CSE 691 COURSE DESCRIPTION FOR SPRING SEMESTER 2022 (3 Credits)

TOPICS IN REINFORCEMENT LEARNING

ASU/SCAI

Prof. Dimitri P. Bertsekas

Wednesdays 4:30 ASU time (6:30 PM Boston time) to 7:00 ASU time (9:00 PM Boston time)

This course will focus on Reinforcement Learning (RL), a currently very active subfield of artificial intelligence, and it will discuss selectively a number of algorithmic topics couched on approximate Dynamic Programming (DP) methods: approximation in value and policy space, approximate policy iteration, rollout (a one-time form of policy iteration), model predictive control, multiagent methods, applications to challenging combinatorial optimization problems, implementations using simulation and neural network architectures, policy gradient methods, aggregation, and engineering and artificial intelligence applications, such as the high-profile successes of the AlphaZero and TD-Gammon programs, which play chess and backgammon, respectively.

One of our principal aims is to propose and develop a new conceptual framework for RL and approximate DP. This framework centers around two algorithms, which are designed largely independently of each other and operate in synergy through the powerful mechanism of Newton's method. We call these the *off-line training and the on-line play algorithms*; the names are borrowed from some of the major successes of RL involving games, such as AlphaZero and TD-Gammon. In the context of these programs, the off-line training algorithm is the method used to teach the program how to evaluate positions and to generate good moves at any given position, while the on-line play algorithm is the method used to play in real time against human or computer opponents.

One of our principal aims is to show, through the algorithmic ideas of Newton's method and the unifying principles of abstract DP, that the AlphaZero and TD-Gammon methodology of approximation in value space and rollout applies very broadly to deterministic and stochastic optimal control problems, involving both discrete and continuous search spaces, as well as finite and infinite horizon. Moreover, we will show that our conceptual framework can be effectively integrated with other important methodologies such as model predictive and adaptive control, multiagent systems and decentralized control, discrete and Bayesian optimization, and heuristic algorithms for discrete optimization.

The primary emphasis of the course is to encourage graduate research in reinforcement learning through directed reading and interactions with the instructor over zoom. Prerequisites are a full course on calculus, and a first course in probability.

The course will leverage a series of videolectures, slides, and other material from previous ASU offerings of the course, which are posted at

http://web.mit.edu/dimitrib/www/RLbook.html

All class material (videolecture recordings, slides, class notes, etc) will be posted at the same website on Thursdays, as well as at ASU/Canvas.

Textbooks (optional):

(1) D. Bertsekas, "Reinforcement Learning and Optimal Control," Athena Scientific, 2019.
(2) D. Bertsekas, "Rollout, Policy Iteration, and Distributed Reinforcement Learning," Athena Scientific, 2020.
(3) D. Bertsekas, "Abstract Dynamic Programming," 3rd Edition, Athena Scientific, 2022 (on-line).
(4) D. Bertsekas, "Lessons from AlphaZero from Optimal, Model Predictive, and Adaptive Control," Athena Scientific, to appear in 2022 (an on-line draft is available).
(5) D. Bertsekas, On-Line Class Notes.

Supplementary material:

(1) Sutton, R., and Barto, A., "Reinforcement Learning," 2nd Edition, MIT Press, Cambridge, MA (on-line; a valuable resource that approaches the subject from the AI point of view).

(2) Research papers and monographs, available on-line.

Structure:
One 2-hour lecture per week by the instructor, except for the last lecture, which will involve research presentations by student participants. Three-four homeworks (30 percent of the grade), and a research project or term paper (70 precent of the grade).

Topics to be covered:
(1) Introduction to exact and approximate dynamic programming
(2) Approximation in value and policy space
(3) Off-line training, on-line play, and Newton's method
(4) Rollout, approximate policy iteration, and Newton's method
(5) Model predictive and adaptive control
(6) Multiagent and multiprocessor reinforcement learning
(7) Training of feature-based approximation architectures and neural networks
(8) Policy networks and approximation in policy space
(9) Aggregation and other problem approximation architectures
(10) Applications in engineering, artificial intelligence, and discrete optimization

The first four lectures will aim to provide an introduction and overview of the subject, which will facilitate selecting and focusing on some research area. The remaining lectures will develop the topics listed above in greater depth.