Reinforcement Learning for Control

Dates and times

05-04-2027 - Jilles Dibangoye
12-04-2027 - Jilles Dibangoy
19-04-2027 - Saeed Ahmed
10-05-2027 - Saeed Ahmed
from 13.45-16.00 hrs

Course location

Cursus- en Vergadercentrum Domstad, Koningsbergerstraat 9, 3531 AJ Utrecht

ECTS

3 ECTS if the homework is completed successfully

1 ECTS for auditing the entire course

Lecturers

Jilles Dibangoye
Saeed Ahmed

Objective

This course provides a control-oriented journey through reinforcement learning. It introduces the formal framework of Markov Decision Processes, value functions, policies, Bellman equations, and the main algorithmic families of reinforcement learning for control. In this course, a policy is interpreted as a feedback controller: it maps the current state or observation of a dynamical system to a control action. Rather than treating each algorithm as an isolated recipe, the course emphasizes the common templates behind value-based, policy-based, actor–critic, and model-based methods. Representative algorithms such as Monte Carlo methods, temporal-difference learning, SARSA, Q-learning, policy gradients, DDPG, TD3, SAC, PPO, and model-based rollouts are used to illustrate these templates. Participants then explore additional variants through exercises and hands-on work. The course culminates in a deep reinforcement learning case study for UAV control, showing how RL theory can be translated into policies that are trained, evaluated, and refined for dynamical systems.

Contents

Lecture	Lecture content	Short description
Lecture 1 05-04, Afternoon, Jilles	**Formal and Algorithmic Foundations of Value-Based Reinforcement Learning for Control**	This lecture introduces reinforcement learning as a framework for learning policies through interaction with a dynamical system. We first present the control-oriented RL setting: agent, environment, state, action, reward, trajectory, return, and policy. In this setting, the policy plays the role of a feedback controller. We then formalize sequential decision-making problems as Markov Decision Processes, and introduce value functions, action-value functions, Bellman expectation equations, and Bellman optimality equations. The second part develops the

Lecture	Lecture content	Short description
		algorithmic foundations of value-based reinforcement learning. The emphasis is on the general algorithmic template of this family: value estimation from sampled experience, policy evaluation, policy improvement, and generalized policy iteration. Monte Carlo and Temporal-Difference learning are introduced as the main mechanisms for estimating value functions. We then study a small number of representative examples, including SARSA as an on-policy TD control method and Q-learning as an off-policy TD control method. Other value-based variants are left for guided exercises, where participants explore how the same template leads to related algorithms. The lecture concludes by explaining why tabular value-based methods do not scale directly to large or continuous state spaces, and how function approximation and neural networks lead to deep value-based reinforcement learning while introducing new stability issues.
Lecture 2 12-04, Afternoon, Jilles	**Policy-Based, Actor–Critic, and Model-Based Reinforcement Learning for Control**	This lecture studies the main alternatives and complements to value-based reinforcement learning. We begin by explaining the limitations of value-based methods in large or continuous action spaces, especially the difficulty of greedy maximization over continuous actions. This motivates policy-gradient methods, where a parameterized policy is optimized directly. We introduce stochastic policies, performance objectives, the policy-gradient theorem, and the use of baselines for variance reduction. We then present actor–critic methods, in which the actor represents the policy and the critic estimates value or advantage information used to improve it. This provides the conceptual bridge to modern deep reinforcement learning methods for continuous control. We position DDPG, TD3, SAC, and PPO within this taxonomy, highlighting deterministic policy gradients, twin critics and delayed updates, maximum-entropy actor–critic learning, and clipped policy optimization. The lecture concludes with model-based reinforcement learning, where a given or learned model is used for planning, simulated rollouts, or policy improvement, and with its connection to classical control and model predictive control.
Lecture 3 19-04, Afternoon, Saeed	**Reinforcement Learning and Control: A Unified Feedback Perspective**	This lecture reinterprets the reinforcement-learning framework from the viewpoint of feedback control and closed-loop system design. Building on the formal and algorithmic foundations introduced in Lectures 1 and 2, it explains how RL and control theory address the same fundamental problem: choosing actions over time to regulate the behaviour of a dynamical system. The lecture connects key RL objects to control concepts: the policy as a feedback controller, the reward as a performance index, the value function as a cost-to-go function, and learning as data-driven controller improvement. It then discusses the relation between RL, optimal control, adaptive control,

Lecture	Lecture content	Short description
		dynamic programming, or model predictive control. Special attention is given to control-oriented issues that are often hidden in algorithmic RL presentations, including stability, robustness, constraints, actuator saturation, safety, excitation, and closed-loop performance evaluation.
Lecture 4 10-05, Afternoon, Saeed	**RL-Control Design in Practice: UAVs and Intelligent Transportation Systems**	This lecture shows how the RL-control connection is used in practical cyber-physical systems. The main case study focuses on UAV control, where an RL policy is designed as a feedback controller for altitude, attitude, or trajectory tracking under nonlinear dynamics, actuator limits, disturbances, and safety constraints. The lecture walks through the formulation of the control problem, the definition of observations and actions, reward design, training, closed-loop evaluation, and comparison with classical controllers. The second part briefly discusses how the same design principles extend to intelligent transportation systems, including vehicle platooning, lane-changing decisions, traffic signal control, and energy management. Across these examples, the emphasis is on identifying when RL is beneficial, when it is risky, and how learned policies should be evaluated before deployment in real control applications.

Course materials

- *Reinforcement Learning: An Introduction* – Sutton & Barto (required)
- Recommended
  - *Dynamic Programming and Optimal Control* – Bertsekas
  - *Algorithms for Reinforcement Learning* – Szepesvari
- Recht, B. (2019). A tour of reinforcement learning: The view from continuous control. Annual Review of Control, Robotics, and Autonomous Systems, 2, 253–279.
- Online resources: RL Course by David Silver, Spinning Up by OpenAI, etc.

Prerequisites

- Mathematics: Linear Algebra (required), Statistics and Probability (preferred), Optimization (preferred)
- Linear Control Theory
- Basics of Machine Learning (Neural Networks)
- Programming: MATLAB (required), Python (optional)

Homework assignments

Homework 1 — Foundations and Algorithmic Templates of Reinforcement Learning
This homework supports Lectures 1 and 2. Students practise the formal and algorithmic foundations of reinforcement learning for control, including MDP formulation, Bellman equations, value functions, policy evaluation and improvement, Monte Carlo and temporal-difference learning, and the comparison between value-based, policy-based, actor–critic, and model-based reinforcement learning. Core exercises focus on representative algorithms such as SARSA and Q-learning, while other variants are explored through shorter conceptual or guided exercises.

Homework 2 — Reinforcement Learning and Control-Oriented Case Studies
This homework supports Lectures 3 and 4. Students apply reinforcement learning concepts from a control perspective by formulating a dynamical system as an RL problem, defining observations, actions, rewards,

constraints, and performance measures. Exercises address the connection between RL and feedback control, stability and robustness considerations, reward design, comparison with classical controllers, and application-oriented case studies such as UAV control and intelligent transportation systems.