Reinforcement Learning Methods for Multibody Systems evaluated with Controlled Multi-Link Inverted Pendulum on the Cart

Oleg Rogov[1], Peter Manzl[2], Johannes Gerstmayr[2], Grzegorz Orzechowski[1]

[1] Department of Mechanical Engineering
LUT University
Yliopistonkatu 34, 53850 Lappeenranta, Finland
[oleg.rogov, grzegorz.orzechowski]@lut.fi

[2] Institute of Mechatronics
University of Innsbruck
Technikerstraße 13, A-6020 Innsbruck, Austria
[peter.manzl, johannes.gerstmayr]@uibk.ac.at

EXTENDED ABSTRACT

1 Introduction

It is possible to model complex structures using multibody systems, from a pendulum through robots to vehicles. Such systems can be further employed as a base of control applications. However, these systems require appropriate control algorithms to function as desired. Reinforcement Learning (RL) has recently gained popularity and proven capable of solving complex control tasks. Therefore, the aim of this work is to determine if we can effectively combine the advantages of RL methods with a multi-body model for robust control of the system. In this study, we will analyze systems of inverted pendulums on a cart, controlled by three different RL methods: Advantage Actor-Critic (A2C), Deep Q-Network (DQN), and Proximal Policy Optimization (PPO). Investigations start with the single inverted pendulum, which is a standard test case for RL methods. Hereafter, complexity is increased by using a higher number of links, such as double and triple inverted pendulums on a cart, representing problems closer to real-life multibody systems. We use the Exudyn framework [2] for the generic creation of the rigid multibody model, using a minimum coordinate formulation for the chain topology and allowing future extension to other configurations or flexible bodies. For the reinforcement learning aspect, we use Stable-Baselines3 (SB3) [3], a reliable and user-friendly tool that implements state-of-the-art RL methods.

2 Multi-link inverted pendulum on a cart

The base for the *n*-link inverted pendulum is the control model from Barto *et al.* [1]. An *n*-link inverted pendulum on a cart is a physical setup where *n* rigid bodies (poles) are attached to a moving cart via revolute joints in series, see Figure 1. The pole is initially placed close to the upright position, with random initial states provided by the RL method. The cart can move along a frictionless track on a horizontal axis.

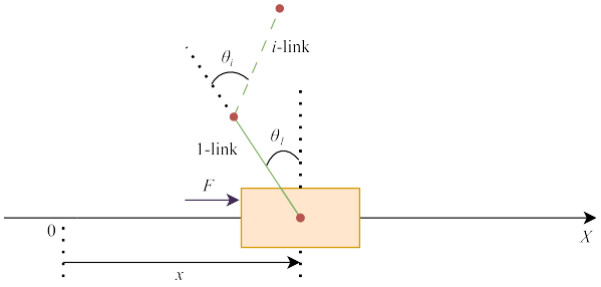


Figure 1: *N*-link inverted pendulum on a cart representation with link length of 1 m and link mass of 0.1 kg

The goal of the control task is to balance the pendulum by applying horizontal force to the cart. The force can be applied either left or right, and it has a constant, predefined magnitude. The state space of the single inverted pendulum consists of four variables: the position and velocity of the cart and the angle and angular velocity of the pendulum. In this study, we examine the one-link inverted pendulum on a cart system, and more complex systems are under further examination. The system is developed with Python programming language, where the Exudyn model is created as an OpenAI gym environment, which is directly used by SB3 algorithms. The parameter variation integrated into Exudyn then allows to run a set of trainings in parallel, only limited by the number of available CPU cores.

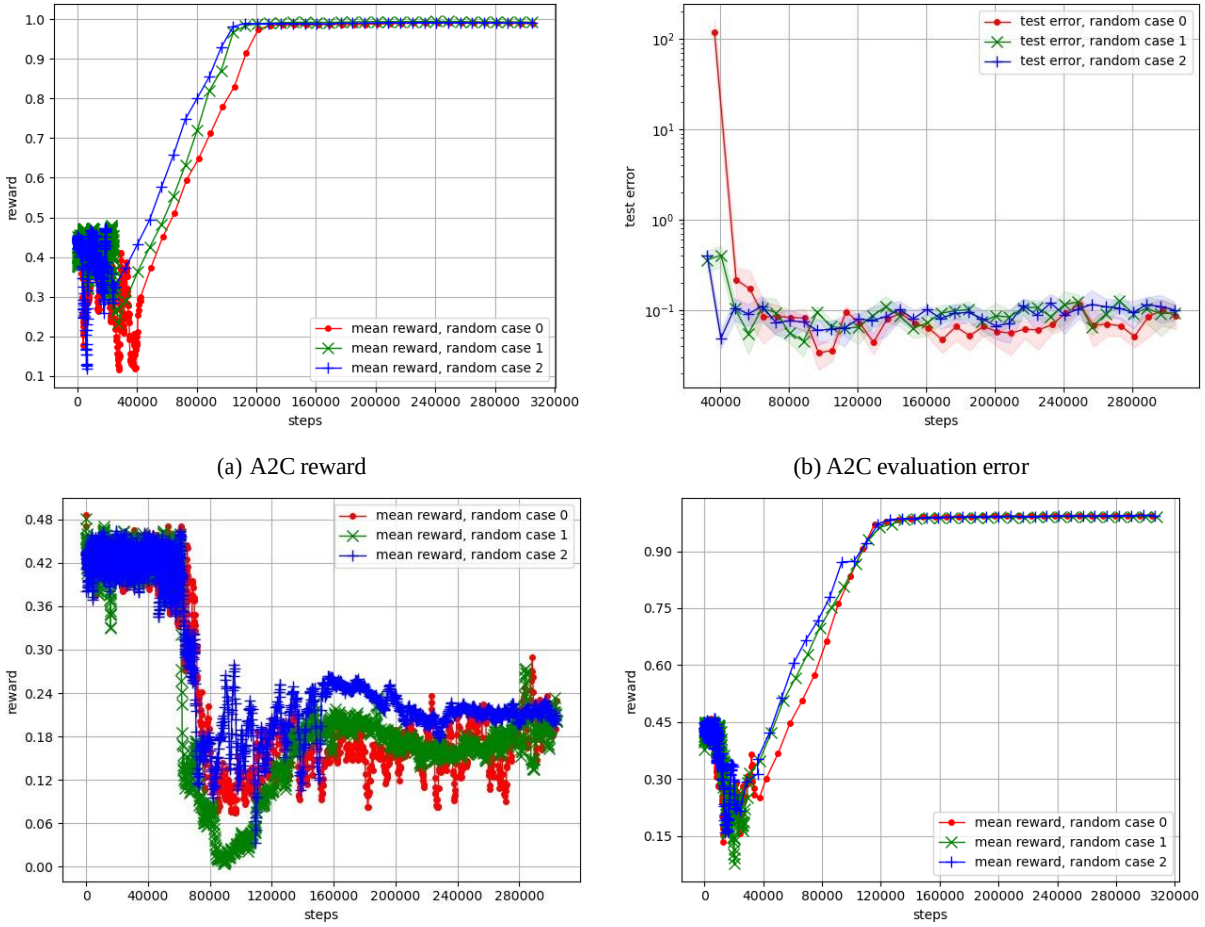
3 Control algorithms

The selected control algorithms for this study are three reinforcement learning methods A2C, DQN, and PPO. The RL algorithms aim to select the actions that maximize the expected cumulative reward. A2C employs policy-based (Actor) and value-based (Critic) networks. The actor generates actions to control the system, while the critic evaluates the quality of the actions taken by the actor. The actor is updated based on the evaluation of the critic to improve the quality of the actions taken. DQN uses a deep neural network to approximate the Q-function, which describes the expected reward for taking a certain action in a

certain state. PPO is the last considered control algorithm. It uses a policy-based approach. The algorithm optimizes the policy network to move towards the best-performing policies while limiting the size of the change in one iteration to maintain stability. Implementation of those algorithms is done using Stable-Baselines3 [3] library.

4 Results and conclusions

Figure 2 shows behavior of the reward over total timesteps on the example of a one-link inverted pendulum with the same training parameters for three different random seeds (cases). From it we can see that A2C algorithm is considered to be fully trained in all three random cases with approximately 110.000 timesteps, while PPO achieves the same reward behavior at nearly 145.000 timesteps. DQN could not achieve successful training results within the given 300.000 timesteps. The evaluation error, only shown for A2C, is based on 8 tests with randomized initialization. The error is computed from the maximum of position (m) and angle (rad) errors within the final 2.5 s of a 10 s test, showing mean error (solid) and 95% confidence interval (shadowed).



(a) A2C reward

(b) A2C evaluation error

(c) DQN reward

(d) PPO reward

Figure 2: Rewards over timesteps for A2C, DQN and PPO and evaluation error for A2C, using mean value of 10 consecutive steps

Overall, our results highlight the importance of selecting the appropriate algorithm for a given control problem. The results of our study can be used to guide future research into the control of more general multibody systems. In the further work, we will show in detail the results of two- and three-link systems, where the increasing complexity considerably impairs the performance of the RL methods.

References

[1] A. G. Barto, R. S. Sutton and C.W. Anderson. Neuronlike adaptive elements that can solve difficult learning control problems. IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-13, 5:834-846, 1983.

[2] J. Gerstmayr. Exudyn – Flexible Multibody Dynamics Systems with Python and C++. https://github.com/jgerstmayr/EXUDYN (accessed on January 31, 2023).

[3] A. A. Raffin, A. Hill, A.Gleave, A. Kanervisto, M. Ernestus and N. Dormann. Stable-Baselines3: Reliable Reinforcement Learning Implementations. Journal of Machine Learning Research, vol. 22, 268:1-8, 2021.