GUC

German University in Cairo
Faculty of Engineering and Material Science
Department of Mechatronics Engineering

Reinforcement Learning and Optimal Control (MCTR 1024)

Quiz 3: Deep Reinforcement Learning Methods and Convex Optimization

Submission Deadline: Tuesday, 13 May, 2025

This take-home quiz is groups of 2 students.

Student 1:

Name:	
GUC ID:	

Student 2:

Name:	
GUC ID:	

Please note that cheating will not be tolerated and that it is your responsibility to ensure the genuineness of your work.

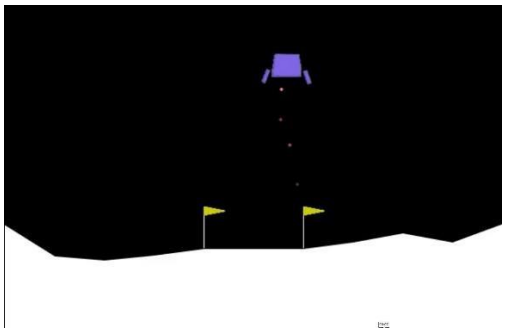
Question 1: DRL on Lunar Lander Environment

Lunar Lander environment is a classic rocket trajectory optimization problem. The lander starts at the top center of the viewport with a random initial force applied to its center of mass. The rewards given by the environment can be classified as follows:

- The reward for moving from the top of the screen to the landing pad and coming to rest is about 100-140 points. If the lander moves away from the landing pad, it loses the reward.
- If the lander crashes, it receives an additional -100 points.
- If it comes to rest, it receives an additional +100 points.
- Each leg with ground contact is +10 points.
- Firing the main engine is -0.3 points each frame.
- Firing the side engine is -0.03 points each frame.
- Solved is 200 points.

There are two versions for the environment:

- **Discrete:** There are four discrete actions available: do nothing, fire left orientation engine, fire main engine, fire right orientation engine.
- **Continuous**: Two actions (corresponding to the throttle of the engines) will be used. The first coordinate of an action determines the throttle of the main engine, while the second coordinate specifies the throttle of the lateral boosters.



Kindly refer to the following link for more information regarding the environment:

https://www.gymlibrary.dev/environments/box2d/lunar_lander/

Required:

Consider the continuous version of the environment:

1. Create a **DDPG** agent to solve the continuous version of the Lunar Lander environment.
2. Consider the following parameters while creating the agent:
   - Learning rate $\epsilon = 0.001$
   - Batch size = 64
   - Discount factor $\gamma = 0.99$
   - Learning starts after 50000 time steps
   - Use a normal distribution function of zero mean and standard deviation equal to 0.1 to act as the action noise
   - Total training timesteps = 250000
3. Show the training results of the agent using **Tensor Board.**
4. Use **stable-baselines3** library to create the agent.
5. It is recommended to write the final python code using **colaboratory** or **jupyter**

References:

- DDPG algorithm: https://stable-baselines3.readthedocs.io/en/master/modules/ddpg.html
- Lunar Lander gym environment: https://www.gymlibrary.dev/environments/box2d/lunar_lander/
- Get started with Tensor Board: https://www.tensorflow.org/tensorboard/get_started#:~:text=TensorBoard%20is%20a%20tool%20for,dimensional%20space%2C%20and%20much%20more.

Question 2: Convex Optimization:

Consider the following function:

$$F(x) = \frac{1}{2} X^T \begin{bmatrix} 10 & -6 \\ -6 & 10 \end{bmatrix} X + [4 \quad 4] X$$

Where, $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Required:

- Prove that this is a convex optimization problem.
- Find X such that $F(x)$ is minimum.
- Use **CasADi** toolbox on Python to solve the optimization problem

References:

- CasADi open source: https://web.casadi.org/

Submission Guidelines:

- Submit a zip file with the following:
  - The quiz (in this form) along with all python codes implemented and resulting figures.
  - The original code files and ".ipynb" notebooks.
  - The training results of the DDPG agent.
- Your submission will be through the following link: https://forms.gle/3gGLky9efagHQvNM7
- The submission deadline is on Tuesday, 13-May-2025, at 23:59.
- Please note that cheating will not be tolerated and that it is your responsibility to ensure the genuineness of your work.

Best of Luck!