Reinforcement learning for optimal control problems

Anton Forsell
Department of Automatic Control
Lund University
Lund, Sweden
an5153fo-s

I. What is reinforcement learning and optimal control

Reinforcement learning (RL) is an interdisciplinary area of machine learning (ML) which in turn is a field of study in artificial intelligence (AI). AI, in its broadest sense, is intelligence exhibited by machines or computers. ML refers to algorithms or methods that can learn from data and generalize to unseen data and thus perform tasks without explicit instructions. RL is a set of algorithms where an intelligent agent (the machine) decides on actions in a dynamic environment to maximize the cumulative reward (learns from data); which is where the name comes from, the agent gets rewarded for good decisions.

We are interested in this short paper on how reinforcement learning methods can be applied to optimal control problems. Optimal control problems refer to finding an "optimal" (there are numerous potential objectives) solution to a control problem; control problems in turn refer to solving issues such as automation, anti-lock braking systems (ABS), and stabilization. This is generally done by a controller which supplies a control signal telling parts of the system such as the actuator and motors what to do.

II. How can reinforcement learning be applied to optimal control

a) Finding optimality: To find the optimality we will use a temporal difference (TD) learning method. A desired property of TD learning is its ability to learn from experiences, without needing a model or description of the problem. This allows us to create a method where we start with a policy and with the gathered data from the system, we can compute a better policy with each iteration. Additionally, data can be gathered before the optimization step allowing for flexible scheduling.

b) Gathering of data: TD learning learns from experiences or data, therefore there is a heavy emphasis on how the data is gathered and that the data sufficiently covers numerous possible situations the system may be in. We will discuss a method of gathering data that utilizes noise to allow a greater exploration of the system. To gather data we will run the system and sample or measure: what state or position the system currently is in, how fast the system is changing, and what control signal was used to get to this state or position. To help diversify the data gathered, we will add stochastic noise (random disturbances) to the control signal. This will result in the system occupying many different states before it reaches equilibrium, generating more data than one without noise.

III. Experiments and results

The method was tried on a five small examples. The initial system was a double integrator with friction. From there it was tested on three more linear systems with signals space and state space of a higher dimensionality. The last experiment was on a non-linear system to highlight the algorithms capabilities.

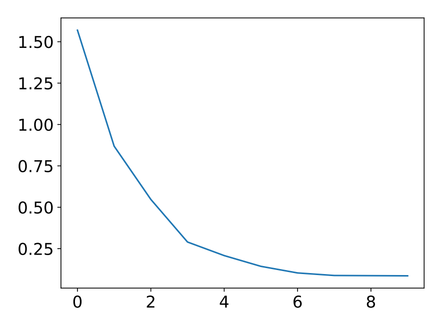

Fig. 1: Example of the algorithms error decreasing for each of the 10 iterations.

Figure. 1, highlights the results from an experiment where the control signal and state space were vector-valued. With each of the 10 iterations, we generate a solution/policy that is better than the previous one.

IV. Conclusion

The results show that the highlighted algorithm produces policies of increasing quality with each iteration, in many cases reaching optimality. The methods model-free approach and ability to find solutions to nonlinear problems, allows it to be used on many different types of problem early in the design process. Creating a valuable tool for solving optimal control problems.