Paper ID: 88540
oral
student

IAF ASTRODYNAMICS SYMPOSIUM (C1)
Interactive Presentations - IAF ASTRODYNAMICS SYMPOSIUM (IP)

Author: Mr. Andrea Forestieri
Politecnico di Torino, Italy

Ms. Alessia De Iuliis
Politecnico di Torino, Italy
Mr. Davide Costigliola
Politecnico di Torino, Italy

# GLOBAL TRAJECTORY OPTIMIZATION WITH ALPHAZERO: A NOVEL APPROACH APPLIED TO THE GTOC 11 CHALLENGE

**Abstract**

In the GTOC 11 challenge, the objective is to construct a Dyson ring using twelve stations. To achieve this, ten spacecraft are dispatched from Earth to explore the asteroid belt. The goal is to visit as many asteroids as possible while minimizing propellant consumption. The visited asteroids can then be activated using a fictional device to complete a low-thrust, constant acceleration spiral to reach a chosen target station and provide building materials for the megastructure. As a whole, this problem is impractical to tackle. Hence, it is helpful to break it down into three subproblems. The first subproblem involves designing asteroid chains to be visited. This can be seen as a combinatorial task that requires identifying favorable asteroid sequences connected by impulsive transfers. The second subproblem involves designing the orbit of the stations for the Dyson Ring. Lastly, the third subproblem requires constructing the stations by developing efficient transfer opportunities for all asteroids to reach one of the twelve target stations.

Since the combinatorial multileg transfer problem often arises in global trajectory optimization problems, this work focuses on tackling the multileg asteroid-to-asteroid trajectories, analyzing the capabilities of reinforcement learning within this framework.

Combinatorial problems related to space trajectory design often employ ad-hoc heuristics in algorithmic solutions. Usually, these heuristics are influenced by prior knowledge of problem characteristics, shaping the algorithm's policy. This paper explores an alternative approach using reinforcement learning, specifically the AlphaZero algorithm, leveraging the dynamics of the environment to optimize decision-making for reward maximization within a predefined timeframe.

AlphaZero introduces a significant shift from classic Monte Carlo Tree Search by replacing random rollouts with a neural network trained on episodes relevant to the problem. In this case, the episodes consist of numerous asteroid-to-asteroid chains of a mothership. The neural network approximates state values (expected cumulative rewards) and policies (probability distributions over actions). This integration enhances the tree policy and rollout policy during tree search, with simulations guided by learned policies offering priors on action probabilities and state values. At its core, AlphaZero seeks its own heuristic through the neural network, distinguishing itself from classical algorithms.

In summary, the incorporation of AlphaZero into the GTOC 11 problem-solving process demonstrates a sophisticated approach to asteroid selection. By combining a tree search method with the power of neural networks, AlphaZero showcases its potential in addressing complex combinatorial challenges.