# Project Proposal: Zap Q-Learning and its applications in Arcade Learning Environment (ALE)

Xuyin Zhu; Muchen Li; Xuanzhao Dong

### I. MOTIVATION & INTRODUCTION

Learning value or Q function plays an integral role in Reinforcement learning. However, all learning control methods face a dilemma: they seek to learn action values conditional on subsequent optimal behavior based on the fact that they first need to behave sub-optimally in order to explore all actions and hence find the optimal ones. As a result, different ways to learn optimal policy while exploring lead to different learning methods. As opposed to the “on-policy” learning approach, which chooses to learn action value function based on near-optimal policy, “off-policy learning’’ consists of two separate policies, one (target policy) that is evaluated and improved towards the optimal policy and another one (behavior policy) that explores and generates behavior. In general, “off-policy learning’’ methods are of greater variance and slower to converge [4]. In an attempt to solve this problem, many improvements have been proposed such as Watkins’ algorithm with a “polynomial learning rate” [5] and Speedy Q-Learning algorithm[6].

As a special case based on the stochastic Newton-Raphson algorithm, Zap Q-learning chooses to use matrix gain techniques to speed up convergence of traditional Watkins’ Q-learning[3]. Furthermore, under two time-scale implementations, the result of Zap Q-learning algorithm will resemble and implementation of Newton-Raphson[3].

The goal of this project is to verify that Zap Q-learning converges faster than Deep Q-Networks (DQN) by comparing their performance on several Atari 2600 computer games. We plan to train agents using Zap Q-learning and DQN that can play Atari game [7] based on in the Arcade Learning Environment(ALE), a simple reinforcement learning testbed that allows researchers and hobbyists to develop AI agents for Atari 2600 games[1]. Now ALE has been included in OpenAI Gym and researchers can easily access the environment by importing a Python library.

### II. Method

We will use OpenAI Gym ATARI as the environment to train two agents, one by Zap Q-learning and the other by DQN. Both agents are trained to play 6 games in Atari: Breakout, Enduro, Pong, Q\*bert, Seaquest and Space Invaders. Normalizing scores[2] will be used to evaluate the performance of our learned policy across games. Training time and/or episodes to converge will be used to evaluate the speed of convergence.

Reference

1.[The Arcade Learning Environment: An Evaluation Platform For General Agents (Extended Abstract) (ijcai.org)](https://www.ijcai.org/Proceedings/15/Papers/585.pdf)

3.Adithya M.Devraj and Sean P.Meyn,’’Zap Q-Learning” ,2017.

[4] Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction Second edition.

[5] E. Even-Dar and Y. Mansour. Learning rates for Q-learning. *Journal of Machine Learning*

*Research*, 5(Dec):1–25, 2003.

[6] M. G. Azar, R. Munos, M. Ghavamzadeh, and H. Kappen. Speedy Q-learning. In *Advances in Neural Information Processing Systems*, 2011.

[7] Mnih, Volodymyr, et al. Playing Atari with reinforcement learning arXiv:1312.5602 (2013).