Multi-Objective Reinforcement Learning for Low-Thrust Transfer Design between Periodic Orbits

Christopher J. Sullivan[1], Natasha Bosanac[1],

Alinda K. Mashiku[2], Rodney L. Anderson[3]

[1]University of Colorado Boulder

[2]NASA Goddard Space Flight Center

[3]Jet Propulsion Laboratory, California Institute of Technology

Motivation

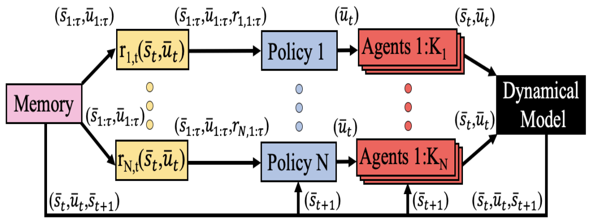
- Low-thrust transfers in multi-body systems exist in high-dimensional, multi-objective solution spaces
- Constructing the transfers and generating the solution spaces may be time consuming and intensive for human trajectory designers
- Developing solutions autonomously, without the need for a predefined reference trajectory or initial guess, aids trajectory designers

Goal: Autonomously generate low-thrust transfers in a multi-objective solution space using multi-objective reinforcement learning (MORL) and validate with optimization scheme

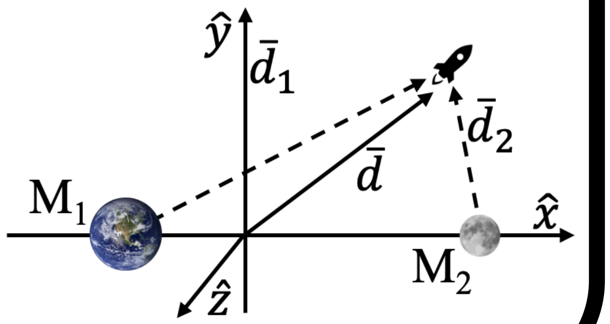
Overview

Technical Background

Reinforcement Learning

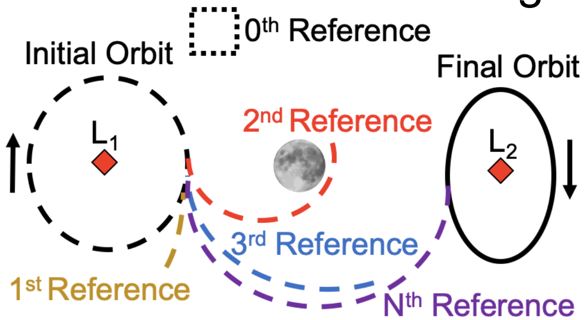


Dynamical Model

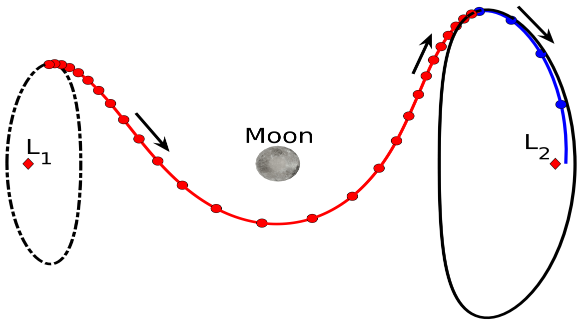


Technical Approach

Reinforcement Learning

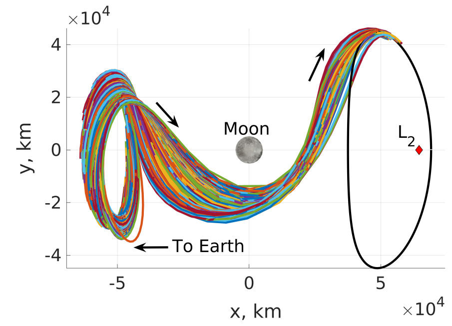


Optimization Scheme

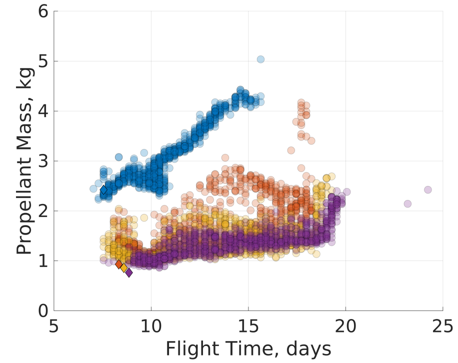


Results

Low-Thrust Transfers



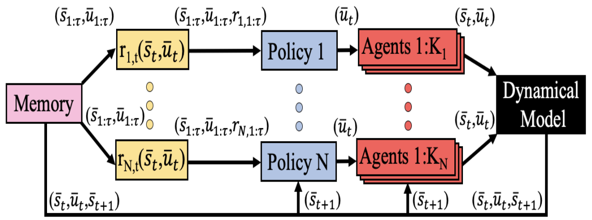
Solution Space



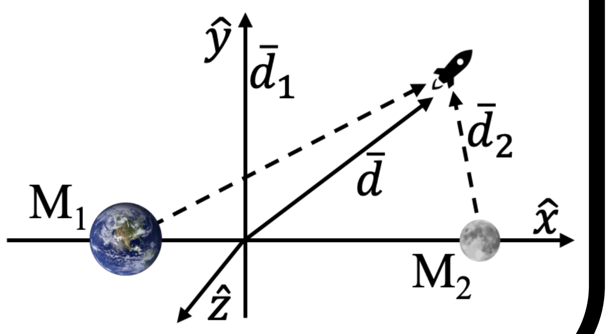
Overview

Technical Background

Reinforcement Learning

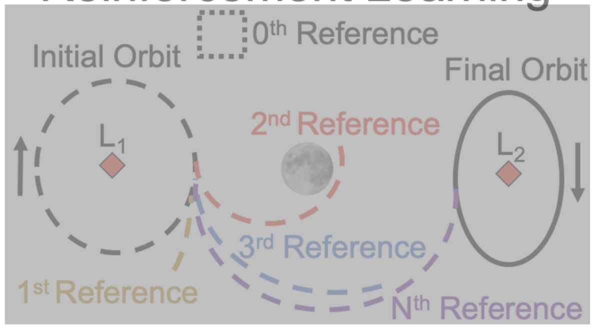


Dynamical Model

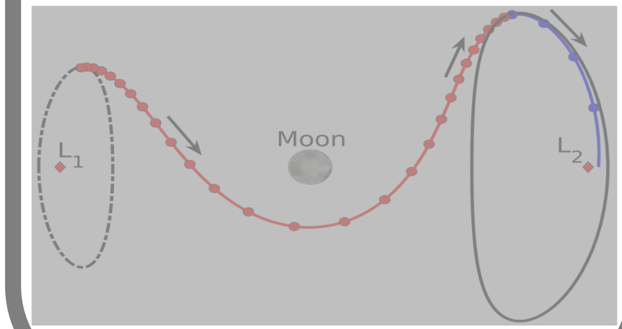


Technical Approach

Reinforcement Learning

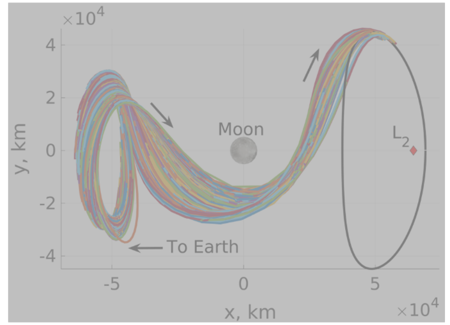


Optimization Scheme

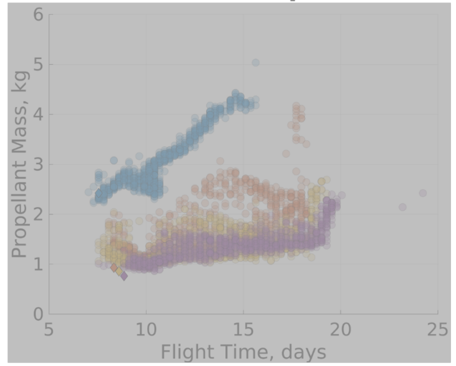


Results

Low-Thrust Transfers

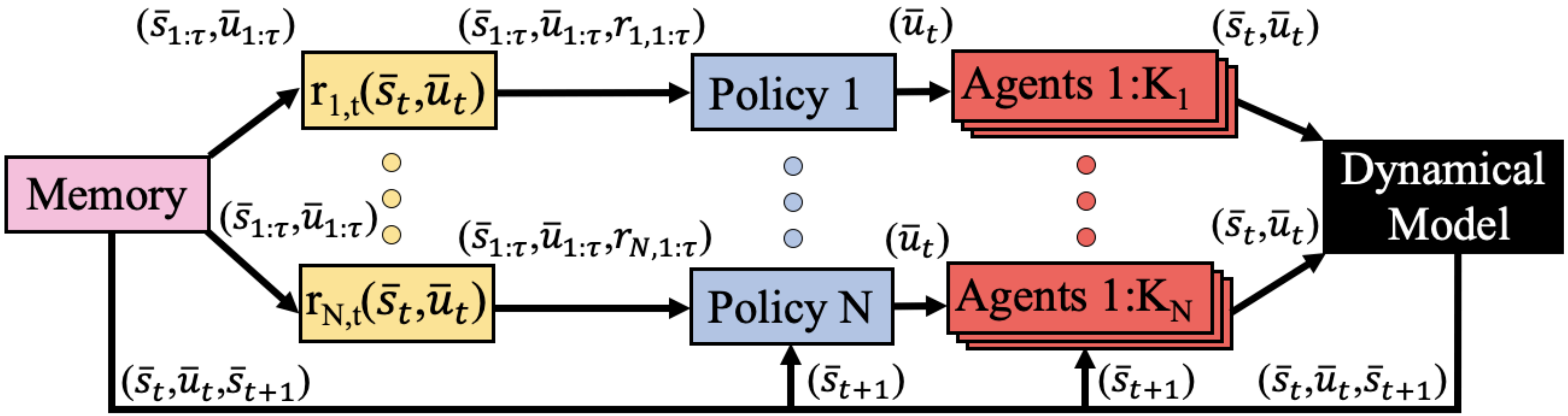


Solution Space



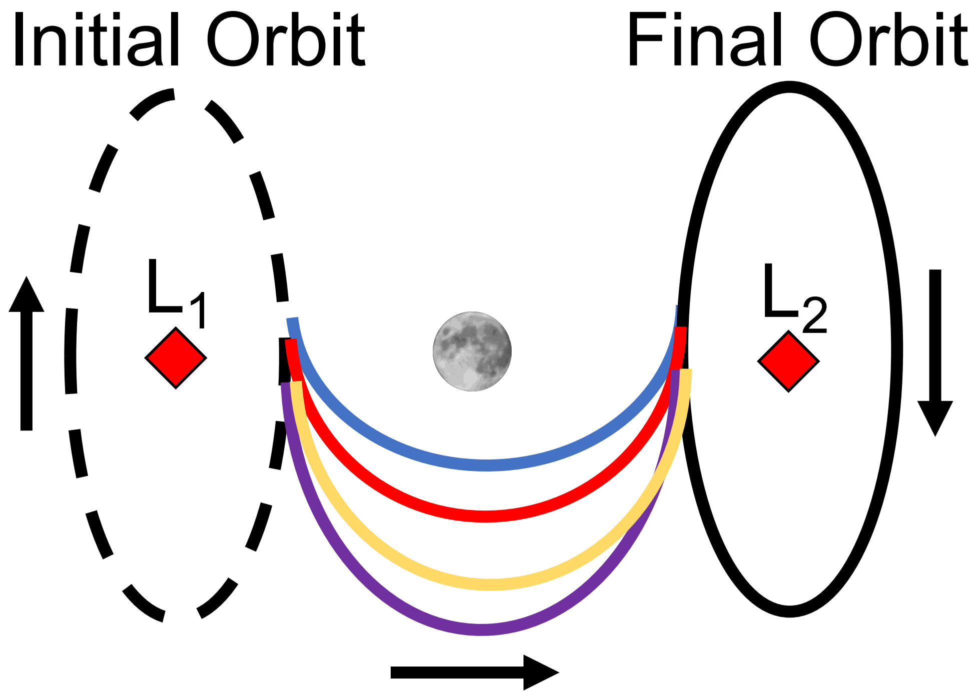
Multi-Reward Proximal Policy Optimization (MRPPO)

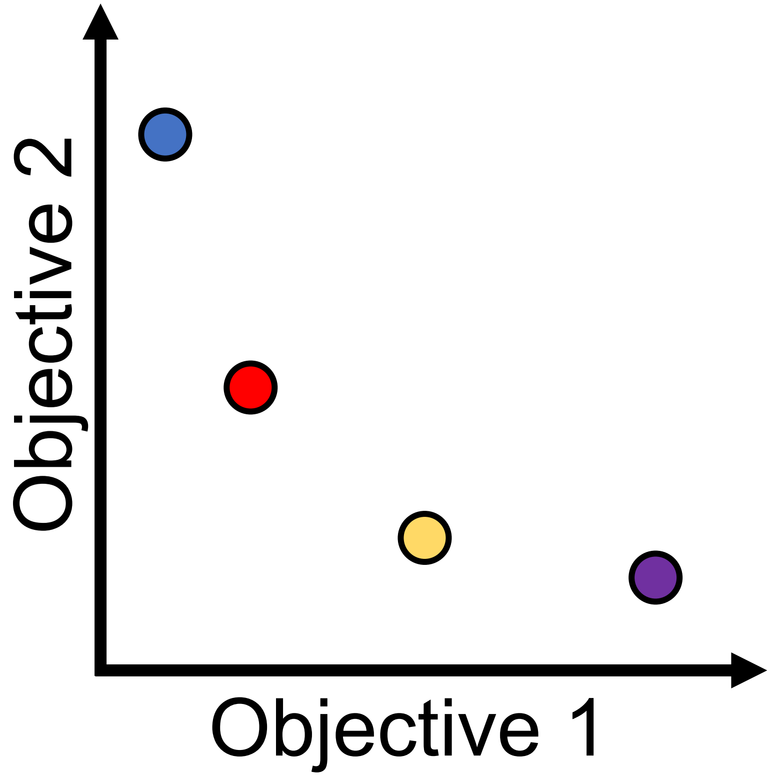
- Multi-objective framework leveraging Proximal Policy Optimization as a foundation
- Enables multiple policies to be more efficiently trained to recover a variety of solutions in the multi-objective solution space



Trajectory Design Application

Goal: Use MRPPO to recover the multi-objective solution space associated with low-thrust transfers between periodic orbits





Dynamical Environment

Low-Thrust Circular Restricted Three-Body Problem (CR3BP)

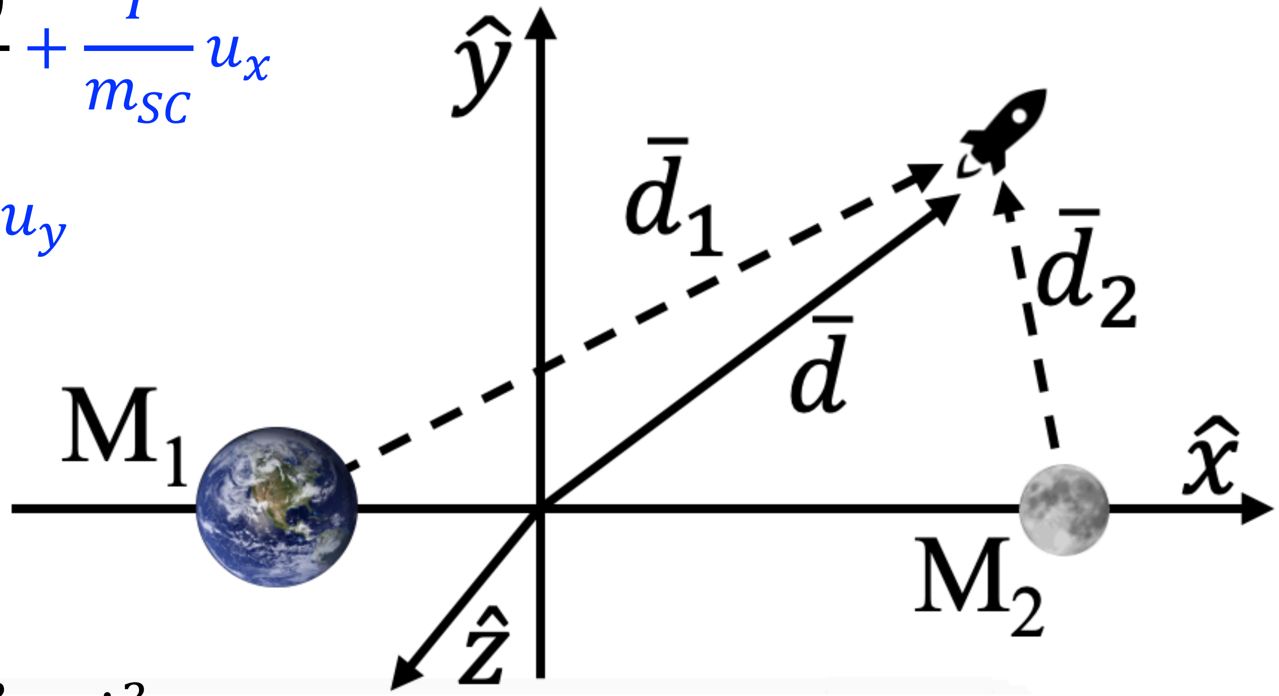
$$\ddot{x} = 2\dot{y} + x - \frac{(1-\mu)(x+\mu)}{d_1^3} - \frac{\mu(x-1+\mu)}{d_2^3} + \frac{T}{m_{SC}}u_x$$

$$\ddot{y} = -2\dot{x} + y - \frac{(1-\mu)y}{d_1^3} - \frac{\mu y}{d_2^3} + \frac{T}{m_{SC}}u_y$$

$$\ddot{z} = -\frac{(1-\mu)z}{r_1^3} - \frac{\mu z}{r_2^3} + \frac{T}{m_{SC}}u_z$$

$$\dot{m}_{SC} = \frac{-T|\bar{u}|}{I_{sp}g_0}$$

$$C_J = x^2 + y^2 + \frac{2(1-\mu)}{d_1} + \frac{2\mu}{d_2} - \dot{x}^2 - \dot{y}^2 - \dot{z}^2$$



Spacecraft Parameters

- ESPA-class SmallSat
  - Thrust direction specified in Earth-Moon rotating frame
  - Single variable thrust, constant Isp engine

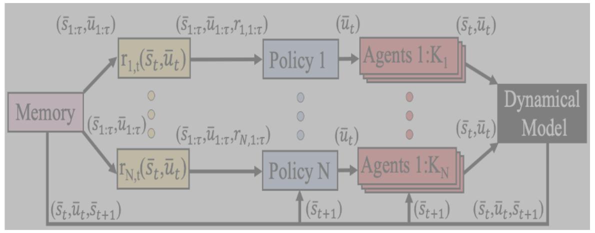
Parameter	Assumed Value
Maximum Thrust (mN)	150
Isp (s)	3,000
Initial Wet Mass (kg)	180

Jet Propulsion Laboratory
California Institute of Technology
University of Colorado **Boulder**
CCAR

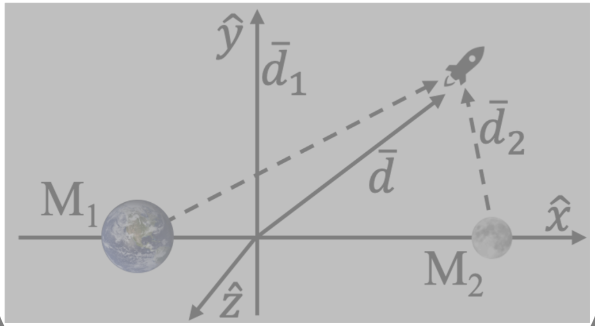
Overview

Technical Background

Reinforcement Learning

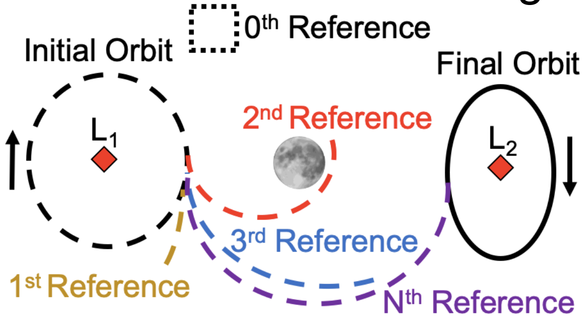


Dynamical Model

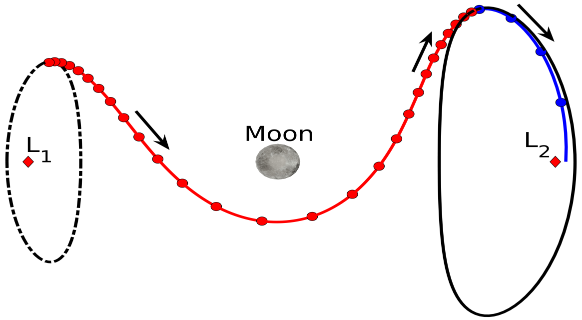


Technical Approach

Reinforcement Learning

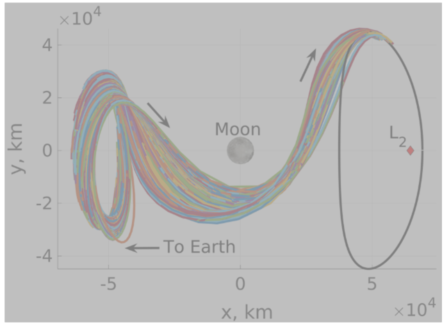


Optimization Scheme

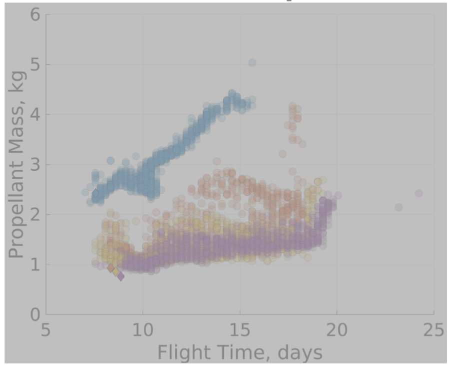


Results

Low-Thrust Transfers

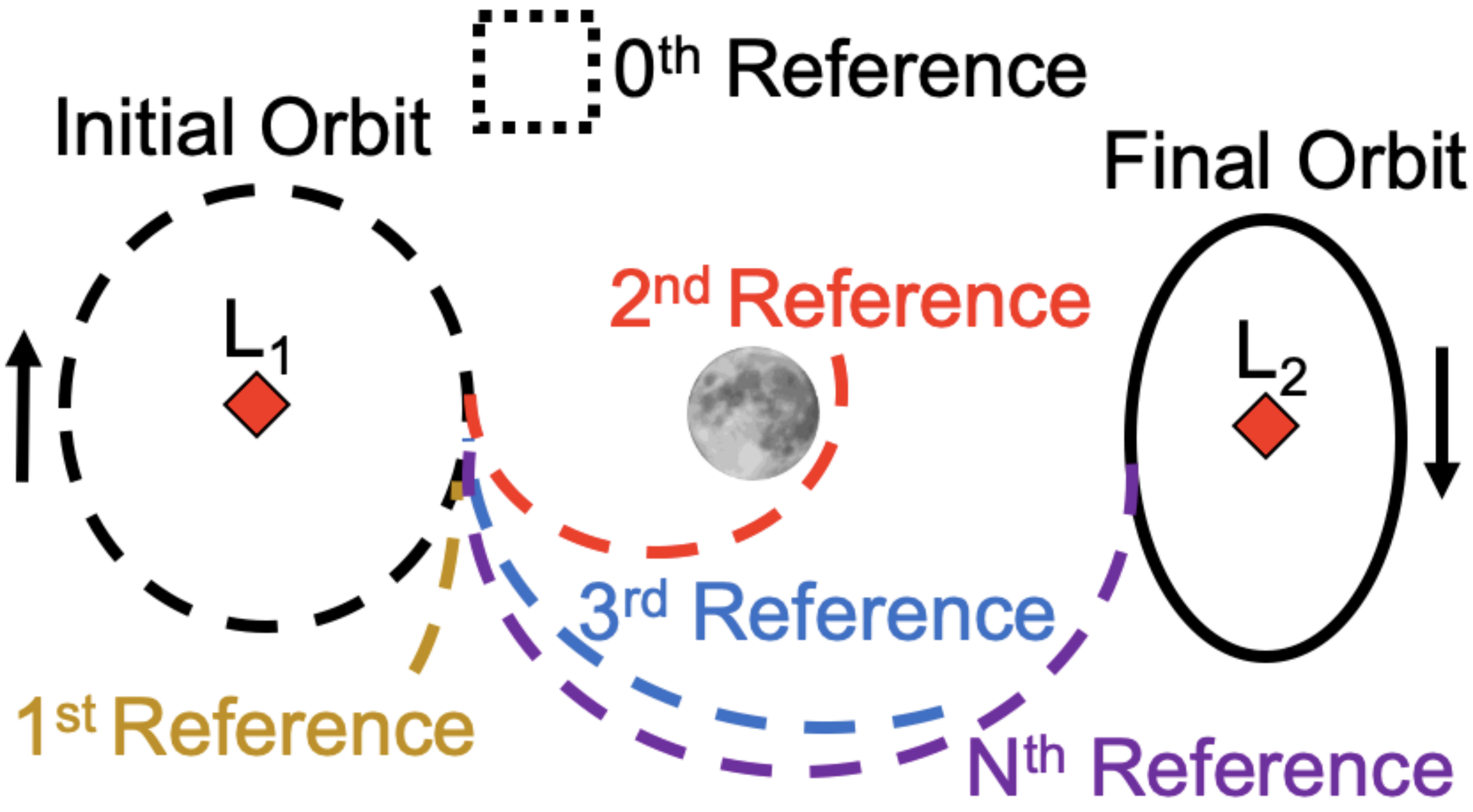


Solution Space



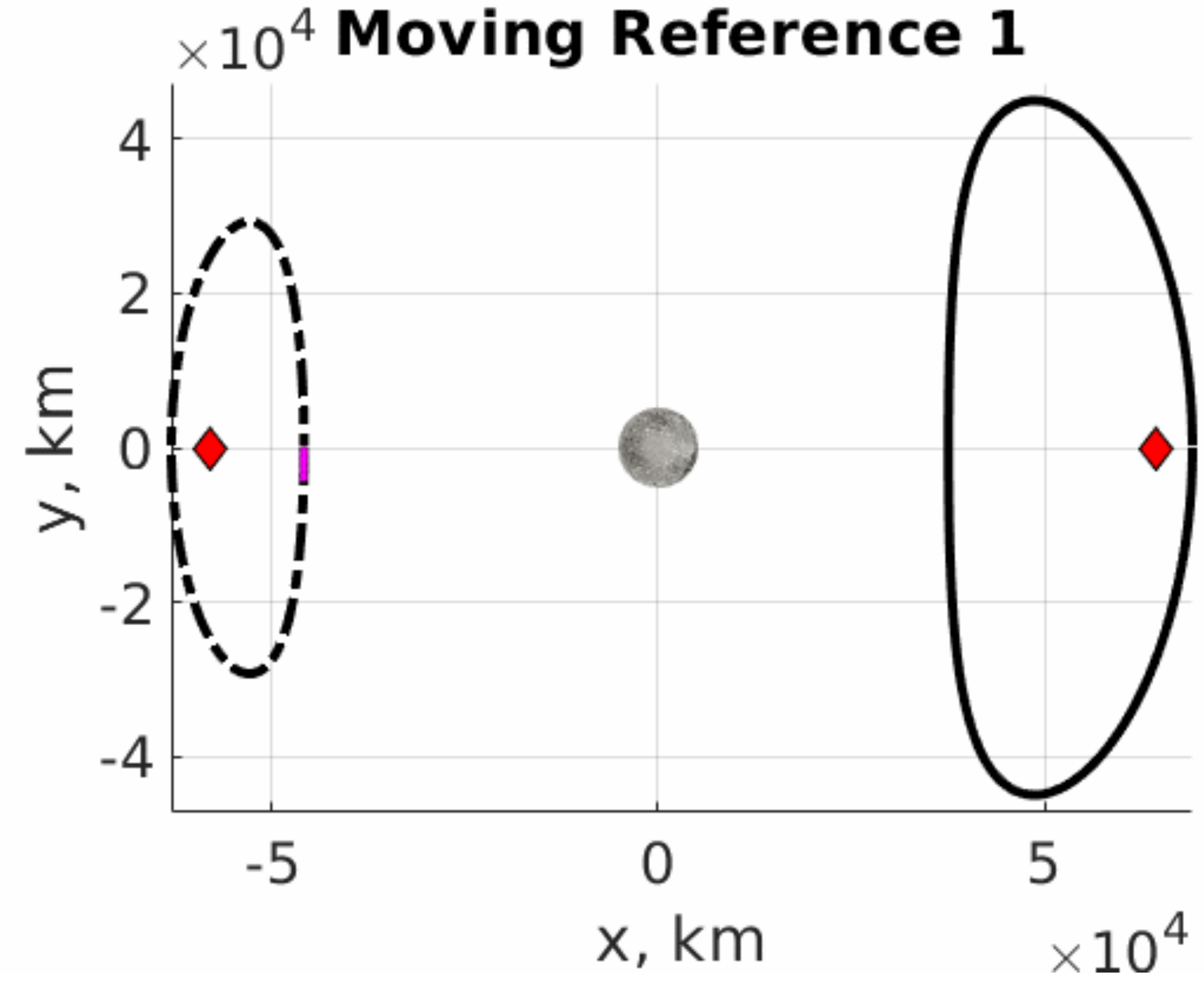
Moving Reference Modification

- Use the best trajectory generated so far as a reference for MRPPO to follow
- Each policy discovers and follows its own reference trajectory
- Aids the policies in autonomously converging on transfers in complex scenarios



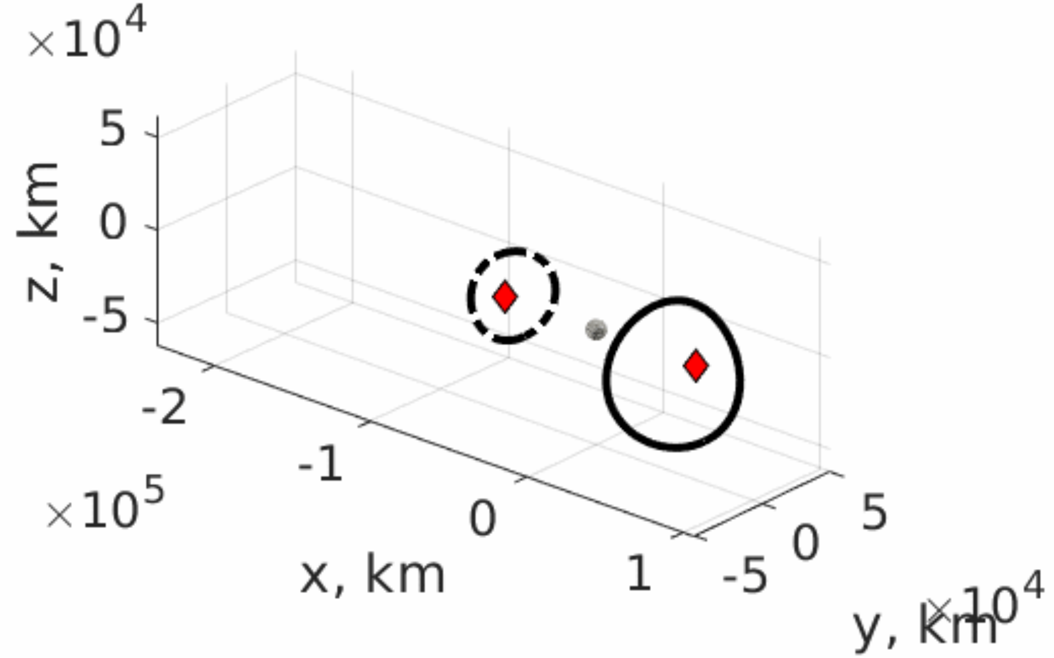
Moving Reference Modification

- Facilitates recovery of a variety of solutions during training
- General required information:
  - Initial orbit states
  - Final orbit states
  - Objectives/constraints
  - Spacecraft parameters



Trajectory Design Scenario

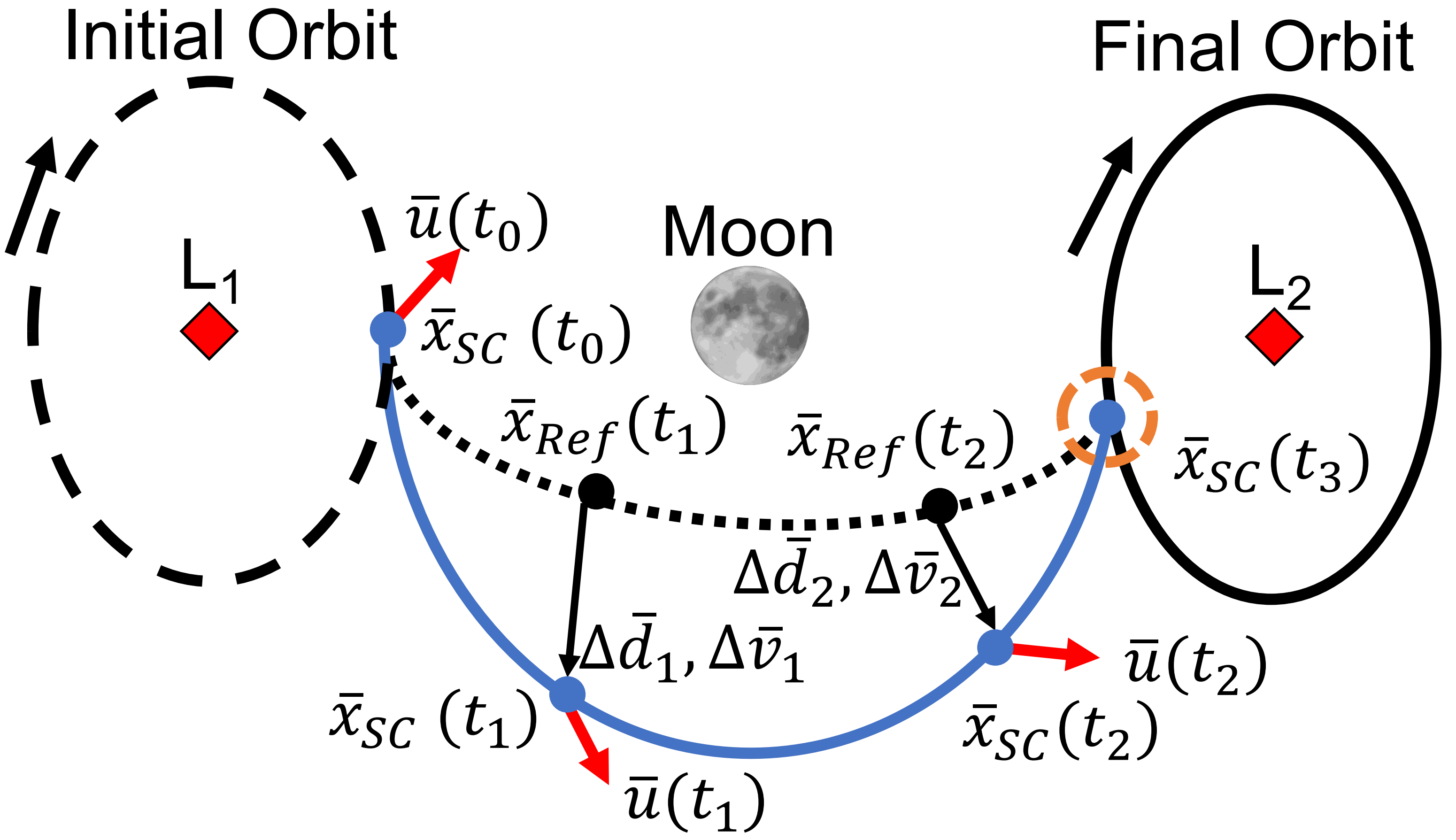
- Initial orbit: L_1 Northern Halo, C_J = 3.15
- Final orbit: L_2 Southern Halo, C_J = 3.11
- Initial perturbations to position and velocity components:
  - Randomly drawn along initial orbit
  - Gaussian distribution N(0,1E-3)
  - Evaluated on 1,000 common IC's



Uncontrolled Perturbed Paths

Trajectory Design Scenario

- Time steps:
  - ~6 hours per action
  - Maximum 150 actions
- Termination Conditions
  1. Reaches 39 day limit
  2. Diverges from reference
  3. Passes within 5E-3 in position and velocity of final orbit
  4. Impacts Moon



Trajectory Design Scenario

- Reward Function:
  - If closest to initial orbit:

    $$r = -8 - 10|\Delta\bar{d}| - |\Delta\bar{v}| - c_{mass}\Delta m + \Omega$$

  - If closest to reference trajectory:

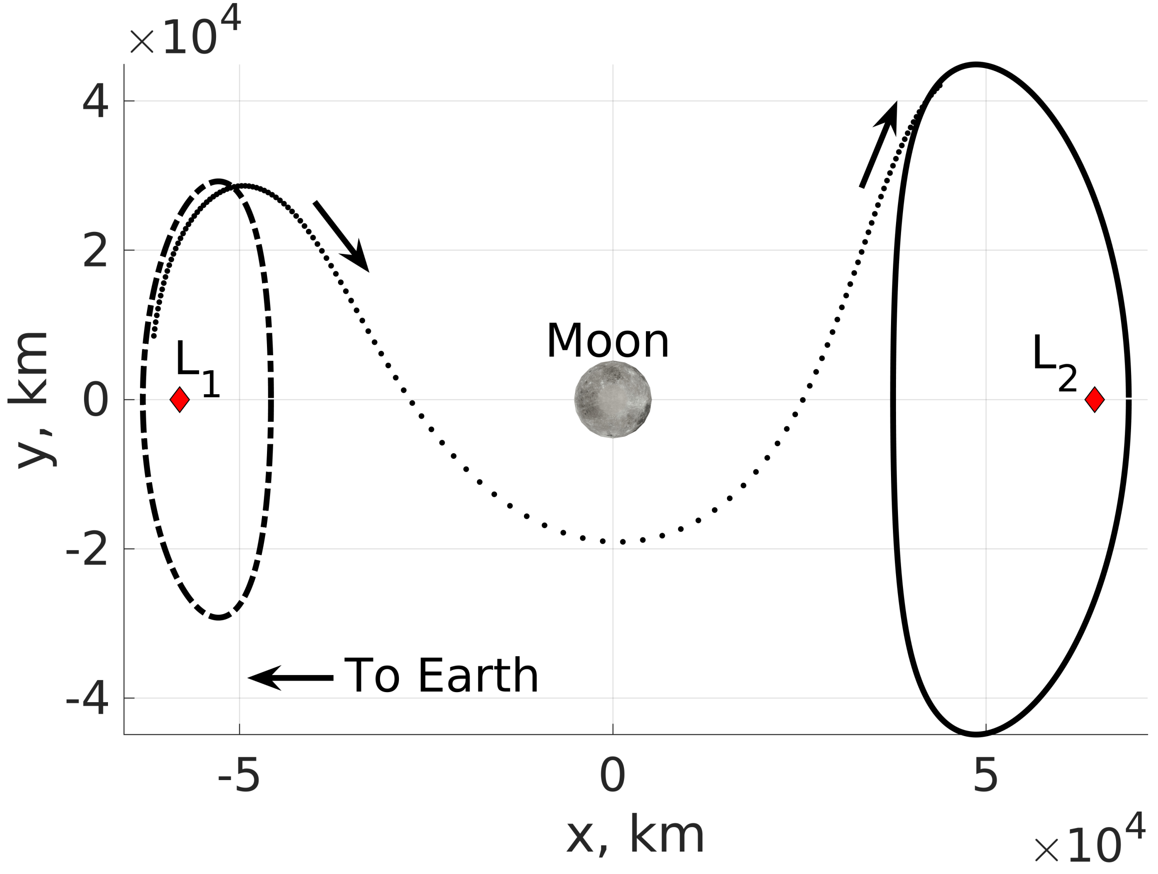
    $$r = -4 - 10|\Delta\bar{d}| - |\Delta\bar{v}| - c_{mass}\Delta m + \Omega$$

  - If closest to final orbit:

    $$r = -100|\Delta\bar{d}| - 10|\Delta\bar{v}| - c_{mass}\Delta m + \Omega$$

  - +1000 bonus for reaching target orbit
  - -1000 penalty for violating constraints
- 4 Policies, c_{mass} = [0, 83, 166, 250]

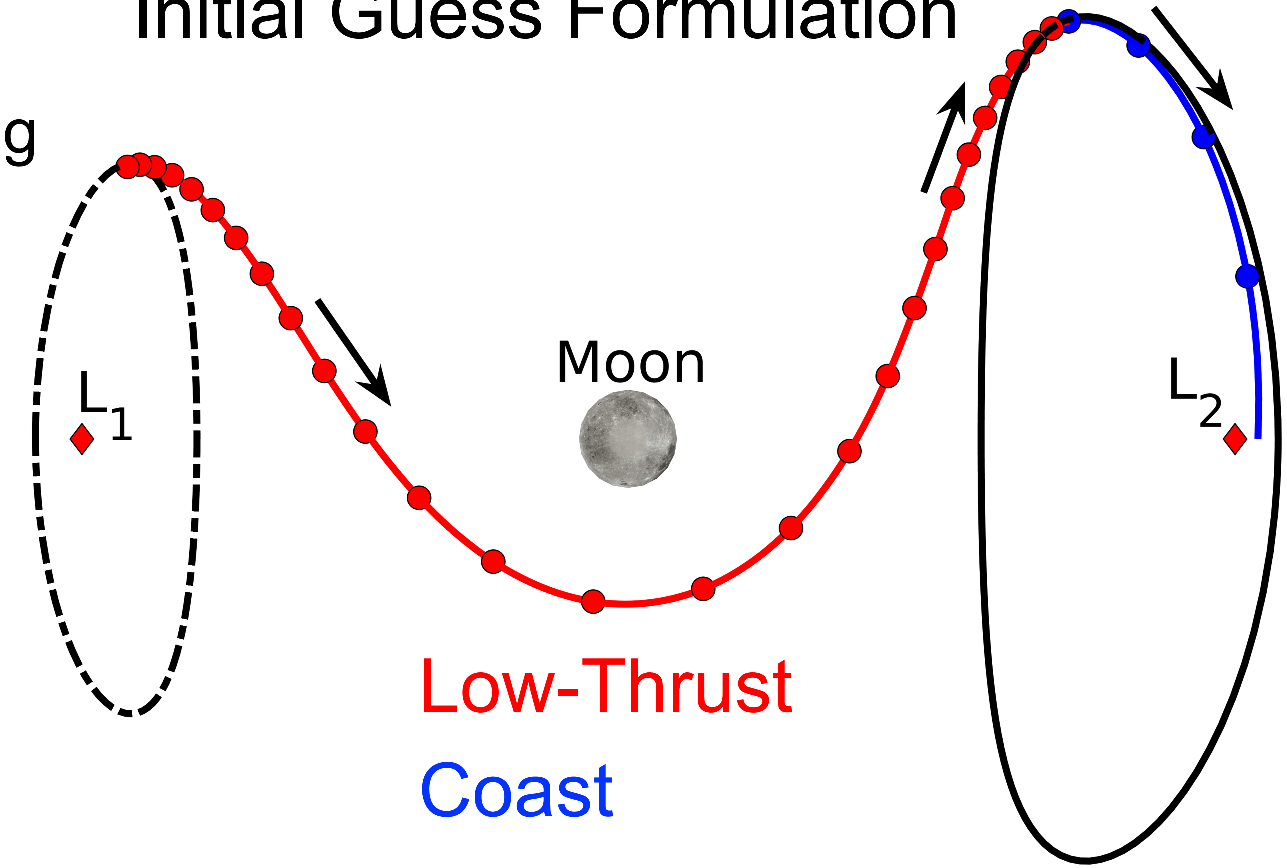
Sample Reference Trajectory



Optimization Scheme

- Reference trajectories from MRPPO serve as initial guesses
- Uses Sequential Quadratic Programming with a multiple shooting algorithm
  - Constraints: State and mass continuity and unit vector thrust direction
- Objective function
  - $J = -m_F$
- Recovers a distinct but related solution space

Initial Guess Formulation

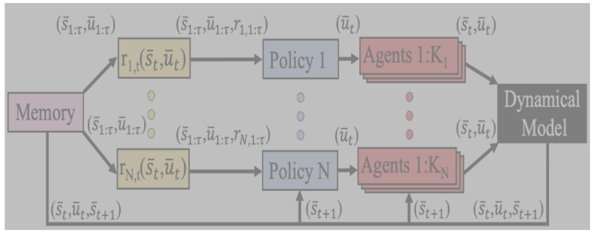


NASA Goddard Space Flight Center | NASA Jet Propulsion Laboratory, California Institute of Technology | University of Colorado Boulder | CCAR

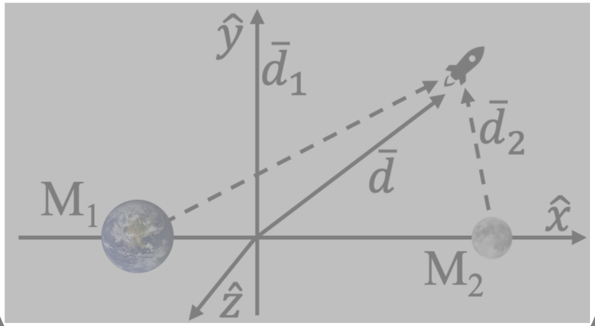
Overview

Technical Background

Reinforcement Learning

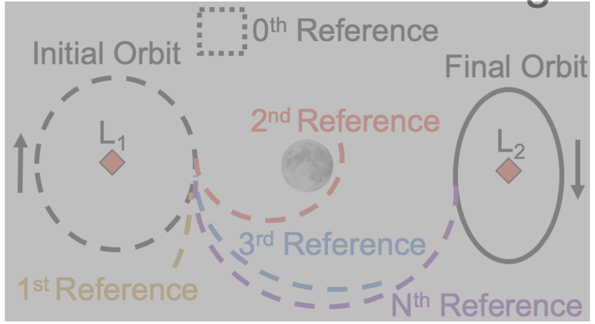


Dynamical Model

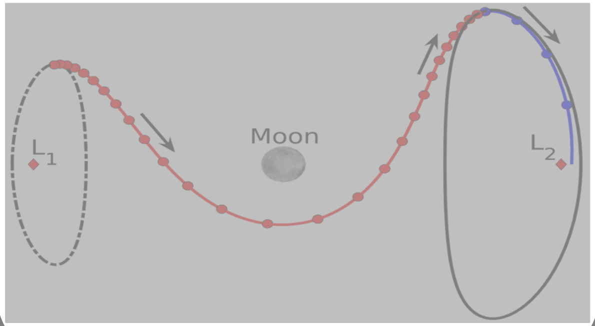


Technical Approach

Reinforcement Learning

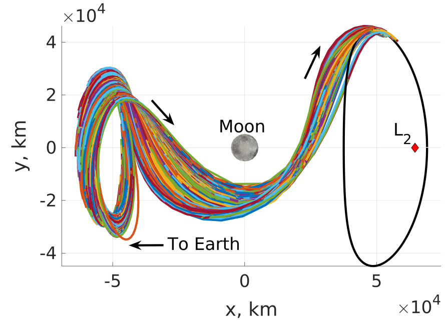


Optimization Scheme

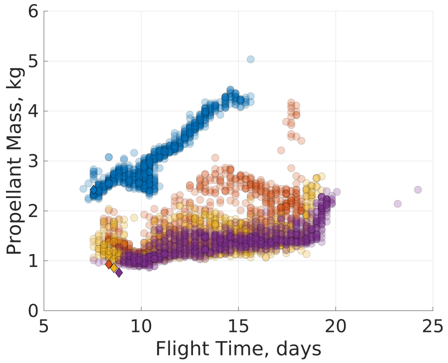


Results

Low-Thrust Transfers

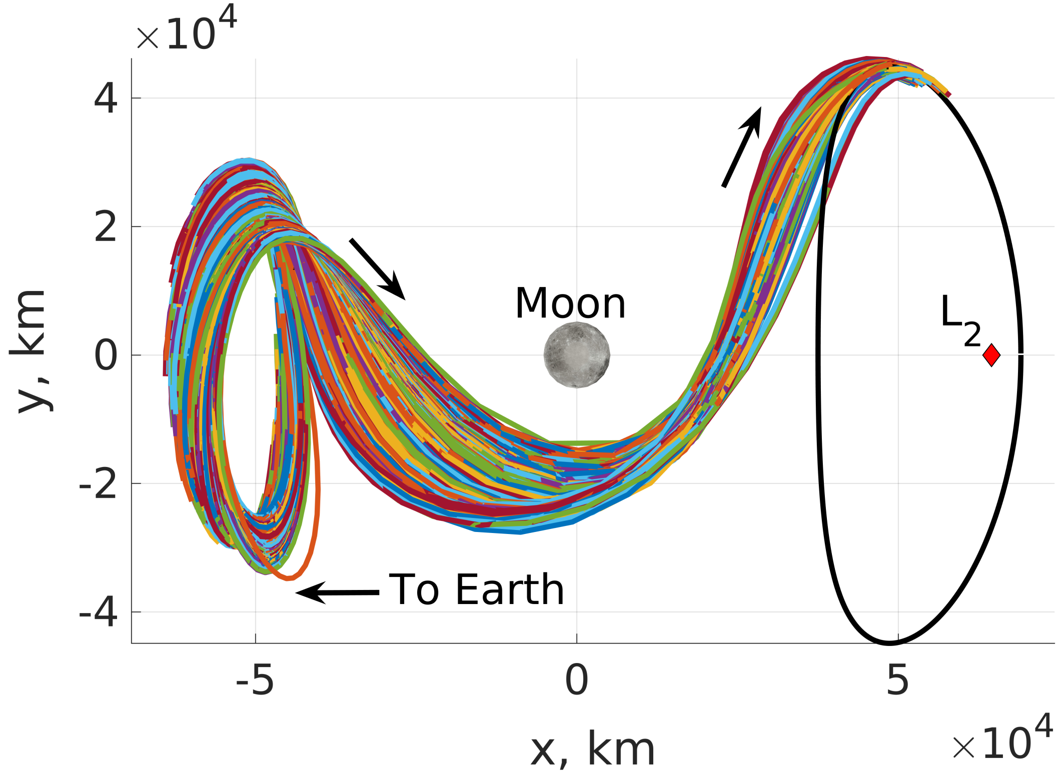


Solution Space

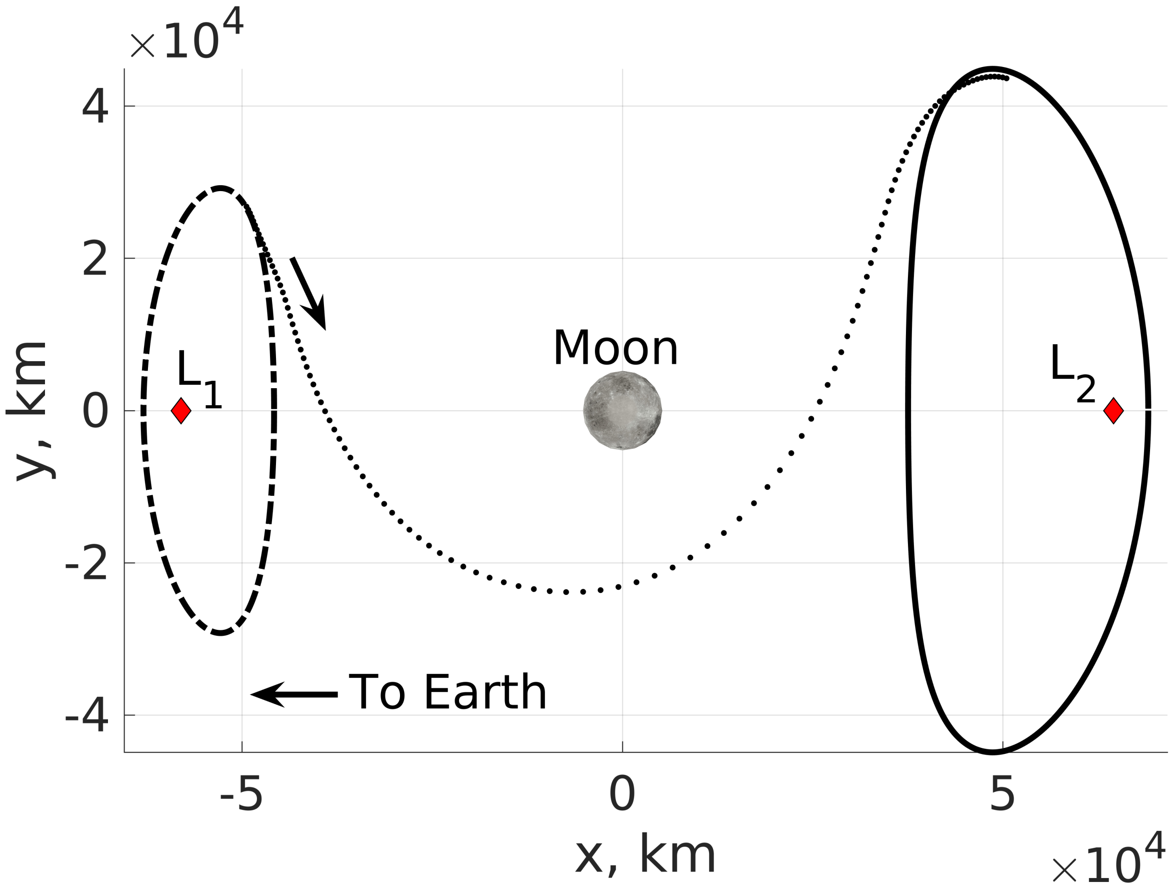


MRPPO with Moving Reference Results

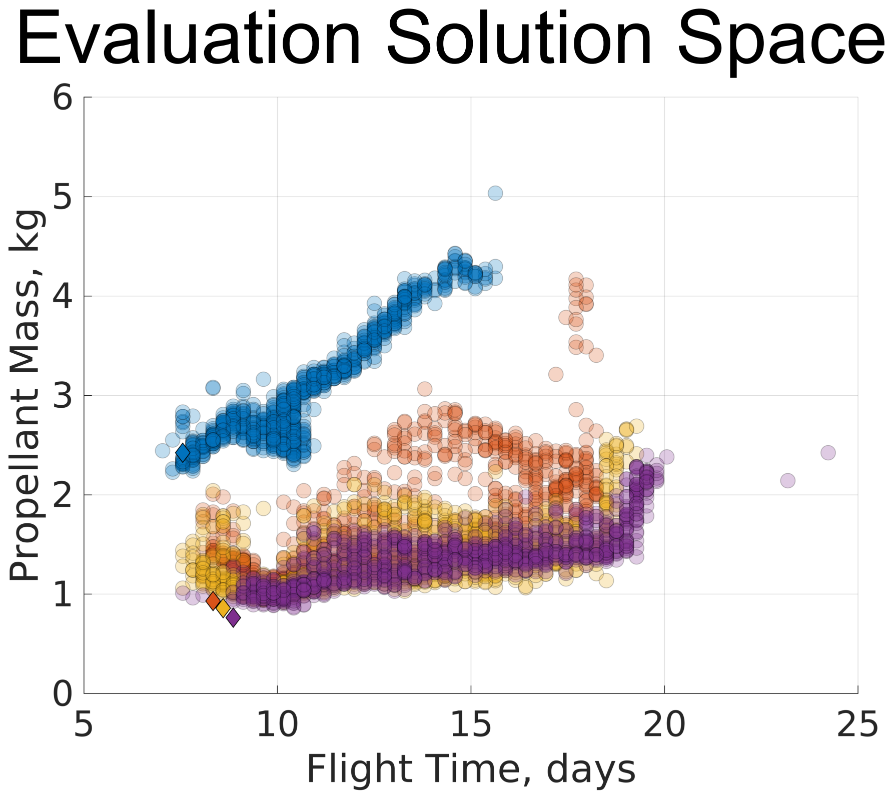
Policy 1 Evaluation Trajectories

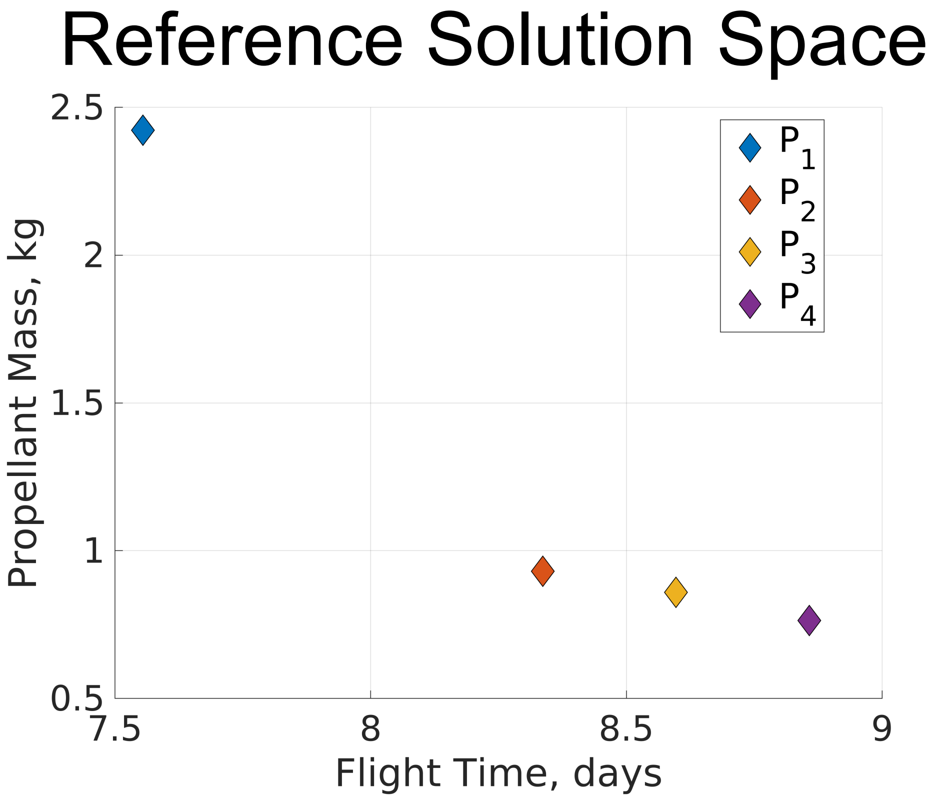


Policy 1 Reference Trajectory



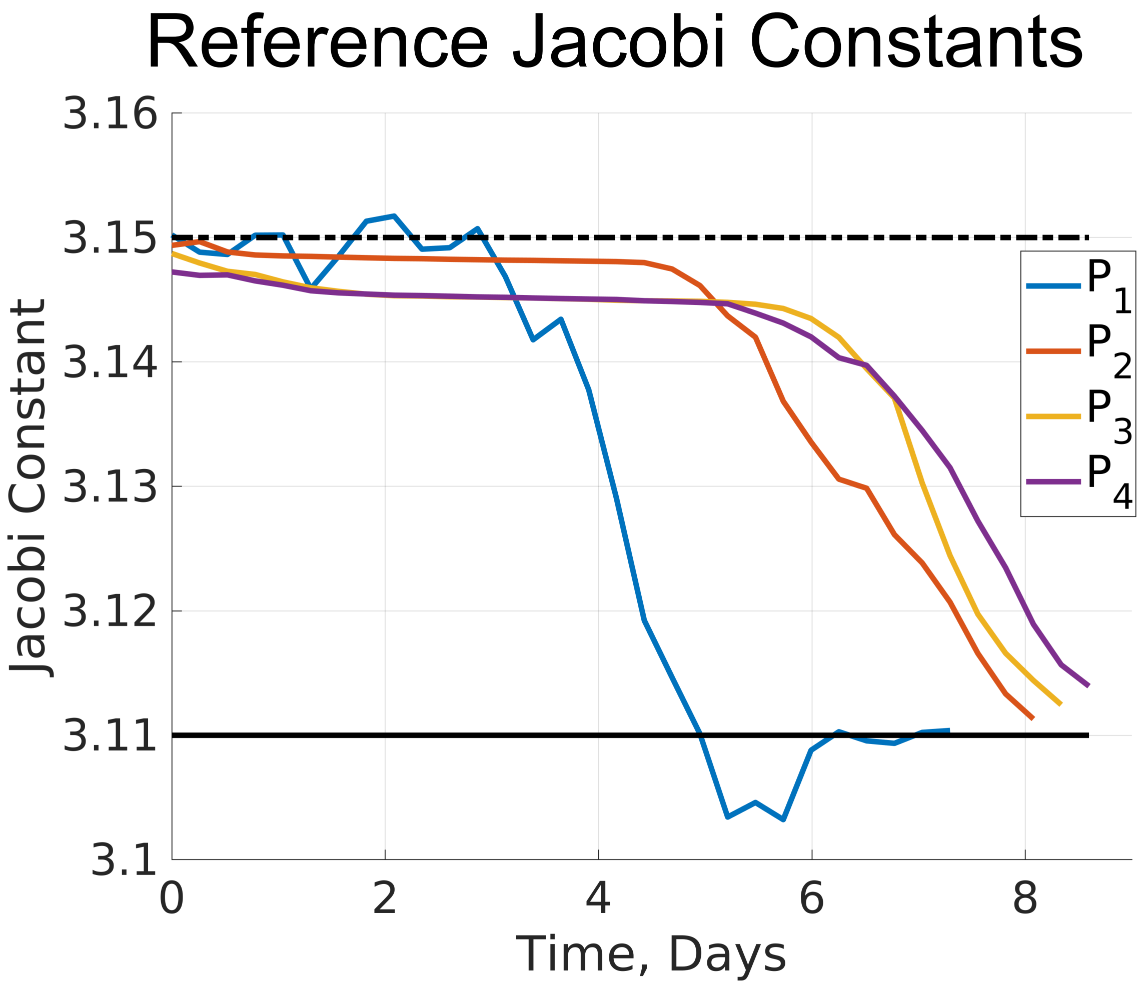
MRPPO with Moving Reference Results

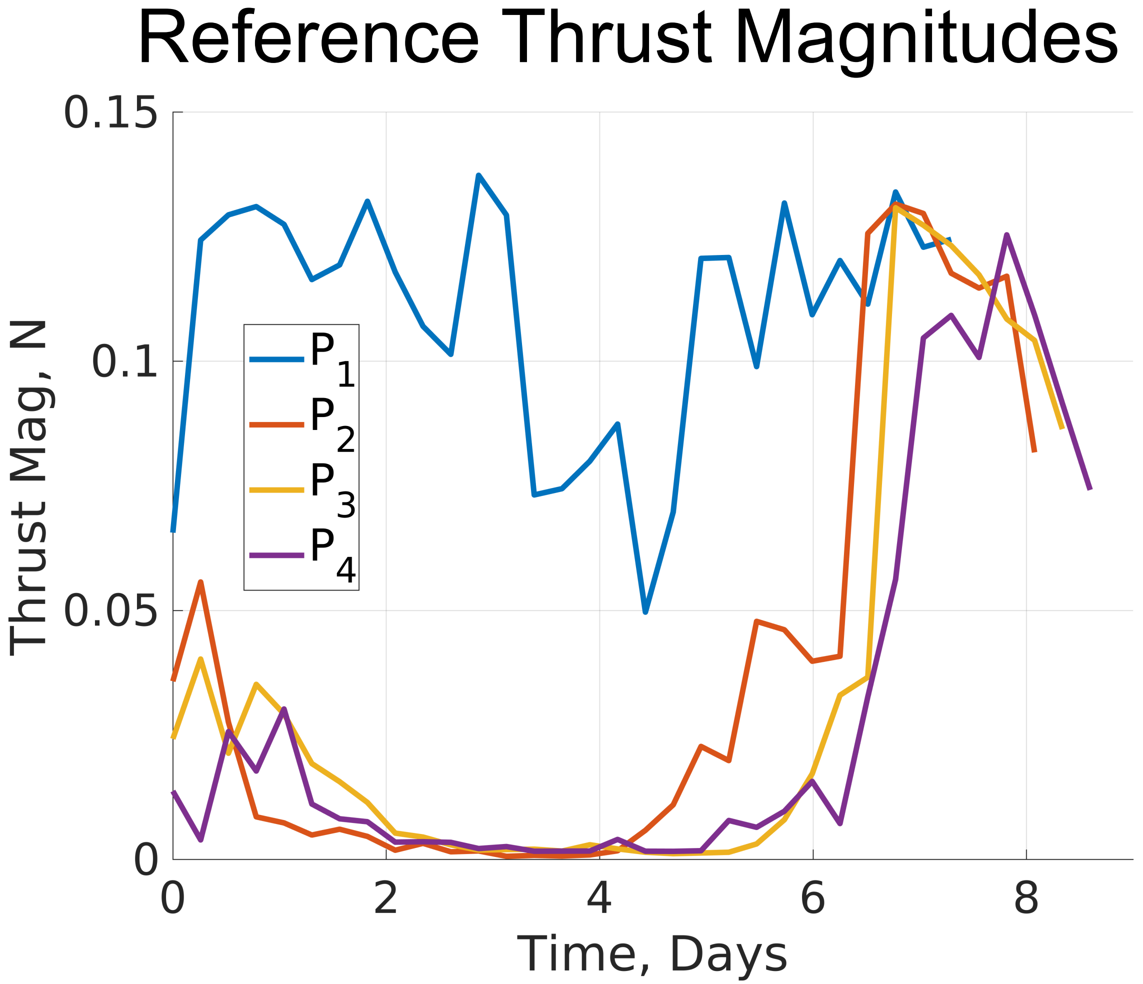




MRPPO with the moving reference modification recovers an array of solutions in the multi-objective space

MRPPO with Moving Reference Results

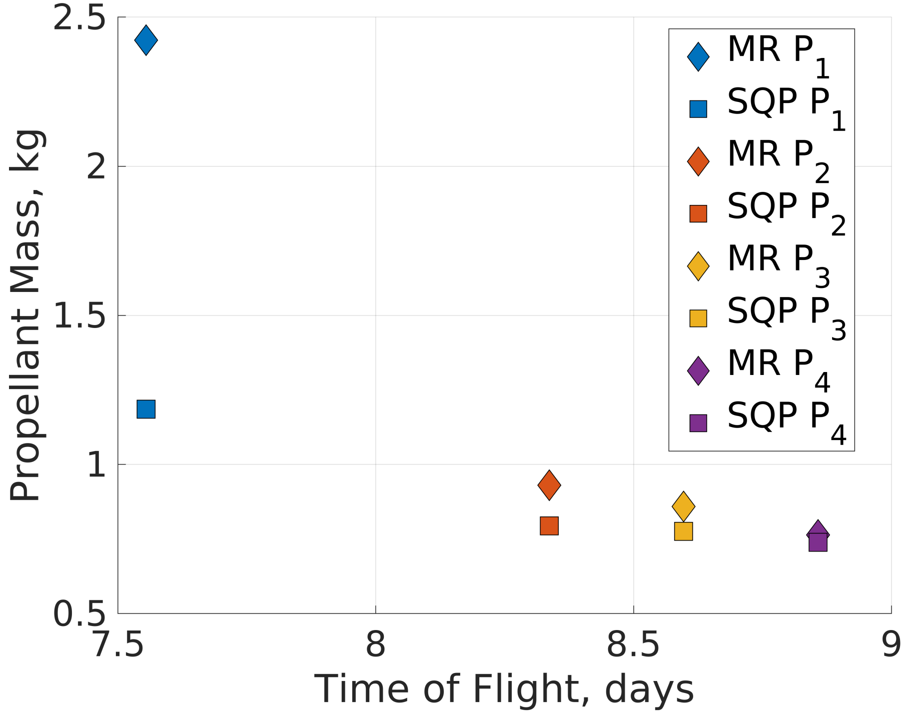




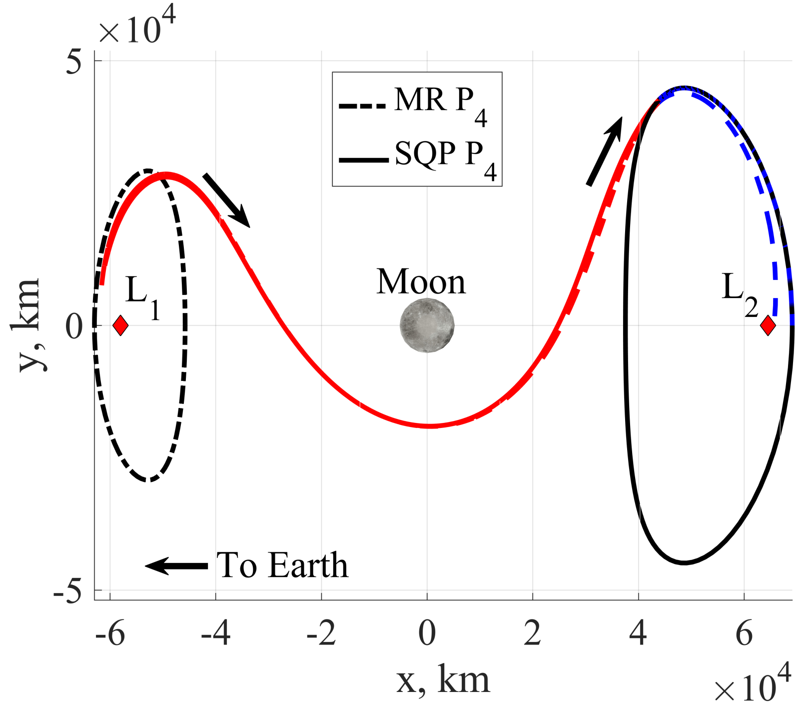
University of Colorado Boulder

Optimization Results

Solution Spaces

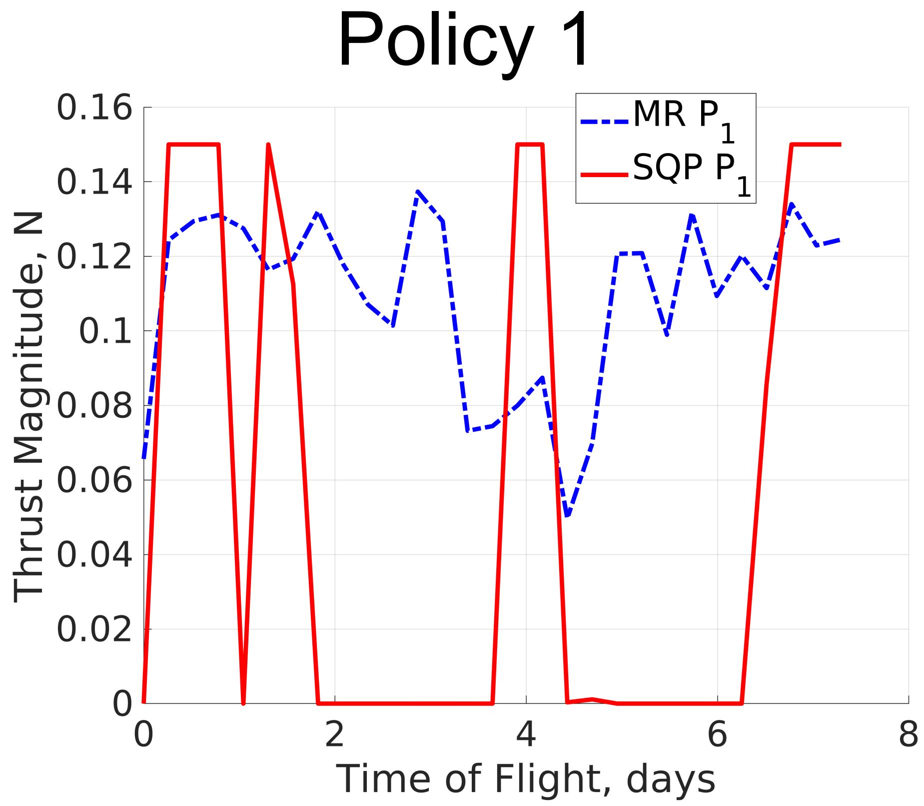


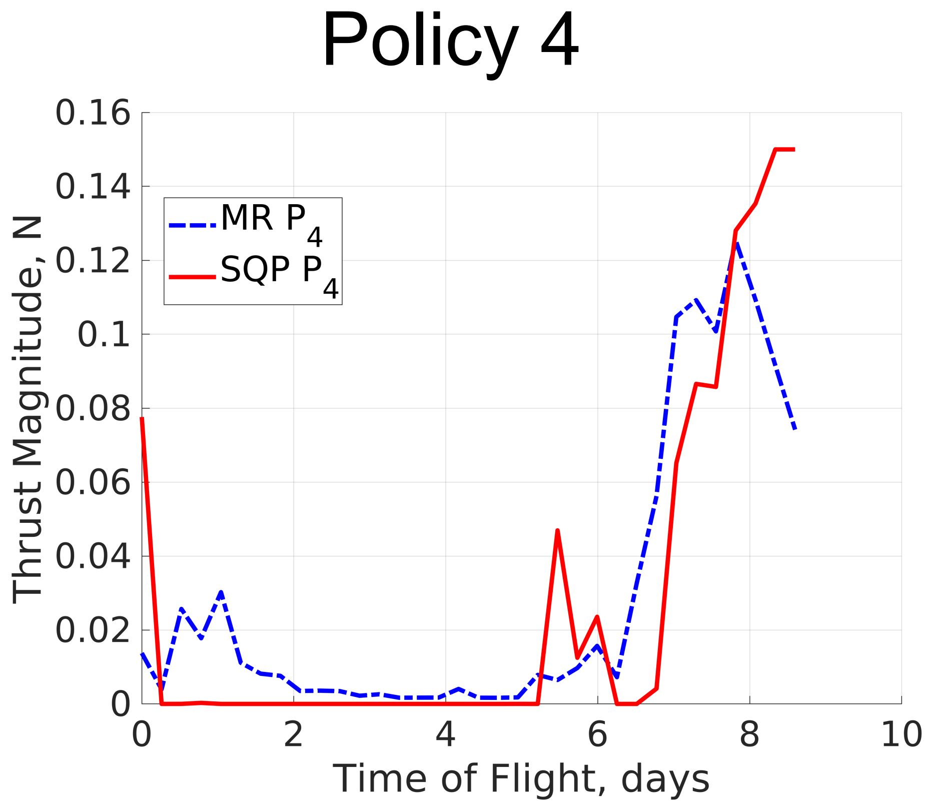
Policy 4 Transfer



Policies with propellant mass usage penalties closely match the transfers produced by the optimization scheme

Optimization Results





Policy 4 produces a thrust profile that closely resembles the thrust profile of the transfer developed by the optimization scheme

Concluding Remarks

- MRPPO with the moving reference modification constructs low-thrust transfers without the need for an initial guess or predefined reference trajectory
- Enables recovery of a subset of the multi-objective solution space demonstrated for a libration point orbit transfer in the Earth-Moon system
- Optimization scheme suggests that the results produced by MRPPO formulation lie close to the local optima of the flight time and propellant mass tradespace

Acknowledgments

- This work was supported by a NASA Space Technology Research Fellowship (80NSSC19K1147).
- Part of this research was carried out at the University of Colorado Boulder.
- Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).
- The High Performance Computing resources used in this investigation were provided by funding from the JPL Office of the Chief Information Officer.

Multi-Objective Reinforcement Learning for Low-Thrust Transfer Design between Periodic Orbits

Christopher J. Sullivan[1], Natasha Bosanac[1],
Alinda K. Mashiku[2], Rodney L. Anderson[3]

[1]University of Colorado Boulder
[2]NASA Goddard Space Flight Center
[3]Jet Propulsion Laboratory, California Institute of Technology