

# Data-driven optimal control via linear programming

A. Martinelli and M. Gargiani

SA/MA Project

## Abstract

Learning how to optimally regulate a dynamical system from data is a fundamental problem in control theory. This project focuses on investigating new theory and methods about the so-called linear programming approach to approximate dynamic programming.

## 1 Introduction

The term “optimal control” came into use in the 1950s to describe the problem of designing a controller to minimize a measure of a dynamical system’s behavior over time. The problem formulation is widely applicable and it arises in many disciplines, from robotics to bioengineering to finance, to name a few. In the same years, R. Bellman developed a method, based on the principle of optimality, that uses the concept of value function to define a functional equation – the Bellman equation. The class of methods for solving optimal control problems by solving this equation came to be known as dynamic programming (DP). For computing the value function, DP methods typically rely on three fundamental approaches: value iteration, policy iteration and linear programming (LP) [2]. The mathematical foundations of these approaches lie in the monotonicity and contractivity properties of a functional operator, implicitly defined in the Bellman equation. DP methods suffer from what Bellman called the “curse of dimensionality”, meaning that the computational requirements grow exponentially with the number of state and input variables. Several approximation methods have been developed to mitigate this curse of dimensionality, collectively known as approximate dynamic programming (ADP).

In this context, our work focuses on the LP approach to ADP. The LP approach, introduced in the 60s, exploits the properties of the Bellman operator to build linear programs whose solution coincides with the optimal value function. In the context of ADP with continuous state or action spaces, the exact infinite dimensional LPs are approximated by tractable finite dimensional ones [4].

In many real-world applications one has to operate a system without the knowledge of its dynamical equation. Learning how to optimally regulate a system is possible by letting it interact with the environment and collecting costs (or rewards) over time, an approach classically known as reinforcement learning (RL) (see Fig. 1). To extend RL methods in the LP approach framework, one can reformulate the Bellman equation in terms of the  $q$ -function, and set up a new class of LPs based on the Bellman operator for  $q$ -functions [3]. The main advantage of the  $q$ -function formulation is that one can extract a policy directly from  $q$ , without knowledge of the system dynamics or stage cost. The problem of learning the optimal  $q$ -function from data with the LP approach has been recently addressed for both deterministic [1] and stochastic [5] systems.

The LP approach has a long history in the theory of ADP. When it comes to computation, however, the LP approach often suffers from poor scalability. In [5], we introduce a relaxed version of the Bellman operator for  $q$ -functions and prove that it is still a monotone contraction mapping with a

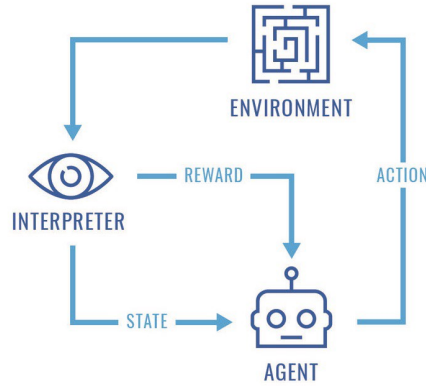


Figure 1: Classical scheme of a reinforcement learning process.

unique fixed point. In the spirit of the LP approach, we exploit the new operator to build a relaxed linear program (RLP). Compared to the standard LP formulation, our RLP has only one family of constraints and half the decision variables, making it more scalable and computationally efficient. For deterministic systems, the RLP trivially returns the correct  $q$ -function. For stochastic linear systems in continuous spaces, the solution to the RLP preserves the minimizer of the optimal  $q$ -function, hence retrieves the optimal policy. Theoretical results are backed up in simulation where we solve sampled versions of the LPs with data collected by interacting with the environment. For general nonlinear systems, we observe that the RLP again tends to preserve the minimizers of the solution to the LP, though the relative performance is influenced by the specific geometry of the problem.

## 2 Goals

1. Understanding fundamental concepts in the theory of ADP and its methods [2, 4].
2. Focus on the linear programming approach, in particular on its relaxed formulation [5] and iterative implementation [1].
3. Theory: extend the results in [5] by relaxing the linear quadratic assumptions in the fixed point analysis to, e.g., linear affine dynamics and generalized quadratic costs.
4. Theory: extend the approach in [1] to stochastic systems by exploiting the formulations in [3, 5].
5. Coding/implementation: convergence and performance analysis of the iterative and non-iterative version of the LP approach. Comparison with other existing ADP techniques, such as approximate policy iteration and value iteration.
6. Theory (advanced): derive performance bounds for the RLP in [5] that link solution quality and online performance, on the line of [3].

## 3 Project information

The project can be carried out in a more theoretical perspective, focusing on points 3 and 4, in a more applied perspective, focusing on point 5, or both, depending on the student's personal preferences.

**Qualifications:** We are looking for a highly motivated student with background in systems & control

and optimization. Sufficient mathematical maturity is required, Matlab/Python knowledge is necessary to develop point 5.

**Corona disclaimer:** This project can be done in person at the Automatic Control Laboratory, hybrid, or completely remotely, depending on the current ETH regulations. Most importantly, we can change between these forms whenever needed.

**Project publication:** If the results are promising they can be turned into a publication.

**How to apply:** Please send your CV and transcript of records to A. Martinelli ([andremar@ethz.ch](mailto:andremar@ethz.ch)) and M. Gargiani ([gmatilde@ethz.ch](mailto:gmatilde@ethz.ch)).

## References

- [1] G. Banjac and J. Lygeros. A data-driven policy iteration scheme based on linear programming. In *58th IEEE Conference on Decision and Control*, pages 816–821, 2019.
- [2] D.P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. II*. Athena Scientific, 3rd edition, 2007.
- [3] P.N. Beuchat, A. Georghiou, and J. Lygeros. Performance guarantees for model-based approximate dynamic programming in continuous spaces. *IEEE Transactions on Automatic Control*, 65(1):143–158, 2020.
- [4] D.P. de Farias and B. Van Roy. The linear programming approach to approximate dynamic programming. *Operations Research*, 51(6):850–865, 2003.
- [5] A. Martinelli, M. Gargiani, and J. Lygeros. Data-driven optimal control with a relaxed linear program. *ArXiv preprint arXiv2003.08721*, 2020.