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

Publications

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). However, the iterative BOND algorithm is prohibitively expensive in practice due to the sample and computation inefficiency. This paper addresses the problem by revealing a unified game-theoretic connection between iterative BOND and self-play alignment, which unifies seemingly disparate algorithmic paradigms.
- Based on the established connection, 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 the parameter space. We provides provable sample efficiency guarantee for one of the WIND variant with the square loss objective. The experimental results confirm that our algorithm not only accelerates the computation, but also achieves superior sample efficiency compared to existing methods.

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 paper investigates the training dynamics of transformers by gradient descent through the lens of non-linear regression tasks. The contextual generalization here can be attained via learning the template function for each task in-context, where all template functions lie in a linear space with m basis functions. We analyze the training dynamics of one-layer multi-head transformers to in-contextly predict unlabeled inputs given partially labeled prompts, where the labels contain Gaussian noise and the number of examples in each prompt are not sufficient to determine the template.

- Under mild assumptions, we show that the training loss for a one-layer multi-head transformer converges linearly to a global minimum. Moreover, the transformer effectively learns to perform ridge regression over the basis functions. To our knowledge, this study is the first provable demonstration that transformers can learn contextual (i.e., template) information to generalize to both unseen examples and tasks when prompts contain only a small number of query-answer pairs.

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

- We develop federated vanilla and entropy-regularized natural policy gradient (NPG) methods in the tabular setting under softmax parameterization, where gradient tracking is applied to estimate the global Q-function to mitigate the impact of imperfect information sharing. We establish non-asymptotic global convergence guarantees under exact policy evaluation, where the rates are nearly independent of the size of the state-action space and illuminate the impacts of network size and connectivity. To the best of our knowledge, this is the first time that near dimension-free global convergence is established for federated multi-task RL using policy optimization.
- We further go beyond the tabular setting by proposing a federated natural actor critic (NAC) method for multi-task RL with function approximation, and establish its finite-time sample complexity taking the errors of function approximation into account.

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

- We develop an interior-point approach to solve constrained variational inequality (cVI) problems. 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 method for cVIs, that we refer to as ADMM-based interior-point method for constrained VIs (ACVI). We provide convergence guarantees for ACVI in two general classes of problems: (i) when the operator is ξ-monotone, and (ii) when it is monotone, some constraints are active and the game is not purely rotational. When the operator is, in addition, L-Lipschitz for the latter case, we match known lower bounds on rates for the gap function of $\mathcal{O}(1/\sqrt{K})$ and $\mathcal{O}(1/K)$ for the last and average iterate, respectively.
- To the best of our knowledge, this is the first presentation of a first-order interior-point method for the general cVI problem that has a global convergence guarantee. Moreover, unlike previous work in this setting, ACVI provides a means to solve cVIs when the constraints are nontrivial. Empirical analyses demonstrate clear advantages of ACVI over common first-order methods. In particular, (i) cyclical behavior is notably reduced as our methods approach the solution from the analytic center, and (ii) unlike projection-based methods that zigzag when near a constraint, ACVI efficiently handles the constraints.

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

- Yang et al. (2023) recently showed how to use first-order gradient methods to solve general variational inequalities (VIs) under a limiting assumption that analytic solutions of specific subproblems are available. In this paper, we circumvent this assumption via a warm-starting technique where we solve subproblems approximately and initialize variables with the approximate solution found at the previous iteration. We prove the convergence of this method and show that the gap function of the last iterate of the method decreases at a rate of $\mathcal{O}(1/\sqrt{K})$ when the operator is L-Lipschitz and monotone. In numerical experiments, we show that this technique can converge much faster than its exact counterpart.
- Furthermore, for the cases when the inequality constraints are simple, we introduce an alternative variant of ACVI and establish its convergence under the same conditions. Finally, we relax the smoothness assumptions in Yang et al., yielding, to our knowledge, the first convergence result for VIs with general constraints that does not rely on the assumption that the operator is L-Lipschitz.

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.
- We provide a theoretical guarantee of the proposed algorithm and empirically validate it on different inverse problems. As a result, our approach outperforms baseline methods qualitatively and quantitatively by a large margin.

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

- In this paper, we introduce a unified approach to online and offline RLHF – value-incentivized preference optimization (VPO) – which regularizes the maximum-likelihood estimate of the reward function with the corresponding value function, modulated by a sign to indicate whether the optimism or pessimism is chosen. VPO also directly optimizes the policy with implicit reward modeling, and therefore shares a simpler RLHF pipeline similar to direct preference optimization.
- Theoretical guarantees of VPO are provided for both online and offline settings, matching the rates of their standard RL counterparts. Moreover, experiments on text summarization and dialog verify the practicality and effectiveness of VPO.

Competition Experience

- The first prize of China Undergraduate Mathematical Contest in Modeling in Shaanxi division, 2019.
- The first prize of China Undergraduate Mathematical Contest in Modeling in Shaanxi division, 2020.
- Honorable Mention of Mathematical Contest In Modeling, 2020.
- The second prize of The Chinese Mathematics Competitions (Mathematics Group), 2019.
- The first prize of Campus Mathematical Contest in Modeling, 2019.
- The second prize of Campus Collegiate Programming Contest, 2021.

Scholarships & Awards

- 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.