ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

Tackling Multi-Agent Policy-Gradient Variance with Stochastic Variance-Reduced Estimators

Multi-Agent Reinforcement Learning (MARL) achieves strong results on cooperative control benchmarks, yet training remains unstable due to **high variance in multi-agent policy-gradient (MAPG) estimators**. Two principal contributors are **temporal sampling noise** (stochastic trajectories, non-stationary policies) and **credit-assignment ambiguity** (attributing a joint reward to individual agents). Counterfactual/optimal baselines (e.g., COMA/OB) mitigate inter-agent attribution variance, while **stochastic variance-reduced (SVR)** policy-gradient methods (e.g., SVRG/SVRPG) stabilize gradients across iterations. Their combined potential in MARL has not been systematically studied.

Project Objective. Credit-assignment baselines primarily target *inter-agent* variance; SVR estimators target *temporal* estimator variance. We will **systematically investigate a hybrid estimator** that integrates SVR with counterfactual/optimal baselines, and characterize when (and why) this combination improves stability and sample efficiency over standard MAPG baselines.

Interested? Please contact us for more details!

Contact

- Flint Xiaofeng Fan: xiafan@ethz.ch, ETZ G97

References