# Lahore University of Management Sciences

# EE569/CS6314 – Dynamic Programming and Reinforcement Learning
# Fall 2023

| | |
|---|---|
| Instructors | Hassan Jaleel |
| Room No. | 9-245 A |
| Office Hours | TBA |
| Email | hassan.jaleel@lums.edu.pk |
| Telephone | +92 (42) 3560-8466 |

**Course Teaching Methodology** (Please mention the following details in plain text)

- Teaching Methodology: Synchronous
- Lecture details: All the lectures will be delivered on-campus in a regular classroom setting

| Course Basics | | | | | | |
|---|---|---|---|---|---|---|
| Credit Hours | 3 | | | | | |
| Lecture(s) | Lectures per week | 2 | Duration | 1 hour 15 minutes each | Timings and venue | TuTh: 09:30am - 10:45am |

| Course Distribution | |
|---|---|
| Core | |
| Elective | Electrical Engineering, Computer Science |
| Open for Student Category | Junior year, Senior year, Graduate students |
| Close for Student Category | |

**COURSE DESCRIPTION**

Dynamic programming is a framework for deriving optimal decision strategies in evolving and uncertain environments. In the first part of the course, we will cover the theoretical foundations of dynamic programming in detail. Topics include the principle of optimality in deterministic and stochastic settings, LQR control, and value and policy iteration.
In the second part of the course, we will focus on approximation techniques and simulation-based methods such as online reinforcement learning, approximate dynamic programming, and model predictive control. We will cover recent advances in these online algorithms that have resulted in a breakthrough in the practical application of dynamic programming concepts to complex systems. Throughout the course, we will draw examples from robotics, data and sensor networks, and social and economic models.

**Course Prerequisite(s)/Co-Requisite(s)**

Pre-requisites: None
Co-requisites: None
Recommended: Basic optimization; Basic probability; Linear algebra; Differential equations; MATLAB readiness.

**Grading Breakup and Policy**

Assignments: (5 -6): 30% (Assignments will include Analytical problems as well as programming exercises.)

Midterm exam: 40%

Final Exam: 30% (Final exam will either be a regular 3 hour inclass exam or a take home exam. Details will be finalized based on the final course enrollment)

**Examination Detail**

# Lahore University of Management Sciences

| | |
|---|---|
| Midterm Exam | Duration: 2 hours<br>Exam Specifications: closed book, closed notes, and help sheets. |

| Course Learning Outcomes | |
|---|---|
| | The students should be able to: |
| CLO1 | Formulate decision-making problems under dynamic and uncertain conditions in a standard dynamic programming setting. |
| CLO2 | Analyze dynamic programs over finite and infinite time horizons using mathematical tools such as the principle of optimality and contraction mappings |
| CLO3 | Implement algorithms to find solutions for both model-based and model-free problem formulations. |

Relation to EE Program Outcomes

| EE-567 CLOs | Related PLOs | Levels of Learning | Teaching Methods | CLO Attainment checked in |
|---|---|---|---|---|
| CLO1 | PLO1 | Cog 3 | Instruction, Assignments | Assignments, midterms, and project |
| CLO2 | PLO2 | Cog 4 | Instruction, Assignments | Assignments, midterms, and project |
| CLO3 | PLO3 | Cog 3 | Instruction, Assignments | Assignments, midterms, and project |

Course Overview

| **No. of weeks** | **Topics** | **Related CLOs & Additional Remarks** |
|---|---|---|
| 1.5 | Multi-stage decision problems, Principle of optimality, Deterministic dynamic programming framework | CLO1 |
| 1.0 | Worst case dynamic program: Problem Formulation and analysis | CLO1 & CLO2 |
| 1.0 | Stochastic shortest path problems, Label correcting algorithm and its variations (Dijkstra, A*) | CLO1 & CLO3 |
| 1.0 | Probability review | CLO2 |
| 2 | Finite-state Markov Chain, Stochastic dynamic program, Example of Stochastic dynamic programs: LQR optimal control | CLO1 |
| 1.0 | Value iteration | CLO2 & CLO3 |
| 1.0 | Policy iteration | CLO3 & CLO3 |
| | Q Factor, Reinforcement learning introduction, | |

| | | |
|---|---|---|
| 1.5 | Q learning: Off policy and On policy,<br>Actor critic learning | CLO2 & CLO3 |
| 1.5 | Temporal difference learning: Policy evaluation TD(0), Approximate policy evaluations, Policy evaluation variants (TD(0), TD(1), TD(λ)) | CLO3 |
| 1.0 | Policy gradient | CLO3 |
| 1.5 | Model Predictive Control (MPC) | CLO2 & CLO3 |

| Textbook(s)/Supplementary Readings |
|---|
| **Text book:**<br>• Dynamic Programming and Optimal Control, by D. P. Bertsekas<br>• Reinforcement Learning by Sutton and Barton.<br><br>**Supplementary Reading:** Lecture handouts and research papers will be provided throughout the semester for supplementary reading. |

| | |
|---|---|
| **Prepared by:** | **Hassan Jaleel** |
| **Date:** | **July 2023** |