2026 TYMRC Workshop on
Large Language Models and Optimization



天元數學國際交流中心

Tianyuan Mathematics Research Center

Feb 1-7, 2026

KUNMING, CHINA

http://tianyuan.amss.ac.cn/ztyt/info/2025/146120.html

Contents

Organizing Committee

Yurii Nesterov, Corvinus University of Budapest

Zaiwen Wen, Peking University

Kun Yuan, Peking University

Contact Information

If you need any help, please feel free to contact:

Kun Yuan, Peking University: kunyuan@pku.edu.cn, 13140810278

Organization Of the Program

The program is divided into eight sessions, with each session comprising:

- Two 40-minute talks (approximately 30 minutes for the presentation, followed by time for questions).
- A half-hour break.
- Two 40-minute talks, or one 40-minute talk plus two 20-minute student talks.

Agenda

	Feb 2, Monday	Feb 3, Tuesday	Feb 4, Wednesday	Feb 5, Thursday	Feb 6, Friday
08:30-09:50	Chair: Deren Han Talk: Anthony Man-Cho So Talk: Zizhuo Wang	Chair: Anthony Man-Cho So Talk: Deren Han Talk: Liwei Zhang	Chair: Minru Bai Talk: Qing Ling Talk: Songtao Lu	Chair: Qing Ling Talk: Yuhong Dai Talk: Yancheng Yuan	Chair: Hailin Sun Talk: Kuang Bai Talk: Haishan Ye
09:50-10:20	Break				
10:20-11:40	Chair: Zizhuo Wang Talk: Kun Yuan Talk: Xiangfeng Wang	Chair: Cong Sun Talk: Minru Bai Talk: Shixiang Chen	Chair: Liping Wang Talk: Hailin Sun Talk: Ziyan Luo	Chair: Ziyuan Luo Talk: Shenglong Hu Talk: Liang Chen	Chair: Kun Yuan Talk: Siqi Zhang Talk: Yang Liu
11:40-14:00	Lunch				
14:00-15:20	Chair: Liang Chen Talk: Cong Sun Talk: Rujun Jiang	Chair: Jianlin Jiang Talk: Bo Jiang (SUFE) Talk: Zi Xu	Free discussion 1	Chair: Bo Jiang (NJNU) Talk: Ruoyu Sun Talk: Xiao Li	Free discussion 2
15:20-15:50	Break			Break	
15:50-17:10	Chair: Jiaxin Xie Talk: Wen Huang Talk: Jiang Hu	Chair: Rujun Jiang Talk: Jiaojiao Zhang Student Talk: Yutong He, Benqi Liu		Chair: Jiaojiao Zhang Talk: Qi Deng Student Talk: Qiming Dai, Mingyu Mo	
17:10-19:00	Dinner			Dinner	
20:00-21:00	Free group discussion 1	Free group discussion 2		Free group discussion 3	

Workshop Schedule

Date	Time	Workshop Information
Feb 2, Monday	08:30-09:50	Chair: Deren Han
		Talk: Anthony Man-Cho So On Efficiently Computable Approximate Stationarity Concepts in Bilevel Optimization
		Talk: Zizhuo Wang Large Language Models for Optimization Modeling
	09:50-10:20	Break
	10:20-11:40	Chair: Zizhuo Wang
		Talk: Kun Yuan Memory-Efficient Optimization for Training Large Language Models
		Talk: Xiangfeng Wang Verifiably Improving Optimization Algorithms via LLMs
	11:40-14:00	Lunch
	14:00-15:20	Chair: Liang Chen
		Talk: Cong Sun Cyclic stochastic gradient method
		Talk: Rujun Jiang Loss Landscape and Error Bound Analysis of Regularized Deep Matrix Factorization
	15:20-15:50	Break
	15:30-17:10	Chair: Jiaxin Xie
		Talk: Wen Huang TBA
		Talk: Jiang Hu Retraction-free optimization over the Stiefel manifold with application to the LoRA fine-tuning
	17:10-19:00	Dinner
	20:00-21:00	Free group discussion
Feb 3, Tuesday	08:30-09:50	Chair: Anthony Man-Cho So
		Talk: Deren Han

<table>
<tr><td rowspan="19"></td><td rowspan="2"></td><td>TBA</td></tr>
<tr><td>Talk: Liwei Zhang
TBA</td></tr>
<tr><td>09:50-10:20</td><td>Break</td></tr>
<tr><td rowspan="3">10:20-11:40</td><td>Chair: Cong Sun</td></tr>
<tr><td>Talk: Minru Bai
Group Sparse-Based Tensor CP Decomposition: Model, Algorithm, and Applications</td></tr>
<tr><td>Talk: Shixiang Chen
Alternating-Projections-Type Retractions</td></tr>
<tr><td>11:40-14:00</td><td>Lunch</td></tr>
<tr><td rowspan="3">14:00-15:20</td><td>Chair: Jianlin Jiang</td></tr>
<tr><td>Talk: Bo Jiang (SUFE)
New Results on the Polyak Stepsize: Tight Convergence Analysis and Universal Function Classes</td></tr>
<tr><td>Talk: Zi Xu
Gradient Norm Regularization Second-Order Algorithms for Solving Nonconvex-Strongly Concave Minimax Problems</td></tr>
<tr><td>15:20-15:50</td><td>Break</td></tr>
<tr><td rowspan="4">15:50-17:10</td><td>Chair: Rujun Jiang</td></tr>
<tr><td>Talk: Jiaojiao Zhang
Locally Differentially Private Online Federated Learning with Correlated Noise</td></tr>
<tr><td>Student Talk 1: Yutong He
Subspace Optimization for Large Language Models with Convergence Guarantees</td></tr>
<tr><td>Student Talk 2: Benqi Liu
TBA</td></tr>
<tr><td>17:10-19:00</td><td>Dinner</td></tr>
<tr><td>20:00-21:00</td><td>Free group discussion</td></tr>
<tr><td rowspan="3">Feb 4,
Wednesday</td><td rowspan="3">08:30-09:50</td><td>Chair: Minru Bai</td></tr>
<tr><td>Talk: Qing Ling
TBA</td></tr>
<tr><td>Talk: Songtao Lu
Bilevel Optimization with Game-Theoretic Lower-Level Problems for Model Training</td></tr>
</table>

	09:50-10:20	Break
	10:20-11:40	Chair: Liping Wang
		Talk: Hailin Sun Solutions of Two-stage Stochastic Minimax Problems
		Talk: Ziyan Luo TBA
	11:40-14:00	Lunch
	14:00-20:30	Free discussion
Feb 5, Thursday	08:30-09:50	Chair: Qing Ling
		Talk: Yuhong Dai TBA
		Talk: Yancheng Yuan A Perturbed DCA for Computing d-Stationary Points of Nonsmooth DC Programs
	09:50-10:20	Break
	10:20-11:40	Chair: Ziyan Luo
		Talk: Shenglong Hu TBA
		Talk: Liang Chen An efficient second-order cone programming approach for dynamic optimal transport on staggered grid discretization
	11:40-14:00	Lunch
	14:00-15:20	Chair: Bo Jiang (NJNU)
		Talk: Ruoyu Sun TBA
		Talk: Xiao Li Optimizers in Post-Training and Pre-Training: An Introduction
	15:20-15:50	Break
	15:50-17:10	Chair: Jiaojiao Zhang
		Talk: Qi Deng On some universal methods for convex optimization
		Student Talk 1: Qiming Dai TBA Student Talk 2: Mingyu Mo TBA
	17:10-19:00	Dinner

	20:00-21:00	Free group discussion
Feb 6, Friday	08:30-09:50	Chair: Hailin Sun
		Talk: Kuang Bai Directional combined approach for bilevel optimization
		Talk: Haishan Ye Explicit and Non-asymptotic Query Complexities of Rank-Based Zeroth-order Algorithms on Smooth Functions
	09:50-10:20	Break
	10:20-11:40	Chair: Kun Yuan
		Talk: Siqi Zhang Recent Progress on Nonconvex Minimax Optimization and Beyond
		Talk: Yang Liu Solver-Aware High-Order Optimization for Large-Scale Nonconvex Problems
	12:00-14:00	Lunch
	14:00-20:30	Free discussion

Titles and Abstracts

08:30-09:10, Feb. 2, Monday
Speaker: Anthony Man-Cho So
Title: On Efficiently Computable Approximate Stationarity Concepts in Bilevel Optimization

Abstract: Bilevel optimization (BO), which concerns optimal decision making in processes that involve an upper-level decision maker (the leader) and a lower-level decision maker (the follower), has attracted much interest lately due to its many applications in machine learning and signal processing. One of the current research directions is the design of efficient iterative methods with complexity guarantees for computing approximate stationary points of structured BO problems. However, existing results in this direction either make the restrictive assumption that the lower-level solution mapping is a singleton or only establish the tractability of rather weak stationarity concepts. In this talk, we introduce a new regularity property of set-valued mappings called set smoothness and show that if a BO problem has, among other things, a set-smooth lower-level solution mapping, then the task of finding an approximate Clarke stationary point of the problem is tractable. Our results significantly sharpen those in the literature and suggest several directions for further research.

09:10-09:40, Feb. 2, Monday
Speaker: Zizhuo Wang
Title: Large Language Models for Optimization Modeling

Abstract: This talk presents recent advances in the application of large language models (LLMs) to mathematical optimization modeling, with a focus on industrial and operational research contexts. We introduce ORLM (Operations Research Language Model) — the first open-source LLMs fine-tuned specifically for optimization tasks. To mitigate the scarcity of domain-specific training data, the proposed framework includes OR-Instruct, a semi-automated pipeline designed to generate instruction-style datasets across a broad range of problem types. OR-Instruct employs two core strategies: Expansion and Augmentation. Expansion leverages advanced LLMs to generate new problem scenarios and question types from seed data, while Augmentation diversifies the dataset by modifying objectives and constraints, rephrasing problem statements, and incorporating varied modeling techniques. ORLMs trained under this framework achieved superior performance on public benchmarks such as NL4OPT and MAMO, surpassing most of the leading LLMs. We also develop the IndustryOR benchmark, encompassing real-world optimization scenarios from 13 industries, spanning five categories and three levels of complexity.

10:20-11:00, Feb. 2, Monday
Speaker: Kun Yuan
Title: Memory-Efficient Optimization for Training Large Language Models

Abstract: Recent years have seen remarkable progress in large language model research. Scaling laws suggest that continued performance gains require jointly increasing model parameters and training data. However, this scaling also dramatically increases memory demands, raising both training and inference costs. To address this challenge, we focus on the intrinsic structural properties of large models and propose several memory-efficient training methods. First, we uncover low-rank structure in gradient updates and develop a subspace projection-based training approach. Second, we analyze sparsity patterns in feed-forward networks and introduce an importance sampling strategy that significantly reduces activation memory overhead. Third, leveraging cross-layer low-rank structure, we present a parameter-efficient pretraining architecture that reduces both model and gradient memory footprint. Through systematic analysis of large model architectures, this work explores new avenues for memory optimization and provides both theoretical insights and practical solutions for efficient large-scale training.

11:00-11:40, Feb. 2, Monday
Speaker: **Xiangfeng Wang**
Title: **Verifiably Improving Optimization Algorithms via LLMs**

Abstract: Recent breakthroughs in LLM-driven mathematical discovery and evolutionary architectures—exemplified by frameworks such as FunSearch and AlphaEvolve—have demonstrated significant efficacy across various applications. Building on these advancements, this research extends similar methodologies to the LLM-aided design of optimization methods, with a specific focus on refining specialized techniques within optimization algorithms. We will discuss leveraging the code-generation capabilities of Large Language Models (LLMs) to enhance and accelerate the Alternating Direction Method of Multipliers (ADMM). To ensure the reliability of the newly discovered methods, we integrate formal verification languages such as Lean to provide rigorous provability. By employing formal methods as a constraint, we can establish a framework for trustworthy code generation, ensuring that the LLM-optimized algorithms are both efficient and mathematically verifiable.

14:00-14:40, Feb. 2, Monday
Speaker: **Cong Sun**
Title: **Cyclic stochastic gradient method**

Abstract: The cyclic stepsize update strategy is proposed for stochastic gradient method. The stepsize is updated cyclicly, where the first two stepsizes use the approximated Cauchy steps and the rest apply the fixed stepsize. The step-ahead BB stepsize is used for the Cauchy step approximation. The method combines with both monotone and nonmonotone linesearches. The convergence properties are analyzed under different types of problems, where the theoretical results are proved without the interpolation condition assumption. The numerical tests show good performances of the proposed methods compared to other first order stochastic methods.

14:40-15:20, Feb. 2, Monday
Speaker: **Rujun Jiang**
Title: **Loss Landscape and Error Bound Analysis of Regularized Deep Matrix Factorization**

Abstract: Deep matrix factorization (DMF) is a fundamental model underlying many applications, including deep linear neural networks. Despite its simplicity, the regularized DMF problem exhibits a highly nonconvex optimization landscape that is not yet fully understood. In this talk, we analyze the loss landscape and local geometry of regularized deep matrix factorization. We characterize all critical points and identify conditions a critical point is a local minimizer, a global minimizer, a strict saddle point, or a non-strict saddle point. We further establish an error bound around the critical point set, which leads to linear convergence guarantees for gradient-based methods. Our results provide theoretical insights into why first-order methods perform well for regularized DMF and offer a unified perspective on the optimization behavior of deep linear networks as an important application.

15:50-16:30, Feb. 2, Monday
Speaker: **Wen Huang**
Title: **TBA**
Abstract: TBA

16:30-17:10, Feb. 2, Monday
Speaker: **Jiang Hu**
Title: **Retraction-free optimization over the Stiefel manifold with application to the LoRA fine-tuning**
Abstract: Optimization over the Stiefel manifold plays a significant role in various machine learning tasks. Existing methods either use the retraction operators, requiring costly orthonormalization for large-scale matrices, or employ landing methods that rely on careful step size selection and penalty parameter tuning. To address these challenges, we propose a retraction-free and penalty parameter-free algorithm that directly lands on the manifold. By leveraging the strongly-convex-like property of the quadratic penalty function and the proximal smoothness of the Stiefel manifold, we establish global convergence guarantees with the best-known iteration complexities under both constant and diminishing step sizes. Then, we reformulate the low-rank adaptation (LoRA) fine-tuning problem for large language models as a manifold optimization problem, introducing Manifold-LoRA for geometry-accelerated adaptation. This approach employs the proposed landing technique and a carefully designed step size strategy to accelerate the training process. Numerical experiments on benchmark datasets demonstrate the superior efficiency and generalization performance of the proposed method.

08:30-09:10, Feb. 3, Tuesday
Speaker: **Deren Han**
Title: **TBA**
Abstract: TBA

09:10-09:50, Feb. 3, Tuesday
Speaker: **Liwei Zhang**

Title: **Large Language Models for Optimization Modeling**

Abstract: TBA

10:20-11:00, Feb. 3, Tuesday

Speaker: **Minru Bai**

Title: Group Sparse-Based Tensor CP Decomposition: Model, Algorithm, and Applications

Abstract: The CANDECOMP/PARAFAC (or canonical polyadic, CP) decomposition of tensors has numerous applications in various fields, including chemometrics, signal processing, and machine learning. Tensor CP decomposition assumes the knowledge of the exact CP rank, i.e., the total number of rank-one components of a tensor. However, accurately estimating the CP rank is very challenging. In this work, to address this issue, we prove that the CP rank can be exactly estimated by minimizing the group sparsity of any one of the factor matrices under the unit length constraints on the columns of the other factor matrices. Based on this result, we propose a CP decomposition model with group sparse regularization, which integrates rank estimation and tensor decomposition as an optimization problem, whose set of optimal solutions is proven to be nonempty. To solve this optimization problem, we propose a double-loop block-coordinate proximal gradient descent algorithm with extrapolation, and prove that each accumulation point of the sequence generated by the algorithm is a stationary point of the proposed model. Furthermore, we incorporate a rank reduction strategy into the algorithm to reduce the computational complexity. Finally, we test and evaluate the proposed model and algorithm on randomly generated tensors and real-world data from chemometrics. Numerical experiment results demonstrate the robustness and effectiveness of the proposed method.

11:00-11:40, Feb. 3, Tuesday

Speaker: **Shixiang Chen**

Title: **Alternating-Projections-Type Retractions**

Abstract: Alternating projections and their variants are a classical tool for computing a point in the intersection of two sets. For smooth manifolds, their local behavior is typically studied under transversality or weaker regularity notions such as intrinsic transversality, leading to linear or quadratic convergence results. In this work we give a geometric interpretation that unifies a broad class of (possibly inexact) alternating-projections-type algorithms. Assume that two C^2 embedded submanifolds $\mathcal{M}_1, \mathcal{M}_2 \subset \mathbb{R}^n$ intersect cleanly at $\bar{x} \in \mathcal{M}_1 \cap \mathcal{M}_2$. We show that the associated alternating mapping admits a well-defined local limit $\psi_{\bar{x}}$ on the intersection manifold $\bar{\mathcal{M}} = \mathcal{M}_1 \cap \mathcal{M}_2$, and that $\psi_{\bar{x}}$ is a retraction on $\bar{\mathcal{M}}$. If, in addition, $\mathcal{M}_1$ and $\mathcal{M}_2$ are C^3, then $\psi_{\bar{x}}$ is a second-order retraction. As a consequence, several alternating-projection-type schemes that exhibit quadratic or superlinear local behavior under transversality can be viewed as implementing second-order retractions on $\bar{\mathcal{M}}$.

14:00-14:40, Feb. 3, Tuesday

Speaker: **Bo Jiang (SUFE)**

Title: New Results on the Polyak Stepsize: Tight Convergence Analysis and Universal Function Classes

Abstract: This work revisits the classical Polyak stepsize and establishes several new results. We establish the tightness of the current convergence results and provide new convergence rates for new function classes. An interesting future direction is how to construct a universal worst-case function for general adaptive stepsizes, beyond just the Polyak stepsize.

14:40-15:20, Feb. 3, Tuesday

Speaker: Zi Xu

Title: Gradient Norm Regularization Second-Order Algorithms for Solving Nonconvex-Strongly Concave Minimax Problems

Abstract: In this talk, we study second-order algorithms for solving nonconvex-strongly concave minimax problems, which have attracted much attention in recent years in many fields, especially in machine learning. We propose a gradient norm regularized trust-region (GRTR) algorithm to solve nonconvex-strongly concave minimax problems, where the objective function of the trust-region subproblem in each iteration uses a regularized version of the Hessian matrix, and the regularization coefficient and the radius of the ball constraint are proportional to the square root of the gradient norm. The iteration complexity of the proposed GRTR algorithm to obtain an $\mathcal{O}(\epsilon,\sqrt{\epsilon})$-second-order stationary point is proved to be upper bounded by $\tilde{\mathcal{O}}(\ell^{1.5}\rho^{0.5}\mu^{-1.5}\epsilon^{-1.5})$, where μ is the strongly concave coefficient, ℓ and ρ are the Lipschitz constants of the gradient and Jacobian matrix respectively, which matches the best known iteration complexity of second-order methods for solving nonconvex-strongly concave minimax problems. We further propose a Levenberg-Marquardt algorithm with a gradient norm regularization coefficient and use the negative curvature direction to correct the iteration direction (LMNegCur), which does not need to solve the trust-region subproblem at each iteration. We also prove that the LMNegCur algorithm achieves an $\mathcal{O}(\epsilon,\sqrt{\epsilon})$-second-order stationary point within $\tilde{\mathcal{O}}(\ell^{1.5}\rho^{0.5}\mu^{-1.5}\epsilon^{-1.5})$ number of iterations. Furthermore, we propose two inexact variants of the above two algorithms, namely the IGRTR algorithm and the ILMNegCur algorithm, which allow to approximately solve the subproblems and still obtain $\mathcal{O}(\epsilon,\sqrt{\epsilon})$-second-order stationary points with high probability, but only require $\tilde{\mathcal{O}}(\ell^{2.25}\rho^{0.25}\mu^{-1.75}\epsilon^{-1.75})$ Hessian-vector products and $\tilde{\mathcal{O}}(\ell^{2}\rho^{0.5}\mu^{-2}\epsilon^{-1.5})$ gradient ascent steps. Numerical results show the efficiency of the proposed algorithms.

15:50-16:30, Feb. 3, Tuesday

Speaker: Jiaojiao Zhang

Title: Locally Differentially Private Online Federated Learning with Correlated Noise

Abstract: In this talk, we introduce a locally differentially private (LDP) algorithm for online federated learning that employs temporally correlated noise to improve utility while preserving privacy. To address challenges posed by the correlated noise and local updates with streaming non-IID data, we develop a perturbed iterate analysis that controls the impact of the noise on the utility. Moreover, we demonstrate how the drift errors from local updates can be effectively managed for several classes of nonconvex loss functions. Subject to a given LDP budget, we establish a dynamic regret bound that quantifies the impact of key parameters and the intensity of changes in the

dynamic environment on the learning performance. Numerical experiments confirm the efficacy of the proposed algorithm.

16:30-16:50, Feb. 3, Tuesday
Speaker: Yutong He
Title: Subspace Optimization for Large Language Models with Convergence Guarantees

Abstract: Subspace optimization algorithms, such as GaLore (Zhao et al., 2024), have gained attention for pre-training and fine-tuning large language models (LLMs) due to their memory efficiency. However, their convergence guarantees remain unclear, particularly in stochastic settings. In this paper, we reveal that GaLore does not always converge to the optimal solution and provide an explicit counterexample to support this finding. We further explore the conditions under which GaLore achieves convergence, showing that it does so when either (i) a sufficiently large mini-batch size is used or (ii) the gradient noise is isotropic. More significantly, we introduce GoLore (Gradient random Low-rank projection), a novel variant of GaLore that provably converges in typical stochastic settings, even with standard batch sizes. Our convergence analysis extends naturally to other subspace optimization algorithms. Finally, we empirically validate our theoretical results and thoroughly test the proposed mechanisms. Codes are available at https://github.com/pkumelon/Golore.

16:50-17:10, Feb. 3, Tuesday
Speaker: Benqi Liu
Title: TBA

Abstract: TBA

08:30-09:10, Feb. 4, Wednesday
Speaker: Qing Ling
Title: TBA

Abstract: TBA

09:10-09:50, Feb. 4, Wednesday
Speaker: Songtao Lu
Title: Bilevel Optimization with Game-Theoretic Lower-Level Problems for Model Training

Abstract: In this talk, I will present recent progress on casting model training as a bilevel optimization problem in which the lower-level problem is a game. The resulting formulation requires optimizing over the set of equilibrium solutions, with a general nonconvex (or weakly convex) objective at the upper level and a lower-level game such as a min–max game or a multi-objective game. Under mild conditions on the lower-level loss functions with respect to the lower-level decision variables, we show that first-order methods can find approximately stationary

solutions for this class of bilevel problems with provable theoretical guarantees. In the deterministic setting, our methods require on the order of one over epsilon iterations to reach an epsilon-stationary point, which matches the best-known complexity guarantees for this setting. We also report numerical experiments demonstrating the effectiveness of the proposed algorithms and their advantages over related methods in the literature.

10:20-11:00, Feb. 4, Wednesday
Speaker: Hailin Sun
Title: Solutions of Two-stage Stochastic Minimax Problems

Abstract: This paper introduces a class of two-stage stochastic minimax problems where the first-stage objective function is nonconvex-concave while the second-stage objective function is strongly convex-concave. We establish properties of the second-stage minimax value function and solution functions, and characterize the existence and relationships among saddle points, minimax points, and KKT points. We apply the sample average approximation (SAA) to the class of two-stage stochastic minimax problems and prove the convergence of the KKT points as the sample size tends to infinity. An inexact parallel proximal gradient descent ascent algorithm is proposed to solve this class of problems with the SAA. Numerical experiments demonstrate the effectiveness of the proposed algorithm and validate the convergence properties of the SAA approach.

11:00-11:40, Feb. 4, Wednesday
Speaker: Ziyan Luo
Title: TBA

Abstract: TBA

08:30-09:10, Feb. 5, Thursday
Speaker: Yuhong Dai
Title: TBA

Abstract: TBA

09:10-09:50, Feb. 5, Thursday
Speaker: Yancheng Yuan
Title: A Perturbed DCA for Computing d-Stationary Points of Nonsmooth DC Programs

Abstract: This talk introduces an efficient perturbed difference-of-convex algorithm (pDCA) for computing d-stationary points of an important class of structured nonsmooth difference-of-convex problems. Compared to the principal algorithms introduced in [J.-S. Pang, M. Razaviyayn, and A. Alvarado, Math. Oper. Res. 42(1):95--118 (2017)], which may require solving several subproblems for a one-step update, pDCA only requires solving a single subproblem. Therefore, the computational cost of pDCA for one-step update is comparable to the widely used difference-of-convex algorithm (DCA) introduced in [D. T. Pham and H. A. Le Thi, Acta Math. Vietnam. 22(1):289--355 (1997)] for computing a critical point. Importantly, under practical assumptions,

we prove that every accumulation point of the sequence generated by pDCA is a d-stationary point almost surely. Numerical experiment results on several important examples of nonsmooth DC programs demonstrate the efficiency of pDCA for computing d-stationary points. This talk is based on the joint work with Zhangcheng Feng.

10:20-11:00, Feb. 5, Thursday
Speaker: Shenglong Hu
Title: TBA
Abstract: TBA

11:00-11:40, Feb. 5, Thursday
Speaker: Liang Chen
Title: An efficient second-order cone programming approach for dynamic optimal transport on staggered grid discretization
Abstract: This paper proposes an efficient numerical method based on second-order cone programming (SOCP) to solve dynamic optimal transport (DOT) problems with quadratic cost on staggered grid discretization. By properly reformulating the discretized DOT problem into an equivalent linear SOCP, we develop a highly efficient implementation using an inexact decomposition-based proximal augmented Lagrangian method. The proposed approach is provided as an open-source software package to facilitate reproducibility and further research. Numerical experiments on a diverse range of DOT problems demonstrate that our method significantly outperforms several state-of-the-art solvers in terms of computational efficiency. Furthermore, the approach exhibits robust performance when handling problems involving measures that are not strictly positive or irregular domains with obstacles.

14:00-14:40, Feb. 5, Thursday
Speaker: Ruoyu Sun
Title: TBA
Abstract: TBA

14:40-15:20, Feb. 5, Thursday
Speaker: Xiao Li
Title: Optimizers in Post-Training and Pre-Training: An Introduction
Abstract: This talk provides an overview of representative optimization methods used in both post-training and pre-training of large language models (LLMs). Optimizer design in these two regimes is driven by different objectives and constraints. In post-training, the primary concern is memory efficiency, as fine-tuning is often performed under limited hardware budgets. This has led to parameter-efficient methods such as LoRA. In contrast, pre-training operates at a much larger scale, where memory-saving techniques alone are insufficient. Instead, the focus shifts toward

designing optimization schemes that can improve convergence speed, stability, and final performance beyond the standard Adam optimizer. I will introduce several such methods, including Shampoo and Muon, and discuss the design principles that distinguish post-training and pre-training optimizers.

15:50-16:30, Feb. 5, Thursday
Speaker： Qi Deng

Title： On some universal methods for convex optimization

Abstract： We present some new parameter-free first-order methods for convex optimization problems in which the objective function exhibits Hölder smoothness. Inspired by the recently proposed distance-over-gradient (DOG) technique, we propose an accelerated distance-adaptive method which achieves optimal anytime convergence rates for Hölder smooth problems without requiring prior knowledge of smoothness parameters or explicit parameter tuning. We also consider convex stochastic optimization, for which we further present a parameter-free accelerated method that eliminates the need for line-search procedures. Preliminary experimental results show the effectiveness of our approach on convex nonsmooth problems and its advantages over existing parameter-free or accelerated methods.

16:30-16:50, Feb. 5, Thursday
Speaker： Qiming Dai

Title： TBA

Abstract： TBA

16:50-17:10, Feb. 5, Thursday
Speaker： Mingyu Mo

Title： TBA

Abstract： TBA

08:30-09:10, Feb. 6, Friday
Speaker： Kuang Bai

Title： Directional combined approach for bilevel optimization

Abstract： TBA

09:10-09:50, Feb. 6, Friday
Speaker： Haishan Ye

Title： Explicit and Non-asymptotic Query Complexities of Rank-Based Zeroth-order Algorithms on Smooth Functions

Abstract: Rank-based zeroth-order (ZO) optimization---which relies only on the ordering of function evaluations---offers strong robustness to noise and monotone transformations, and underlies many successful algorithms such as CMA-ES, natural evolution strategies, and rank-based genetic algorithms. Despite its widespread use, the theoretical understanding of rank-based ZO methods remains limited: existing analyses provide only asymptotic insights and do not yield explicit convergence rates for algorithms selecting the top-k directions. This work closes this gap by analyzing a simple rank-based ZO algorithm and establishing the first *explicit*, and *non-asymptotic* query complexities. For a d-dimension problem, if the function is L-smooth and μ-strongly convex, the algorithm achieves $\tilde{\mathcal{O}}\left(\frac{dL}{\mu}\log\frac{dL}{\mu\delta}\log\frac{1}{\varepsilon}\right)$ query complexity to find an ε-suboptimal solution, and for smooth nonconvex objectives it reaches $\mathcal{O}\left(\frac{dL}{\varepsilon}\log\frac{1}{\varepsilon}\right)$. Notation $\mathcal{O}(\cdot)$ hides constant terms and $\tilde{\mathcal{O}}(\cdot)$ hides extra $\log\log\frac{1}{\varepsilon}$ terms. These query complexities hold with a probability at least $1-\delta$ with $0<\delta<1$. The analysis in this paper is novel and avoids classical drift and information-geometric techniques. Our analysis offers new insight into why rank-based heuristics lead to efficient ZO optimization.

10:20-11:00, Feb. 6, Friday
Speaker: **Siqi Zhang**
Title: **Recent Progress on Nonconvex Minimax Optimization and Beyond**
Abstract: TBA

11:00-11:40, Feb. 6, Friday
Speaker: **Yang Liu**
Title: **Solver-Aware High-Order Optimization for Large-Scale Nonconvex Problems**

Abstract: Many modern learning workloads, including large-model training, can be formulated as large-scale, generally nonconvex optimization problems under stringent memory and compute budgets. This talk presents a solver-aware perspective on scalable high-order optimization, coupling algorithmic design with Hessian- and tensor-free subproblem solvers. First, we discuss second-order methods with full-curvature awareness. By detecting and handling nonpositive curvature on the fly within Krylov-subspace iterations for solving the Newton system, we establish methods with global complexity guarantees and superlinear local convergence under standard regularity conditions. Moving to third-order and beyond (arbitrary order p), we present practical and theoretical results for adaptively regularized tensor methods (ARp). We introduce improved strategies including efficient regularization updates and a novel pre-rejection mechanism. Furthermore, we establish a sharp local pth-order convergence rate for ARp, contingent on the right choice of local subproblem minimizer. The discussed frameworks offer principled paths towards more efficient and robust optimizers for resource-constrained large-model training.

List of Participants

姓名	单位
白旷	湖南大学
白敏茹	湖南大学
陈亮	湖南大学
陈士祥	中国科学技术大学
程郁琨	江南大学商学院
戴其铭	北京大学
戴彧虹	中国科学院数学与系统科学研究院
邓琪	上海交通大学
刁若愉	中国科学院数学与系统科学研究院
韩德仁	北京航空航天大学
何雨桐	北京大学
胡胜龙	国防科技大学
户将	清华大学
黄文	厦门大学数学科学学院
江波	上海财经大学
江如俊	复旦大学
姜波	南京师范大学
蒋建林	南京航空航天大学
李肖	香港中文大学（深圳）
凌青	中山大学
刘本起	北京大学
刘洋	大湾区大学
卢松涛	香港中文大学
罗自炎	北京交通大学
莫明煜	北京大学
倪洁	北京理工大学
苏文藻	香港中文大学
孙聪	北京邮电大学
孙海琳	南京师范大学
孙若愚	香港中文大学（深圳）
王丽平	南京航空航天大学

王祥丰	华东师范大学
王子卓	香港中文大学（深圳）
谢家新	北京航空航天大学
徐姿	上海大学
叶海山	西安交通大学
袁坤	北京大学
袁雁城	香港理工大学
张娇娇	大湾区大学
张立卫	大连理工大学
张思奇	南京大学

Access Guide

· 地理位置

中心位于云南省昆明市宜良县柴石滩水库库区内国家一级公益林中；三面环山，一面向水，环境优美，风景秀丽、气候宜人；海拔为1700米，距离昆明长水国际机场约90公里，乘车时间约1.5小时。距离石林风景区25公里，距离九乡风景区17公里。

· 专车服务

中心周日全天接机、接站，周五下午、周六全天安排车辆送机、送站。其他时间需自行解决交通问题，中心可协助联系车辆，但费用需自理。

由于中心外部道路尚未完善，建议参会人员尽量不要选择夜间行车。

· 出租车网约车

昆明长水国际机场或昆明南站，乘坐出租车网约车到柴石滩水库风景区天元数学国际研究交流中心下车。中途经过宜良县城，宜良县城至天元数学国际研究交流中心，有两条线路可供选择：

（一）宜良县城上汕昆高速，往石林方向，石林风景区下高速，走九石阿公路到达柴石滩水库风景区天元数学国际研究交流中心。这条线路，**路面宽，急弯道少，路况相对较好。**

昆明长水国际机场至天元数学国际研究交流中心全程 99 **公里**，一般需要 1 **小时** 30 **分钟**；

昆明南站至天元数学国际研究交流中心全程 92 **公里**，一般需要 1 **小时** 20 **分钟**。

（二）走古柴段线路到达柴石滩水库风景区天元数学国际研究交流中心。这条线路，需要经过村镇，路面窄，急弯道多，路况相对较差。

昆明长水国际机场至天元数学国际研究交流中心全程 74 **公里**，需要 1 **小时** 25 **分钟**;

昆明南站至天元数学国际研究交流中心全程 78 **公里**，需要 1 **小时** 35 **分钟**。

· **公共交通**

1、昆明长水国际机场乘坐地铁 6 **号线**到昆明东部客运站下车，乘坐昆明至宜良大巴车到宜良客运站下车，或昆明南站乘坐昆明高铁南客运中心

至宜良大巴车到宜良客运站下车，乘坐20路公交车到北古城镇下车，乘坐5 路公交车到柴石滩水库下车。下车终点站距离天元数学国际研究交流中心约4 公里，联系工作人员接送。

2、公共交通工具运营时间

地铁6 号线首班时间6：20，末班时间23：00，运营间隔为25 分钟。

昆明至宜良大巴车首班时间 7：00，末班时间 21：00，运营间隔为 1 小时。

宜良 20 路公交车首班时间 7：00，末班时间 19：00，运营间隔为 30 分钟。

宜良 5 路公交车首班时间 8：00，末班时间 18：00，运营间隔为 2 小时。

注：宜良5 路、20 路，逢日期尾号3、6、9 才有班车

· 自驾

自驾可使用导航搜索“天元数学国际研究交流中心”定位，依据导航指引可到达天元数学国际研究交流中心。

Tianyuan Mathematics Research Center

天元数学国际交流中心由中国科学院、国家自然科学基金委员会、中国数学会、中国科学院数学与系统科学研究院，以及昆明市有关部门共同支持建设的数学与交叉科学交流机构。旨在搭建数学及其跨学科应用领域的学术交流平台，提升我国数学整体研究水平，成为国际一流的数学交流与合作研究中心。

该中心位于云南省昆明市宜良市柴石滩水库库区的国家级公益林内，三面环山，一面环水。该中心总用地面积约 2.7 万平方米，有研究楼、专家楼和后勤楼三栋两层主楼，可容纳近 200 人进行学术活动，并配有图书馆和阅览室、餐厅以及一定数量的办公和住宿用房。

该中心将聚焦数学科学的重大前沿方向和重大问题，组织开展形式多样的国内外优秀专家学术交流与合作研究活动，推动实质性合作研究形成优势方向，促进数学学科发展。同时，该中心将重点培养青年人才，普及数学科学，增强公众对数学科学的认识，提高我国数学的整体研究水平，努力建设国际一流的数学交流与合作研究中心。

该中心将支持数学科学的主要方向。在基础数学领域，应该布局对未来数学有重大领导作用的方向，包括数论和代数、几何和拓扑学、现代分析和数学物理、概率论和随机分析。在应用数学领域，应布局国家战略急需的应用数学关键通用方法领域，包括数据科学与人工智能数学理论、科学与工程计算方法、复杂系统优化与控制理论、计算机数学与密码学等。该中心还支持多样化的学术研究，不仅在数学学科，而且在与数学交叉的领域，如物理、医学、生物学和信息等。该中心将邀请相关领域的顶尖专家，特别是活跃在其领域前沿的杰出年轻研究人员，进行合作和交流，从而通过多学科的交叉融合，产生新的思想，推动重大问题的解决。

中心将借鉴国际顶尖学术交流机构的成功经验，建立合理的机构运作和活动组织机制。该中心不会设立常设研究职位，工作人员的流动性将与学术活动的多样性相匹配。中心将组织学术活动，包括研讨会、著名学者

系列讲座和青年研讨会、暑期研究生课程等。中心学术委员会将负责规划和验证中心的学术主题和活动。所有活动都向国内数学界开放。在中心成立初期，学术活动主要以每周一次的研讨会为基础。