

332:515 FINAL TERM PAPER or PROJECT – Fall 2024

The final term paper or project (worth 25% of the course grade) will be due on the last day of the final exams. The topic of the term paper may be selected by the students so that the paper is beneficial to your study and your future research. **Each graduate student must prepare an individual report.** However, the students may work on the same topic for their term papers. **The undergraduate students will work in teams, 2-3 students per team, and submit one report per team.**

1) The final **term paper** will be based on a journal paper (or papers) related to the material covered in class starting. The list of potential topics is presented at the end of this document.

2) The **term paper** may be also on any of chapter 7-13 from the Sutton and Barto textbook, the chapters that we did not cover in the course (this might be appropriate for the computer science and ECE undergraduate students).

3) The **term paper** may be based on an overview RL paper (or papers) that were tacitly mentioned in the lectures. These papers are posted on Canvas, and listed at the end of this document.

4) The **term paper** may be based on any topic of your interest on RL material related to the topics covered in this class, assuming it is discussed and approved by Professor Gajic.

5) The **final project** will be an extension of Project1, to design reinforcers for linear-quadratic zero-sum differential games and Nash (non-zero sum) differential games. The project formulation is presented at the end of this document.

In general, the final term paper should be based on any reinforcement learning topic relevant to this course and/or approved by Professor Gajic, unless it is in the list of topics given in the follow-up. The topics are supposed to be beyond the material covered in the course. **The paper should be typed and its pdf file uploaded on Canvas by the last day of the final exams for Fall semester of 2024.**

The paper must contain all parts of a standard conference/journal paper:

- **Abstract;**
- **Introduction including the relationship to the material covered in this course;**
- **Methods and/or algorithms developed;**
- **Discussion of analytical results obtained;**
- **Discussion of numerical results obtained (if any);**
- **Conclusions;**
- **References.**

Potential Topics for the Final Term Paper

- 1) **Applications of Nash Differential Games to Aerospace.** Following the theory of policy iterations for Nash differential games consider the linear-quadratic (LQ) Nash differential game problem for attitude takeover control of failed space craft, as presented in the paper:
Y. Chai, J. Luo, N. Han, and J. Xie, "Linear differential game approach for attitude takeover control of failed spacecraft," *Acta Astronautica*, Vol. 175, 142-154, 2020.
X. Lang and A. de Ruiter, "Non-cooperative differential game based output feedback control for spacecraft attitude regulation," *Acta Astronautica*, vol. 193, 370-380, 2022.
Provide a detailed review of the papers with emphases on the policy iterations for the N -agent LQ Nash differential game problem.
- 2) **Policy Iterations in Affine Nonlinear Nash Differential Games.** Using the knowledge that we got in this course about approximate dynamic programming for affine nonlinear systems, present in detail policy iterations for affine nonlinear Nash games by mostly following the paper:
K. Vamvoudakis and F. Lewis, "Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton-Jacobi equations," *Automatica*, Vol. 47, 1556-1569, 2011.
- 3) **Nash Games with Unknown Dynamics.** The interesting papers on the topic of solving the Nash games on-line when the system model is not available are:
D. Vrabie and F. Lewis, "Integral Reinforcement Learning for Finding Online the Feedback Nash Equilibrium of Nonzero-Sum Differential Games," 313-330, Chapter 17 in *Advances in Reinforcement Learning*, A. Mellouk (ed.), IntechChina, 2012.
K. Vamvoudakis, "Non-zero sum Nash Q-learning for unknown deterministic continuous-time linear systems," *Automatica*, Vol. 61, 274-281, 2015.
- 4) **Online RL for Affine Nonlinear Systems.** In Lecture 20 we presented RL for partially and completely model free systems (learning on line, learning from data interacting with the system) using the linear-quadratic optimal control problem formulation. An extension to non-linear (affine) systems is presented in Vrabie and Lewis (2012) in the book chapter that can be used as a term paper topic.
D. Vrabie and F. Lewis, "Online Adaptive Optimal Control Based on Reinforcement Learning," in *Optimization and Optimal Control*, A. Chinchuluun et al. (eds.), Springer, 2010.
J. Murray, C. Cox, G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Transactions on Systems, Man, and Cybernetics- Part C: Applications and Reviews*, Vol. 32, 140-153, 2002.
- 5) **RL for Graphical Games.** Several papers by Lewis and his coworkers considered RL for graphical games (interactions among the agents are constrained to a fixed strongly connected graph with the agent dynamics represented by independent linear systems):
F. Lewis, H. Zhang, Hengster-Movric, and A. Das, "Graphical Games: Distributed Multiplayer Games on Graphs," pages 181-217, in *Cooperative Control of Multi-Agent Systems: Optimal and Adaptive Design Approaches*, Springer, 2014.
M. Abouheaf, F. Lewis, K. Vamvoudakis, and S. Haesaert, and R. Babushka, "Multi-agent discrete time graphical games and reinforcement learning solutions," *Automatica*, Vol. 50, 3038-3053, 2014;
M.I. Abouheaf, F.L. Lewis, M. S. Mahmoud, and D. G. Mikulski, "Discrete-time dynamic graphical games: model-free reinforcement learning solution," *Control Theory and Technology*, Vol. 13, 55-69, 2015.

- 6) **Discrete-Time Zero-Sum Games.** In Lecture 21, we presented the continuous-time zero sum games. RL for discrete-time LQ games was considered in the paper:
A. Al-Tamimi, F. Lewis, and M. Abu-Khalaf, "Model-free Q-learning for linear discrete-time zero-sum games with applications to H-infinity control," *Automatica*, Vol. 43, 473-481, 2007.
An aircraft example is presented in this paper.
- 7) **Discrete-Time Nash Games.** In Lecture 22, we presented the continuous-time zero sum games. RL for discrete-time LQ Nash games was considered in the paper:
Z. Zhang, J. Xu, and M. Fu, "Q-learning for feedback Nash strategy of finite-horizon nonzero-sum difference games," *IEEE Transactions on Cybernetics*, vol. 52, 9170-978, 2022.
- 8) **RL for Electric Cars (PEM Fuel Cells):** The following two papers present the use of RL for air-fuel sensors control of PEM (proton exchange membrane) fuel cells used for electric cars:
M. Gheisarnejad, J. Boudjadar, and M. Khooban, "A new adaptive type-II fuzzy-based deep reinforcement learning control: Fuel cell and air-feed sensors control," *IEEE Sensors Journal*, Vol. 19, 9081-9089, 2019.
J. Li and T. Yu, "A new adaptive controller based on distributed reinforcement learning for PEMFC air supply system," *Energy Reports*, 1267-1279, 2021.
- 9) **RL for Affine Nonlinear Zero-Sum Differential Games.** We studied in Lecture 21 the LQ zero-sum games. The RL for nonlinear (affine) zero-sum games was presented in:
H. Zhang, Q. Wei, and D. Liu, "An iterative adaptive dynamic programming method for solving a class of nonlinear zero-sum games," *Automatica*, Vol. 47, 207-214, 2011.
Y. Zhu, D. Zhao, and X. Li, "Iterative adaptive dynamic programming for solving unknown nonlinear zero-sum game based on online data," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 28, 714-725, 2017.
- 10) **RL for Kalman Filtering.** Introduction to Kalman filter can be found in slides 15b (not covered in class lecture), Among the first attempts to connect RL and the Kalman filter (linear dynamic stochastic estimator) with potential applications to neuroscience was the paper:
I. Szita and A. Lorincz, "Kalman filter control embedded into the reinforcement learning framework," *Neural Computation*, Vol. 16, 491-499, 2004. Read also:
C. Tripp and R. Shachter, "Approximate Kalman filter Q-learning for continuous-state-space MDPs," Cornell University Archives, 2013.
X. Gao, H. Luo, B. Ning, F. Zhao, L. Bao, Y. Gong, Y. Xiao, and J. jiang, "RL-AKF: An adaptive Kalman filter navigation algorithm based on reinforcement learning for ground vehicles, Remote Sensing, 12, 1704; doi:10.3390/rs12111704, 2020.
The paper by Szita and Lorincz (2004) has the CS approach to RL (temporal difference, cost-to-go) and presents a SARSA algorithm.
- 11) **Comparison of the LQ Zero-Sum Game Algorithms.** Compare analytically and/or numerically the algorithms from Lectures 24a), 24c), and 24d): the modified Anderson *et al.* 2010 sequential algorithm derived by Vrabie and Lewis, simultaneous update algorithm of Wu and Luo , and Li and Gajic's algorithm.

- 12) **Reinforcement Learning for Markov Jump Linear Systems.** A good starting point to learn about this topic is the paper:

S. He, M. Zhang, H. Fang, F. Liu, X. Luan, and Z. Ding, "Reinforcement learning and adaptive optimization of a class of Markov jump systems with completely unknown dynamic information," *Neural Computing and Applications*, Vol.32, 14311-14320, 2020. [The paper was published within the subject: "Extreme Learning Machine and Deep learning Networks].

- 13) **ADP for Weakly Coupled Nonlinear Systems.** The following paper presents reinforcement learning for so called weakly coupled systems were extensively researched by Professor Gajic and his graduate students and colleagues:

L. Carrillo, K. Vamvoudakis, and J. Hespanha, "Approximate optimal adaptive control of weakly coupled nonlinear systems: A neuro-inspired approach," *International Journal on Adaptive Control and Signal Processing*, Vol. 30, 1494-1522, 2016.

- 14) **RL for Output Feedback Control Systems.** This is an important topic for applications of RL for real physical engineering systems. Namely, in general, only in rare cases all state variables are available for feedback, and the system only provides only on its output a certain combination of the state variables, say $y(t) = Cx(t)$ with the rank of the matrix C (the number of linearly independent rows in C) much smaller than the number of the state space variables. How to implement reinforcement learning in this case was discussed in the papers:

F. Lewis and K. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Transactions on Systems, man, and Cybernetics – Part B: Cybernetics*, Vol. 41, 14-25, 2011.

S. Rizvi and Z. Lin, "Output feedback Q-learning control for the discrete-time linear quadratic regulator problem," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, 1623-1536, 2019.

- 15) **Gradient-Based Learning and Differential Games.** The Berkeley researchers in a very recent paper connected continuous-time differential games to the concept of the gradient-based reinforcement learning. In Section 5, they consider the problem within LQ Nash games and algorithm of Li and Gajic (1995). The paper is quite mathematical, but readable.

E. Mazumdar, L. Ratliff, and S. Sastry, "On gradient-based learning in continuous games," *SIAM Journal on Mathematics of Data Science*, Vol. 2, 103-131, 2020.

- 16) **RL for Pareto (Cooperative) Games.** When controllers (agents) cooperate in order to improve their performance criteria over trajectories a dynamic system we have Pareto differential games. A recent paper is a good reference for RL and Pareto differential (dynamic) games:

V. Lopez and F. Lewis, "Dynamic multi-objective control for continuous-time systems using reinforcement learning," *IEEE Transactions on Automatic Control*, Vol. 64, 2869-2874, 2019.

- 17) **RL for Stackelberg Games.** We did not have time to cover Stackelberg differential games in class (differential games with conflict of interest and sequential decision making). You may learn about them from the corresponding class on the game theory that I taught many years ago (uploaded on Sakai). A good paper on reinforcement learning and the Stackelberg games is:

K. Vamvoudakis, F. Lewis, and W. Dixon, "Open-loop learning for hierarchical control problems," *International Journal on Adaptive Control and Signal Processing*, Vol. 33, 285-299, 2017.

18) Reinforcement Learning Policy Iterations for Zero-Sum Differential Games (TERM PAPER OR PROJECT)

In this assignment the students will decide *either* to work on a Final Term Project (if you have good programming skills in MATLAB or Python or Mathematica or any other programming language – similarly to Project 1) *or* Term Paper (if you prefer theoretical considerations). Either choice is worth 25% of the course grade. The topic is the reinforcement learning policy iterations for of the linear-quadratic zero-sum differential games with the policy iterations considered Lecture 21.

Project/Term Paper Report Format: The project/term-paper reports are supposed to be typed and contain standard technical paper parts: introduction, problem formulation, main results, conclusions, references. Students working on the project should include also obtained simulation results, tables, and figures. The program code should be placed at the end in the report as an appendix.

Specific to the Term Paper: Students working on the term papers should read Lecture 21 and references [1]-[3] and prepare the reports based on the class presentation of these topics and the material presented in the papers specifically addressing the issues of *linear-quadratic zero-sum differential games and corresponding policy iterations of reinforcement learning*.

Specific to the Project: It is required to implement *two policy iterations algorithms* using any software: Implement the policy iteration algorithms for *zero-sum games* based on Riccati iterations algorithms from [1]-[2] and the Lyapunov iterations algorithm from [3] (note $\gamma = 1$ in the case of zero-sum games). For both policy iteration algorithms use Example 1 from [2], Lanzon *et al.* (2008).

For each algorithm present two figures and one table. Plot on one figure selected approximates feedback controls obtained for $k = 1, 2, 3, 4, 5$, (or other equally spaced iterations in the case of slow convergence) and on another figure plot the approximate trajectories for a selected (the one that performs the worst in your opinion) state variable for $k = 1, 2, 3, 4, 5$. (or other equally spaced iterations in the case of slow convergence). In the tables provide the values of the performance criteria per iteration and showing the convergence to the optimal values.

- [1] B. Anderson, Y. Feng, and W. Chen, “A Game-theoretic Algorithm to Solve Riccati and Hamilton-Jacobi-Bellman-Isaac (HJBI) Equations in H_∞ Control,” 277-308, in *Optimization and Optimal Control*, A. Chinchuluun *et al.* (eds), Springer, 2010.
- [2] A. Lanzon, Y. Feng, B. Anderson, and W. Rotkowitz, “Computing the positive stabilizing solution of algebraic Riccati equations with an indefinite quadratic term via a recursive method,” *IEEE Transactions on Automatic Control*, Vol. 53, 2280-2291, 2008.
- [3] T. Y. Li and Z. Gajic, “Lyapunov iterations for solving coupled algebraic Riccati equations of Nash differential games and the algebraic Riccati equation of zero-sum games,” p. 333-351, in *New Trends in Dynamic Games and Applications*, G. J. Olsder (ed.), Birkhauser, 1995.

19) Reinforcement Learning Policy Iterations for Nash Differential Games (TERM PAPER OR PROJECT)

In this assignment the students will decide *either* to work on the Final Term Project (similarly to our Project 1; – if you are good in programming) *or* Term Paper 2 (if you prefer theoretical considerations). Either choice is worth 25% of the course grade. The topic is the reinforcement learning policy iterations for the linear-quadratic non-zero-sum (conflict of interest) Nash differential games as presented in *Lectures 22*.

Project/Term Paper Report Format: The project/term-paper reports are supposed to be typed and contain standard technical paper parts: introduction, problem formulation, main results, conclusions, references. Students working on the project should include also obtained simulation results, tables, and figures. The program code should be placed at the end in the report in an appendix.

Specific to the Term Paper: Students working on the term papers should read Lectures 22 and reference [1] and prepare the reports based on the class presentation of this topic and the paper [1], specifically addressing the issue of *linear-quadratic non-zero-sum (Nash) differential games and corresponding policy iterations of reinforcement learning*.

Specific to the Project: Implement the policy iteration algorithm for non-zero sum Nash differential games from reference [1]. Good system model to be simulated is given on page 325 of the book chapter by Vrabie and Lewis 2012, reference [2].

Present two figures and one table. Plot on one figure selected approximates feedback controls obtained for $k = 1, 2, 3, 4, 5$, (or other equally spaced iterations in the case of slow convergence) and on another figure plot the approximate trajectories for a selected (the one that performs the worst in your opinion) state variable for $k = 1, 2, 3, 4, 5$. (or other equally spaced iterations in the case of slow convergence). In the tables provide the values of the performance criteria per iteration and showing the convergence to the optimal values.

- 1) T. Y. Li and Z. Gajic, “Lyapunov iterations for solving coupled algebraic Riccati equations of Nash differential games and the algebraic Riccati equation of zero-sum games,” p. 333-351, in *New Trends in Dynamic Games and Applications*, G. J. Olsder (ed.), Birkhauser, 1995.
- 2) D. Vrabie and F. Lewis, “Integral Reinforcement Learning for Finding Online the Feedback Nash Equilibrium of Nonzero-Sum Differential Games,” 313-330, Chapter 17 in *Advances in Reinforcement Learning*, A. Mellouk (ed.), IntechChina, 2012.

20) Value and Policy Iterations for Discrete-Time Systems (for undergraduate students only)

This term paper is based on Lecture 18 and its Examples 1 and 2. The students are supposed to present the value and policy iterations as successive approximation techniques in the discrete-time domain, according to the material presented in this lecture and run 1 or 2 extra iterations for the systems presented in Examples 1 and 2, indicating the convergence to the optimal solutions.

SURVEY/OVERVIEW PAPERS

- [1] L. Kaelbling, M. Littman, and A. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, 237-285, 1996.
- [2] F-Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: An introduction," *IEEE Computational Intelligence Magazine*, 39-47, May 2009.
- [3] F. Lewis and D. Vrabie, "Reinforcement learning and feedback control," *IEEE Circuits and Systems Magazine*, 32-50, Third Quarter, 2009.
- [4] D. Bertsekas, "Approximate policy iteration: a survey and some new results," *Journal of Control Theory and Applications*, Vol. 9, 310-335, 2011.
- [5] F. Lewis, D. Vrabie, and K. Vamvoudakis, "Reinforcement learning and feedback control," *IEEE Control Systems Magazine*, 76-105, Dec. 2012.
- [6] Z-P. Jiang and Y. Jiang, "Robust adaptive dynamic programming for linear and nonlinear systems: An overview," *European Journal of Control*, Vol. 19, 417-425, 2013.
- [7] K. Vamvoudakis, H. Modares, B. Kiumarsi, and F. Lewis, "Game theory-based control system algorithms with real-time reinforcement learning: How to solve multiplayer games on line," *IEEE Control Systems Magazine*, 33-52, 2017.
- [8] B. Kiumarsi, K. Vamvoudakis, H. Modares, and F. Lewis, "Optimal and autonomous control using reinforcement learning: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 29, 2042-2062, 2018.

All the papers cited in this document can be downloaded from the Rutgers University libraries website. Particularly, going to the New Brunswick libraries, the students may search some standard databases like: *IEEE Xplore*, *ScienceDirect* (book and journal Elsevier publisher database), *Web of Science*, and *Wiley Online Library*. Prof. Gajic will make efforts to upload all the papers in the document to the final term paper file.