Notes on Variance Reduction Techniques for Gradient Estimates in Reinforcement Learning

compiled by D.Gueorguiev, 4/16/2025

# Introductory Notes

Policy gradient methods for reinforcement learning avoid some of the undesirable properties of the value function approaches, such as policy degradation. However, the variance of the performance gradient estimates obtained from the simulation is sometimes excessive. Two commonly used policy gradient techniques will be discussed in this document – the *baseline* and *actor-critic* methods.

## System Model

A partially observable Markov decision process (POMDP) can be modelled by a system consisting of a state space, , an action space, , and an observation space, , all of which will be considered finite here. State transitions are governed by a set of probability transition matrices , where , components of which will be denoted , where . There is also an observation process , where is the space of probability distributions over , and a reward function . Together these define the POMDP .

A policy for this POMDP is a mapping where denotes the space of all finite sequences of observations and is the space of probability distributions over . If only the set of reactive policies

# References

[1] [Variance Reduction Techniques for Gradient Estimates in Reinforcement Learning, Evan Greensmith, Peter L. Bartlett, Jonathan Baxter, JMLR, 2004](https://github.com/dimitarpg13/reinforcement_learning_and_game_theory/blob/main/articles/ReinforcementLearning/Variance_Reduction_Techniques_for_Gradient_Estimates_in_Reinforcement_Learning_Greensmith_2004.pdf)

[2] [Infinite-Horizon Policy-Gradient Estimation, J. Baxter, P. Bartlett, 2001](https://github.com/dimitarpg13/reinforcement_learning_and_game_theory/blob/main/articles/ReinforcementLearning/Infinite-Horizon_Policy-Gradient_Estimation_Baxter_2001.pdf)

[3] [Planning and acting in partially observable stochastic domains, LP Kaelbling et al, 1998](https://github.com/dimitarpg13/reinforcement_learning_and_game_theory/blob/main/articles/ReinforcementLearning/Planning_and_acting_in_partially_observable_stochastic_domains_aij98-pomdp.pdf)

# Appendix

## Markov Decision Processes and Partially Observable Markov Decision Processes

Markov Decision Processes serve as a basis for solving more complex partially observable problems.

**Definition**: *Markov Decision Process*