Hongpei Li

✉ ishongpeili@gmail.com | ⌂ Github | 🏠 Homepage | ☁ Google Scholar

EDUCATION

Shanghai University of Finance and Economics (SUFE)   2021/09 - 2025/07
Honor Bachelor of Engineering in Data Science and Big Data Technology
Pilot Class of Research Institute for Interdisciplinary Science   Shanghai, China
GPA 3.77/4.0   Average Score 89.44/100   Major GPA 3.87   Major Average Score 91.48/100

Main Courses:
- Mathematics: Linear Algebra, Mathematical Analysis, Probability, Statistics, Stochastic Processes, Operations Management, Operation Research and Numerical Methods.
- Computer Science: Python, C++, Data Structures, Algorithms, Data Mining, and Machine/Deep Learning.
- Economics: Micro/Macroeconomics, Econometrics, and Money & Banking.

Honors & Awards:
- Honor Bachelor Degree, 2025. **(Top 5% in Major)**
- Honor of Excellence Graduation Thesis, 2025. **(Top 3% in Major)**
- Shanghai Municipal Bronze Award in the Internet+ University Student Innovation and Entrepreneurship Competition. **(0.13%)**

RESEARCH INTERESTS

My research interests include Optimization, Artificial Intelligence and the Interdisciplinary of Operations Research and Machine Learning. For example, I am interested in the following topics:
- **Optimization**: Design and implement efficient algorithms for large-scale optimization problems.
- **Artificial Intelligence**: Utilize Artificial Intelligence techniques to improve scientific fields, such as engineering problems.
- **Computing**: Improve the efficiency of training and inference of LLMs and large-scale optimization problems.

PUBLICATIONS

- [P1] Restarted Primal-Dual Hybrid Conjugate Gradient Method for Large-Scale Quadratic Programming (Accepted by **INFORMS Journal on Computing**)
  Y. Huang, W. Zhang, **H. Li**, D. Ge, H. Liu, and Y. Ye. (2024). *arXiv preprint* ([Paper])([Code])

- [P2] Solving Integrated Process Planning and Scheduling Problem via Graph Neural Network Based Deep Reinforcement Learning
  H. Li, H. Zhang, Z. He, Y. Jia, B. Jiang, X. Huang, and D. Ge. (2024). *arXiv preprint* ([Paper])([Code])

- [P3] BenLOC: A Benchmark for Learning to Configure MIP Optimizers
  H. Li, Z. He, Y. Wang, S. Pu, Q. Deng, W. Tu, and D. Ge. (2025). *arXiv preprint* ([Paper])([Code])

- [P4] PDHCG: A Scalable First-Order Method for Large-Scale Competitive Market Equilibrium Computation
  H. Liu, Y. Huang, **H. Li**, D. Ge and Y. Ye (2025). *arXiv preprint* ([Paper])([Code])

- [P5] FMIP: Joint Continuous-Integer Flow for Mixed-Integer Linear Programming (Submitted to ICLR 2026)
  H. Li, H. Yuan, H. Zhang, J. Lin, D. Ge, M. Wang and Y. Ye (2025). *arXiv preprint* ([Paper])([Code])

- [P6] OptPipe: Memory- and Scheduling-Optimized Pipeline Parallelism for LLM Training (Submitted to ICLR 2026)
  H. Li, H. Zhang, H. Liu, D. Ge and Y. Ye (2025). *arXiv preprint* ([Paper])

SELECTED RESEARCH EXPERIENCE

Improvement and Post-processing Phase of Primal-Dual Hybrid Gradient Method For Large-Scale Linear Programming   Oct.2025 - Present
Adviser: Prof. Haihao Lu   Massachusetts Institute of Technology
Further improve the efficiency and design post-processing phase of cuPDLPx, a GPU-accelerated primal-dual hybrid gradient method for large-scale linear programming problems.
- Develop fused GPU kernels, specified for the features of problem instances, to further improve the efficiency of cuPDLPx.
- Design a post-processing phase to further improve the solution quality of cuPDLPx after the major PDHG iterations.

GPU-Accelerated First-Order Method for the SDP Relaxation of Polynomial Optimization Jun.2025 - Present
Adviser: Prof. Yinyu Ye Stanford University

An efficient and GPU-accelerated first order method for large-scale SDP relaxation of polynomial optimization problems.
- Implement whole algorithm framework and solver components.
- Design efficient multi-GPU PSD projection algorithm and other fused GPU kernels.

Improve Pipeline Parallelsim using Mathematic Programming [P6] Mar.2025 - Present
Adviser: Prof. Yinyu Ye Stanford University

Use mathematic programming to optimize the memory and scheduling of pipeline parallelism in LLM training.
- Implement a framework based on Megatron-LM, which including building and solving a mixed-integer programming model.
- Design practical techniques to reduce required solving time and improve the solution quality.

GPU-Accelerated First-Order Method for Large-Scale Fisher Equilibrium Problems [P4] Nov.2024 - Apr.2025
Adviser: Prof. Yinyu Ye Stanford University

An efficient and GPU-accelerated first order method for large-scale Fisher equilibrium problems.
- Implement efficient GPU kernels for efficiency improvement.
- Develop a CUDA-C based solver for large-scale Fisher equilibrium problems.

Generative Models for Linear Programming&Mixed Integer Programming [P5] Sep.2024 - May.2025
Adviser: Prof. Mengdi Wang, Prof. Yinyu Ye Princeton University, Stanford University

Use generative models to accelerate the solving of linear programming (LP) and mixed-integer programming (MIP) problems.
- Build a framework for generating high qulified solution of LPs and MILPs using flow matching.
- Design a joint continuous-integer flow for MILP problems and a optimality-aware training-free guidance strategy.

Deep Reinforcement Learning (DRL) for Scheduling Problems [P2] Mar.2024 - Aug.2024
Adviser: Prof. Dongdong Ge Prof. Bo Jiang Shanghai Jiao Tong University, SUFE, Cardinal Optimizer

Use DRL to solve the integrated process planning and scheduling problem, a kind of realistic and difficult scheduling problem.
- Implement a GPU-based simulation environment for training and Design a novel graph representation of the problem, a dense reward function and a real-time action space reduction method.
- Design a pruning strategy to improve learning-guided searching method.

Primal-Dual Hybrid Conjugate Gradient Method (PDHCG) [P1] Apr.2024 - Oct.2024
Adviser: Prof. Yinyu Ye Prof. Dongdong Ge Prof. Huikang Liu Stanford University, Shanghai Jiao Tong University

An efficient and GPU-accelerated first order method for large-scale quadratic programming problems.
- Implement GPU version and low-rank acceleration of PDHCG solver.
- Implement computational techniques on GPU, including asynchronous computation, fused kernels, memory access optimization.

Machine Learning for MIP Optimizer Configuration [P3] Dec.2023 - Oct.2024
Adviser: Prof. Dongdong Ge Shanghai Jiao Tong University

Use machine learning to help configure MIP optimizers for better performance.
- Build a whole framework from dataset generation, feature engineering, model training to deployment.
- Design GNNs to predict the best configuration of MIP optimizers.

Personal Projects :octocat: Lhongpei

- **DRL for Campus Auto-Delivery Vehicle** *Developed an innovative framework integrating DRL and MILP for campus auto-delivery vehicles.*

- **First-Order Methods for Large-Scale Linear Constrained Nonlinear Programming** *Implemented a GPU-accelerated first-order method for large-scale linearly constrained nonlinear programming problems in CUDA-C.*

- Recurrent Neural Network-Based Quasi-Newton Method: A learning-based Quasi-Newton method.

- PDHCG-Net: A neural network to predict high-quality initial solutions for PDHCG method.

EMPLOYMENT HELD & VOLUNTEER EXPERIENCE

LLMs Technical Intern (Inference Optimization)	Cardinal Optimizer	Feb.2025 - Present
TA of Advanced Operations Research	Shanghai University of Finance and Economics	Feb.2024 - Jun.2024
TA of Object-Oriented Analysis and Design	Shanghai University of Finance and Economics	Sep.2023 - Jan.2024
Peer Tutor in Programming Design Foundations	Shanghai University of Finance and Economics	Sep.2022 - Dec.2022
University Students' Union	Shanghai University of Finance and Economics	Sep.2021 - Sep.2022
Campus Return Recruitment	Shanghai University of Finance and Economics	Dec.2021 - Jan.2022