- Serge Gratton, CERFACS
- Deren Han, Beihang University/Nanjing Normal University
- Xin Liu, Chinese Academy of Sciences
- Yafeng Liu, Chinese Academy of Sciences
- Mingyi Hong, University of Minnesota
- Bo Jiang, Nanjing Normal University
- Bo Jiang, Shanghai University of Finance and Economics
- Thorsten Koch, Zuse Institute Berlin, Technische Universitäet Berlin
- Jon Lee, University of Michigan
- Zhouchen Lin, Peking University
- Cheng Lu, North China Electric Power University
- Zhaosong Lu, Simon Fraser University
- Zhiquan Luo, University of Minnesota/The Chinese University of Hong Kong,Shenzhen
- Andre Milzarek, Peking University
- Shiqian Ma, UC Davis
- Lingfeng Niu, University of Chinese Academy of Sciences
- Jongshi Pang, University of Southern California
- Zheng Peng, Fuzhou University
- Ted Ralphs, Lehigh university
- Cong Sun, Beijing University of Posts and Telecommunications
- Jie Sun, Curtin University/National University of Singapore
- Defeng Sun, The Hong Kong Polytechnic University

- Hailin Sun, Nanjing University of Science and Technology
- Anthony So, The Chinese University of Hong Kong
- Kim Chuan Toh, National University of Singapore
- Ke Wei, Fudan University
- Yong Xia, Beihang University
- Shuhuang Xiang, Central South University
- Zi Xu, Shanghai University
- Ming Yan, Michigan State University
- Junfeng Yang, Nanjing University
- Xinming Yang, Chongqing Normal University
- Yinyu Ye, Stanford University
- Wotao Yin, University of California, Los Angeles
- Yaxiang Yuan, Chinese Academy of Sciences
- Xiaoming Yuan, Hong Kong University
- Zaikun Zhang, The Hong Kong Polytechnic University
- Zhihua Zhang, Peking University
- Hui Zhang, National University of Defense Technology
- Hongchao Zhang, Louisiana State University
- Yin Zhang, The Chinese University of Hong Kong, Shenzhen
- Zhenyue Zhang, Zhejiang University
- Shuzhong Zhang, University of Minnesota

The organizers wishes you a pleasant stay in Wuyishan!