MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS
Learning and Dynamical Systems Group (Dr. Michael Muehlebach)

Master Thesis Project

Can reinforcement learning algorithms learn piecewise linear policies?

The project is motivated by the recent success of reinforcement learning on tasks such as playing the game of Go, mastering several different Atari games, and manipulating objects. We ask the question whether reinforcement learning algorithms, such as deep Q-learning, PILCO, or policy gradient approaches can also solve simple linear problems?
Concretely, the aim will be to evaluate these methods on the constrained linear quadratic regulator problem. The constrained linear quadratic regulator problem is a fundamental problem in control theory and consists of finding the feedback policy ϕ that minimizes a quadratic performance objective, subject to linear dynamics, and input and state constraints. In the absence of constraints, the problem can be stated as

$$\begin{aligned} \inf_{\phi\in\mathcal{F}} \quad & \lim_{T\to\infty} \frac{1}{T}\mathrm{E}[\sum_{k=0}^{T} x_k^{\mathrm{T}} Q x_k + u_k^{\mathrm{T}} R u_k], \\ \text{s.t.} \quad & x_{k+1} = Ax_k + Bu_k + v_k, \qquad y_k = Cx_k + w_k, \\ & u_k = \phi(y_k, y_{k-1}, \dots, y_0), \end{aligned}$$

where the positive definite matrices Q, R define the cost, the matrices A, B, C the dynamics, and x_0, v_k, w_k are mutually independent Gaussian random variables with zero mean. The important feature of these problems is that the optimal feedback policy ϕ can be computed in closed form, provided that A, B, C, Q, R and the covariance matrices of v_k, w_k are known. This provides us with a benchmark against which we can evaluate reinforcement learning algorithms.
Depending on time, the project can be extended in several directions. For example, one could construct simple nonlinear benchmarks, in such a way that the optimal policy is again known ahead of time (for example systems that are feedback linearizable/differentially flat). Since a reinforcement learning problem is at its heart an optimization problem over feedback policies, the project could also experiment with gradient-free optimization algorithms or inexact gradient-based algorithms.

Learning and Dynamical Systems Group
The Learning and Dynamical Systems Group is part of the Max Planck Institute for Intelligent Systems in Tübingen, Germany. Our research lies at the intersection between machine learning, dynamical systems, and mathematical optimization. Further information can be found here: https://lds.is.mpg.de/ .

Prerequisites
Strong analytical skills and programming experience (Python, MATLAB, C/C++ or similar). Background in machine learning, control theory, statistics, or mathematical optimization is a plus.

Contact
If you have any questions do not hesitate to contact us. When applying for a project, please include your CV, bachelor's and master's transcripts, and a one-page letter of motivation describing your research interests and educational background.

Dr. Michael Muehlebach, michael.muehlebach@tuebingen.mpg.de