

Yuan, Kun

| | | |
|---------------------|---|---|
| CONTACT INFORMATION | Center for Machine Learning Research Peking University | E-mail: kunyuan@pku.edu.cn Cellphone: (+86)13140810278 |
| RESEARCH INTERESTS | Signal Processing, Optimization, Machine Learning | |
| EDUCATION | University of California, Los Angeles , Los Angeles <i>Ph.D. in Electrical and Computer Engineering</i> • Advisor: Professor Ali H. Sayed | 2014 – 2019 |
| | University of Science and Technology of China , Hefei <i>M.S. in Electrical Engineering</i> | 2011 – 2014 |
| | Xidian University , Xi'an <i>B.E. in Electrical Engineering</i> | 2007 – 2011 |
| ACADEMIC AWARDS | IEEE CloudCom Distinguished Paper Award | 2025 |
| | IEEE Signal Processing Society Young Author Best Paper Award One of the most prestigious awards in the signal processing society | 2018 |
| | Distinguished Paper Award in ICCM 2017 | 2017 |
| EXPERIENCE | Peking University , Beijing, P. R. China Assistant Professor | Nov. 2022 – Now |
| | DAMO Academy, Alibaba (US) Group , Bellevue, USA Research Scientist | Aug. 2019 – Oct. 2022 |
| | Microsoft Research , Redmond, USA Research Intern | Jun. 2018 – Sep. 2018 |
| | École Polytechnique Fédérale de Lausanne (EPFL) , Lausanne, Switzerland Visiting Researcher | Jan. 2018 – Jun. 2018 |
| JOURNAL PAPERS | Y. He, Q. Shang, X. Huang, J. Liu, and K. Yuan . “A Mathematics-Inspired Learning-to-Optimize Framework for Decentralized Optimization”, <i>IEEE Transactions on Signal Processing</i> , 2025 | |
| | B. Kong, S. Zhu, S. Lu, X. Huang, and K. Yuan , “Decentralized Bilevel Optimization: A Perspective from Transient Iteration Complexity,” <i>Journal of Machine Learning Research</i> , 2025. | |
| | T. Sun, X. Liu, and K. Yuan , “Revisiting Gradient Normalization and Clipping for Nonconvex SGD under Heavy-Tailed Noise: Necessity, Sufficiency, and Acceleration,” <i>Journal of Machine Learning Research</i> , 2025. | |
| | L. Liang, X. Huang, R. Xin, K. Yuan . “Understanding the Influence of Digraphs on Decentralized Optimization: Effective Metrics, Lower Bound, and Optimal Algorithm”, <i>SIAM Journal on Optimization</i> , 2025. | |
| | M. Gurbuzbalaban, Y. Hu, U. Simsekli, K. Yuan , and L. Zhu. “Heavy-Tail Phenomenon in Decentralized SGD”, <i>IISE Transactions</i> , 2024. | |

X. Wang, T. Zhou, J. Zhu, J. Liu, **K. Yuan**, T. Yao, W. Yin, R. Jin, H. Cai. “S3 Attention: Improving Long Sequence Attention with Smoothed Skeleton Sketching”, *IEEE Journal of Selected Topics in Signal Processing*, 2024.

S. A. Alghunaim and **K. Yuan**. “An Enhanced Gradient-Tracking Bound for Distributed Online Stochastic Convex Optimization”, *Signal Processing*, 2023.

K. Yuan, S. A. Alghunaim, and X. Huang. “Removing Data Heterogeneity Influence Enhances Network Topology Dependence of Decentralized SGD,” *Journal of Machine Learning Research (JMLR)*, 2023.

S. A. Alghunaim and **K. Yuan**. “A Unified and Refined Convergence Analysis for Non-convex Decentralized Learning”, *IEEE Transactions on Signal Processing*, 2022.

L. Cassano, **K. Yuan**, and A. H. Sayed. “Multiagent fully decentralized value function learning with linear convergence rates”, *IEEE Transactions on Automatic Control*, 2020.

S. A. Alghunaim, E. K. Ryu, **K. Yuan**, and A. H. Sayed. “Decentralized Proximal Gradient Algorithms with Linear Convergence Rates”, *IEEE Transactions on Automatic Control*, 2020.

X. Mao, **K. Yuan**, Y. Hu, Y. Gu, and A. H. Sayed, W. Yin. “Walkman: A communication-efficient random-walk algorithm for decentralized optimization”, *IEEE Transactions on Signal Processing*, 2020.

K. Yuan, W. Xu, and Q. Ling. “Can Primal Methods Outperform Primal-dual Methods in Decentralized Dynamic Optimization?”, *IEEE Transactions on Signal Processing*, 2020.

K. Yuan, S. A. Alghunaim, and B. Ying, and A. H. Sayed. “On the Influence of Bias-Correction on Distributed Stochastic Optimization”, *IEEE Transactions on Signal Processing*, 2020.

B. Ying, **K. Yuan**, and A. H. Sayed. “Variance-Reduced Stochastic Learning under Random Reshuffling”, *IEEE Transactions on Signal Processing*, 2020.

L. Cassano, **K. Yuan**, and A. H. Sayed. “Multiagent Fully Decentralized Value Function Learning with Linear Convergence Rates”, *IEEE Transactions on Automatic Control*, 2020.

X. Mao, **K. Yuan**, Y. Hu, Y. Gu, A. H. Sayed, and W. Yin. “Walkman: A Communication-Efficient Random-Walk Algorithm for Decentralized Optimization”, *IEEE Transactions on Signal Processing*, 2020.

S. A. Alghunaim, **K. Yuan**, and A. H. Sayed. “A Proximal Diffusion Strategy for Multiagent Optimization with Sparse Affine Constraints”, *IEEE Transactions on Automatic Control*, 2019.

B. Ying, **K. Yuan**, and A. H. Sayed. “Dynamic Average Diffusion with Randomized Coordinate Updates”, *IEEE Transactions on Signal and Information Processing over Networks*, 2019.

K. Yuan, B. Ying, J. Liu, A. H. Sayed, “Efficient Variance-Reduced Learning for Fully Decentralized On-Device Intelligence”, *IEEE Transactions on Signal Processing*, 2018.

B. Ying, **K. Yuan**, and A. H. Sayed. “Supervised Learning under Distributed Features”, *IEEE Transactions on Signal Processing*, 2018.

K. Yuan, B. Ying, X. Zhao, A. H. Sayed, “Exact Diffusion for Distributed Optimization and Learning – Part I: Algorithm Development”, *IEEE Transactions on Signal Processing*, 2018.

K. Yuan, B. Ying, X. Zhao, A. H. Sayed, “Exact Diffusion for Distributed Optimization and Learning – Part II: Convergence Analysis”, *IEEE Transactions on Signal Processing*, 2018.

B. Ying, **K. Yuan**, V. Stefan, and A. H. Sayed. “Stochastic Learning Under Random Reshuffling with Constant Step-sizes”, *IEEE Transactions on Signal Processing*, 2018.

T. Wu, **K. Yuan**, Q. Ling, W. Yin, and A. H. Sayed. “Decentralized Consensus Optimization with Asynchrony and Delays”, *IEEE Transactions on Signal and Information Processing over Networks*, 2017.

K. Yuan, B. Ying, S. Vlaski, and A. H. Sayed. “Stochastic Gradient Descent with Finite Samples Sizes”, *IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, 2016.

K. Yuan, B. Ying, and A. H. Sayed, “On the Influence of Momentum Acceleration on Online Learning”, *Journal of Machine Learning Research*, 2016.

K. Yuan, Q. Ling, and W. Yin, “On the Convergence of Decentralized Gradient Descent”, *SIAM Journal on Optimization*, 2016.

K. Yuan, Q. Ling, and Z. Tian, “A Decentralized Linear Programming Approach to Energy-Efficient Event Detection”, *International Journal of Sensor Networks*, 2015.

K. Yuan, Q. Ling, and Z. Tian, “Communication-efficient Decentralized Event Monitoring in Wireless Sensor Networks”, *IEEE Transactions on Parallel and Distributed Systems*, 2014.

W. Shi, Q. Ling, **K. Yuan**, G. Wu, and W. Yin, “On the Linear Convergence of the ADMM in Decentralized Consensus Optimization”, *IEEE Transactions on Signal Processing*, 2014.

CONFERENCE PAPERS

Y. Liu, J. Deng, Y. He, X. Wang, T. Yao, and **K. Yuan**, “MISA: Memory-Efficient LLMs Optimization with Module-wise Importance Sampling”, *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

R. Hu, Y. He, R. Yan, M. Sun, B. Yuan, and **K. Yuan**, “MeCeFO: Enhancing LLM Training Robustness via Fault-Tolerant Optimization”, *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Z. Wu, Z. Yuan, Y. Dong, C. Zhang, C. Fang, **K. Yuan**, and Z. Lin, “Improving Model Representation and Reducing KV Cache via Skip Connections with First Value Heads”, *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Y. He, P. Li, Y. Hu, C. Chen, and **K. Yuan**, “Subspace Optimization for Large Language Models with Convergence Guarantees”, *International Conference on Machine Learning (ICML)*, 2025.

Y. Chen, Y. Zhang, Y. Liu, **K. Yuan**, and Z. Wen, “A Memory Efficient Randomized Subspace Optimization Method for Training Large Language Models”, *International Conference on Machine Learning (ICML)*, 2025.

Y. Song, P. Li, B. Gao, and **K. Yuan**, “Distributed Retraction-Free and Communication-Efficient Optimization on the Stiefel Manifold”, *International Conference on Machine Learning (ICML)*, 2025.

L. Liang, G. Luo, X. Chen, and **K. Yuan**, “Achieving Linear Speedup and Optimal Complexity for Decentralized Optimization over Row-stochastic Networks”, *International Conference on Machine Learning (ICML)*, 2025.

L. Chen, Q. Xiao, E. H. Fukuda, X. Chen, **K. Yuan**, and T. Chen, “Efficient Multi-Objective Learning under Preference Guidance: A First-Order Penalty Approach”, *International Conference on Machine Learning (ICML)*, 2025.

Y. Chen, Y. Zhang, L. Cao, **K. Yuan**, and Z. Wen, “Enhancing Zeroth-Order Fine-Tuning for Language Models with Low-Rank Structures”, *International Conference on Learning Representations (ICLR)*, 2025.

S. Zhu, B. Kong, S. Lu, X. Huang, and **K. Yuan**, “SPARKLE: A Unified Single-Loop Primal-Dual Framework for Decentralized Bilevel Optimization”, *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Y. He, J. Hu, X. Huang, S. Lu, B. Wang, and **K. Yuan**, “Distributed Bilevel Optimization with Communication Compression”, *International Conference on Machine Learning (ICML)*, 2024.

Elsa Rizk, **K. Yuan**, Ali H. Sayed. “Asynchronous Diffusion Learning with Agent Subsampling and Local Updates”, *The IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2024.

Z. Cheng, X. Huang, P. Wu, and **K. Yuan**. “Momentum Benefits Non-IID Federated Learning Simply and Provably”, *International Conference on Representation Learning (ICLR)*, 2024.

Y. He, X. Huang, and **K. Yuan**. “Unbiased Compression Saves Communication in Distributed Optimization: When and How Much?”, *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

H. Yuan, S. A. Alghunaim, and **K. Yuan**. “Achieving Linear Speedup with Network-Independent Learning Rates in Decentralized Stochastic Optimization”, *IEEE Conference on Decision and Control (CDC)*, 2023.

E. D. H. Nguyen, S. A. Alghunaim, **K. Yuan**, and C. A. Uribe. “On the Performance of Gradient Tracking with Local Updates”, *IEEE Conference on Decision and Control (CDC)*, 2023.

L. Ding, K. Jin, B. Ying, **K. Yuan**, and W. Yin[†]. “DSGD-CECA: Decentralized SGD with Communication-Optimal Exact Consensus Algorithm”, *The International Conference on Machine Learning (ICML)*, 2023.

Y.-F. Zhang, X. Wang, K. Jin, **K. Yuan**, Z. Zhang, L. Wang, R. Jin, and T. Tan. “AdaNPC: Exploring Non-Parametric Classifier for Test-Time Adaptation”, *The International Conference on Machine Learning (ICML)*, 2023.

L. Yang, K. Yu, T. Tang, J. Li, **K. Yuan**, L. Wang, X. Zhang, and P. Chen. “BEVHeight: A Robust Framework for Vision-based Roadside 3D Object Detection”, *The IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2023.

X. Huang, Y. Chen, W. Yin, and **K. Yuan**, “Lower Bounds and Nearly Optimal Algorithms in Distributed Learning with Communication Compression”, *Neural Information Processing Systems (NeurIPS)*, 2022.

Z. Song, W. Li, K. Jin, L. Shi, M. Yan, W. Yin, and **K. Yuan**, “Communication-Efficient Topologies for Decentralized Learning with $O(1)$ Consensus Rate”, *Neural Information Processing Systems (NeurIPS)*, 2022.

K. Yuan, X. Huang, Y. Chen, X. Zhang, Y. Zhang, and P. Pan, “Revisiting Optimal Convergence Rate for Smooth and Non-Convex Stochastic Decentralized Optimization”, *Neural Information Processing Systems (NeurIPS)*, 2022.

K. Yuan, Z. Wu, and Q. Ling, “A Byzantine-Resilient Dual Subgradient Method for Vertical Federated Learning”, *The IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2022.

Z. Hou, M. Qin, F. Sun, X. Ma, **K. Yuan**, Y. Xu, Y.-K. Chen, R. Jin, Y. Xie, and S.-Y. Kung, “CHEX: Channel Exploration for CNN Model Compression”, *The IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR)*, 2022.

X. Ma, M. Qin, F. Sun, Z. Hou, **K. Yuan**, Y. Xu, Y. Wang, Y.-K. Chen, R. Jin, and Y. Xie, “Effective Model Sparsification by Scheduled Grow-and-Prune Methods”, *The International Conference on Learning Representations (ICLR)*, 2022.

K. Yuan, Y. Chen, X. Huang, Y. Zhang, P. Pan, Y. Xu, and W. Yin, “DecentLaM: Decentralized Momentum SGD for Large-batch Deep Training”, *International Conference on Computer Vision (ICCV)*, 2021.

X. Huang, **K. Yuan**, X. Mao, and W. Yin, “Improved Analysis and Rates for Variance Reduction under Without-replacement Sampling Orders”, *Neural Information Processing Systems (NeurIPS)*, 2021.

B. Ying, **K. Yuan**, Y. Chen, H. Hu, P. Pan, W. Yin, “Exponential Graph is Provably Efficient for Decentralized Deep Training”, *Neural Information Processing Systems (NeurIPS)*, 2021.

Y. Chen, **K. Yuan**, Y. Zhang, P. Pan, Y. Xu, and W. Yin, “Accelerating Gossip SGD with Periodic Global Averaging”, *International Conference on Machine Learning (ICML)*, 2021.

S. Alghunaim, **K. Yuan**, A. H. Sayed, “A Linearly Convergent Proximal Gradient Algorithm for Decentralized Optimization”, *Neural Information Processing Systems (NeurIPS)*, 2019.

Y. Liu, **K. Yuan**, G. Wu, Z. Tian, and Q. Ling, “Decentralized Dynamic Admm with Quantized and Censored Communications”, *Asilomar Conference on Signals, Systems, and Computers*, 2019.

S. A. Alghunaim, **K. Yuan**, and A. H. Sayed, “Dual Coupled Diffusion for Distributed Optimization with Affine Constraints”, *The IEEE Conference on Decision and Control (CDC)*, 2018.

S. A. Alghunaim, **K. Yuan**, and A. H. Sayed, “Decentralized Exact Coupled Optimization”, *Allerton Conference on Communication, Control, and Computing (Allerton)*, 2017.

B. Ying, **K. Yuan**, S. Vlaski, and A. H. Sayed. “On the Performance of Random Reshuffling in Stochastic Learning”, *Information Theory and Application Workshop (ITA)*, 2017.

B. Ying, **K. Yuan**, and A. H. Sayed, “Online Dual Coordinate Ascent Learning”, *European Signal Processing Conference (EUSIPCO)*, 2016.

K. Yuan, Q. Ling, W. Yin, and A. Ribeiro, “A Linearized Bregman Algorithm for Decentralized Basis Pursuit”, *European Signal Processing Conference (EUSIPCO)*, 2013.

W. Shi, Q. Ling, **K. Yuan**, G. Wu, and W. Yin, “Linearly Convergent Decentralized Consensus Optimization with the Alternating Direction Method of Multipliers”, *The IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

SOFTWARE

MindOpt: An Industrial-Level Optimization Solver Suite

Aug. 2019 – Oct. 2022

MindOpt is an industrial-level solver suite for mathematical optimization problems. It is independently developed by Decision Intelligence Lab, Alibaba DAMO Academy. It provides linear, nonlinear, and integer programming solutions to users. MindOpt linear programming is among the fastest solvers worldwide. [[official website]]

- Main contributer to the first-order optimization algorithms in MindOpt; built first-order optimization modules that help MindOpt Simplex solver break the world record and top the prestigious Mittelmann benchmark list in December 2020. [[Benchmark list]] [[News]]
- Main contributer to the Parallel Interior Point Method in MindOpt.

Decentralized optimization algorithms are low-communication-overhead alternatives to traditional distributed algorithms using a center to conduct global average. However, the lack of an easy-to-use and efficient software package has kept most decentralized algorithms merely on paper. BlueFog is the first python library for straightforward, high-performance implementations of diverse decentralized algorithms. It is open source at <https://github.com/Bluefog-Lib>. BlueFog was invited to give keynote lectures in the US East Coast Optimization Meeting 2021 and China Symposium on Machine Learning 2020.

- Main contributor to the development of decentralized algorithms utilized in BlueFog.
- Main contributor to the tutorial for BlueFog.

TEACHING

| | |
|-----|--|
| PKU | Optimization in Deep Learning |
| PKU | Introductions to Large Language Models |
| PKU | Optimization in Deep Learning |

| |
|-------------|
| Fall 2024 |
| Spring 2024 |
| Fall 2023 |