Tong Yang

Email: tongyang@andrew.cmu.edu
Personal Site: https://pptmiao.github.io
Research Interests:
Optimization, Machine Learning Theory

Education

Carnegie Mellon University Aug 2023–Present
Ph.D. in Electrical and Computer Engineering

Peking University Sep 2022–June 2023
M.S. in Data Science

University of California, Berkeley Jan 2022–May 2022
Exchange Program, GPA: 4.0/4.0

Xi'an Jiaotong University Sep 2018–July 2022
B.S. in Mathematics (Honors Program), GPA: 4.01/4.3, Major rank:1/50

Work Experience

- Machine learning student researcher in Google Research May 2025–Dec 2025
- Machine learning student researcher in Meta Fair May 2026–Current

Selected Publications

Agentic Transformers Provably Learn to Search via Reinforcement Learning
Preprint, 2026
Authors: **Tong Yang**, Yu Huang, Yingbin Liang, Yuejie Chi
arXiv: https://arxiv.org/abs/2606.00183

Diffusion Controller: Framework, Algorithms and Parameterization
ICML 2026
Authors: **Tong Yang**, Moonkyung Ryu, Chih-Wei Hsu, Guy Tennenholtz, Yuejie Chi, Craig Boutilier, Bo Dai
arXiv: https://arxiv.org/abs/2603.06981

Multi-head Transformers Provably Learn Symbolic Multi-step Reasoning via Gradient Descent
NeurIPS 2025
Authors: **Tong Yang**, Yu Huang, Yingbin Liang Yuejie Chi
arXiv: https://arxiv.org/abs/2508.08222
We aim to provide mechanistic understandings of how transformers acquire on the chain-of-thought multi-step reasoning of the Transformer by theoretically analyzing the Transformer's training dynamics and generalization on symbolic reasoning tasks on trees.

Exploration from a Primal-Dual Lens: Value-Incentivized Actor-Critic Methods for Sample-Efficient Online RL
NeurIPS 2025
Authors: **Tong Yang**, Bo Dai, Lin Xiao, Yuejie Chi
arXiv: https://arxiv.org/abs/2506.22401

Motivated by recent online reinforcement learning developments in exploration via optimistic regularization, we propose a provable efficient and practical actor-critic method to balance the fundamental trade-off between exploration and exploitation.

Incentivize without Bonus: Provably Efficient Model-based Online Multi-agent RL for Markov Games

ICML, 2025

Authors: **Tong Yang**, Bo Dai, Lin Xiao, Yuejie Chi

arXiv: https://arxiv.org/abs/2502.09780

In this paper, we propose a novel model-based algorithm, called VMG, that incentivizes exploration via biasing the empirical estimate of the model parameters towards those with a higher collective best-response values of all the players when fixing the other players' policies, thus encouraging the policy to deviate from its current equilibrium for more exploration.

Faster WIND: Accelerating Iterative Best-of-N Distillation for LLM Alignment

AISTATS 2025

Authors: **Tong Yang**, Jincheng Mei, Hanjun Dai, Zixin Wen, Shicong Cen, Dale Schuurmans, Yuejie Chi, Bo Dai

arXiv: https://arxiv.org/abs/2410.20727

Recent advances in aligning large language models with human preferences have corroborated the growing importance of best-of-N distillation (BOND). We establish a novel framework, WIN rate Dominance (WIND), with a series of efficient algorithms for regularized win rate dominance optimization that approximates iterative BOND.

In-Context Learning with Representations: Contextual Generalization of Trained Transformers

Conference: *NeurIPS 2024*

Authors: **Tong Yang**, Yu Huang, Yingbin Liang Yuejie Chi

arXiv: https://arxiv.org/abs/2408.10147

This work aims to provide theoretical understandings on in-context learning ability of transformers by investigating the training dynamics of transformers through the lens of non-linear regression tasks.

Federated Natural Policy Gradient and Actor Critic Methods for Multi-task Reinforcement Learning

Conference: *NeurIPS 2024*

Authors: **Tong Yang**, Yuejie Chi, Shicong Cen, Yuting Wei, Yuxin Chen

arXiv: https://arxiv.org/abs/2311.00201

Federated reinforcement learning enables collaborative decision making of multiple distributed agents without sharing local data trajectories. In this work, we consider a multi-task setting where the agents target at learning a globally optimal policy that maximizes the sum of the discounted total rewards of all the agents in a decentralized manner, where each agent only communicates with its neighbors over some prescribed graph topology.

A Primal-Dual Approach to Solving Variational Inequalities with General Constraints

Conference: *ICLR 2024*

Authors: Tatjana Chavdarova*, **Tong Yang***, Matteo Pagliardini, Michael I. Jordan (* alphabetical)

arXiv: https://arxiv.org/abs/2210.15659

In this work, we improve ACVI—a first-order gradient methods to solve general variational inequalities (VIs)—by circumventing its limiting assumption that analytic solutions of specific subproblems are available.

Solving Constrained Variational Inequalities via a First-order Interior Point-based Method

Conference: *ICLR 2023 (spotlight)*

Authors: **Tong Yang***, Michael I. Jordan*, Tatjana Chavdarova*

arXiv: https://arxiv.org/abs/2206.10575

Inspired by the efficacy of the alternating direction method of multipliers (ADMM) method in the single-objective context, we generalize ADMM to derive a first-order interior-type method for constrained variational inequality problems.

Optimization for Amortized Inverse Problems
Conference: *ICML 2023*
Authors: Tianci Liu*, **Tong Yang*** Quan Zhang, Qi Lei
arXiv: https://arxiv.org/abs/2210.13983
In this paper, we propose an efficient amortized optimization scheme for inverse problems with a deep generative prior. Specifically, the optimization task with high degrees of difficulty is decomposed into optimizing a sequence of much easier ones.

Value-Incentivized Preference Optimization: A Unified Approach to Online and Offline RLHF
ICLR 2025
Authors: Shicong Cen, Jincheng Mei, Katayoon Goshvadi, Hanjun Dai, **Tong Yang**, Sherry Yang, Dale Schuurmans, Yuejie Chi, Bo Dai
arXiv: https://arxiv.org/abs/2405.19320
Reinforcement learning from human feedback (RLHF) has demonstrated great promise in aligning large language models (LLMs) with human preference. In this paper, we introduce a unified approach to online and offline RLHF – value-incentivized preference optimization (VPO).

Scholarships & Awards

- Gold Reviewer Award of ICML 2026.
- Wei Shen and Xuehong Zhang Presidential Fellowship, 2025.
- Outstanding Student of Xi'an Jiaotong University (2%), 2018-2019, 2019-2020, 2020-2021.
- National Encouragement Scholarship (3%), 2018-2019.
- HIWIN Scholarship (1%), 2019-2020.
- ZhuFeng Scholarship (1%), 2018-2019,2019-2020.
- National Scholarship (1%), 2020-2021.