Algorithmic Complexity in Markov Decision Processes

Supervised by Bruno Gaujal (Polaris team, Inria and LIG, Grenoble)
Email: bruno.gaujal@inria.fr

Description

While the complexity of computing the optimal policy in discounted MDPs is rather well known (both Value iteration and Policy iteration have been investigated ([1] and [2] respectively), the case of undicounted MDPs is still open, in the following sense. The only known polynomial algorithm is based on linear programming and its complexity is very high. In practice, policy iteration and value iteration both converge fast, and they perform better than linear programming. The goal of the internship is to analyze the complexity of both of them in the discounted case and make some progress in their analysis in the undiscounted case. Another objective is to design a new efficient algorithm whose complexity (worse case and average case) is better than linear programming. The work in [3] may be a promising start.

This work could be extended to a three year research project carried in a PhD, for example by investigating smooth complexity issues.

Prerequisites

A taste for algorithmic complexity and modelisation.
Being knowledgeable in Markov decision processes, linear algebra and stochastic matrices will help.

Information and Contact

This internship duration is 5 to 6 months (M2 internship) and can start anytime in 2024. Contact Bruno Gaujal (bruno.gaujal@inria.fr) for more information.

Bibliography

[1] M. L. Puterman. Markov Decision Processes, Discrete Stochastic Programming, Wiley, 2005.

[2] Bruno Scherrer. Improved and generalized upper bounds on the complexity of policy iteration. Mathematics of Operations Research, 2016, 41 (3), pp.758-774.

[3] V. Boone. When do discounted-optimal policies also optimize the gain? https://arxiv.org/abs/2304.08048