Project Proposal: Reinforcement Learning for Spacecraft Mode Control

March 1, 2018

Abstract

1 Background

The high cost of space mission operations has motivated several space agencies to prioritize the development of "autonomous" spacecraft control techniques [**?**]. Of particular note are on-board autonomy techniques, which serve to expand spacecraft capabilities in environments where ground contact is not possible. While these techniques offer considerable promise, concerns remain regarding a variety of factors surrounding their implementation, such as the cost of developing such systems and their tolerance for un-modeled behaviors[**?**]. Many of these issues are addressed in other fields through the application of machine learning (ML) techniques, which can provide autonomous systems with the ability to learn from and adapt to their environments without human intervention. These techniques, coupled with high-fidelity simulation tools, are a potential avenue for reshaping the manner in which flight systems are developed and fielded.

2 Machine Learning Focus

This project topic is primarily motivated by our group's interest in understanding reinforcement learning techniques and their application to problems in aerospace. However, the scope of the problem at hand includes the possibility of adding inference or classification as additional tasks.

Currently, Bayesian estimation is the de facto method used for spacecraft autonomy. By obtaining the maximum likelihood estimate of a spacecrafts states given a set of measurements, Kalman Filters provide an optimal solution for guidance. Although this is a robust solution, these filters often fall short when it comes to fault detection. They also demand expertise when being developed, modified, or maintained. Reinforced learning methods solve the optimization problem in a new and scarcely used way. The reason for their lack of popularity in the space sector is that data is not plentiful, whereas ML techniques rely on the abundance of data. Yet, with improved simulation methods, runs can now be generated swiftly and with high fidelity, hence yielding the data needed for reinforced learning.

Reinforcement learning is broadly defined as a class of machine learning techniques that maximize the earned reward of a software agent that interacts with a given environment. Through the project scope, we intend to explore the space of both different simulation environments (linear vs. nonlinear models, selection of parameters to model in an environment) and various learning algorithms (model-based vs. model free, Q-learning methods, "deep" reinforcement learning methods).

3 Problem Statement

The system we are attempting to control with reinforcement learning consists of a single spacecraft attempting to enter and maintain an orbit about a planet in deep space. At a high level, this consists of a pair of orbit-entry burns, science operations, and minor corrective impulses periodically to counteract the affect of un-modeled dynamics. It is assumed that the spacecraft has states in which it can conduct the low-level

maths required to accomplish these aims; the goal of our software agent is to command the spacecraft to enter the correct mode under the correct circumstances.

3.1 State Description

Environment States

1. Spacecraft Position: $\mathbf{r} \in \mathbb{R}^3$

2. Spacecraft Velocity: $\mathbf{v} \in \mathbb{R}^3$

3. Reference Position: $\mathbf{r_0} \in \mathbb{R}^3$

4. Reference Velocity: $\mathbf{v_0} \in \mathbb{R}^3$

5. Spacecraft Error Indicator: $E \in \{0, 1\}$

On-Board States

1. State Vector: $\mathbf{X} = \begin{bmatrix} \mathbf{r} & \mathbf{v} \end{bmatrix}^T \in \mathbb{R}^6$

2. State Estimate: $\hat{\mathbf{X}} \in \mathbb{R}^6$

3. State Error: $\mathbf{x} \in \mathbb{R}^6$

3.2 Action Model

This project considers a spacecraft with the following operational modes:

1. **Orbit Determination:** In this mode, the spacecraft uses its sensors to reduce the error in its estimate of its orbit. However, perturbing forces such as solar radiation pressure will drive it away from its desired trajectory.

2. **Orbit Control:** In this mode, the spacecraft uses thrusters to command its current state estimate towards its desired trajectory. We model this as bringing the current state-desired error towards the state-estimate error.

3. **Mission Maneuver**: In this mode, the spacecraft conducts a large burn to enter a transfer trajectory or insert itself in an orbit around a planet.

4. **Science Operations:** In this mode, the spacecraft collects a reward proportional to its state error. However, both its orbit determination and orbit control errors increase with time.

5. **Safe Mode:** if the spacecraft enters an error state (i.e., if the error state variable is 1), entering this mode will reset the error state to 0. No other behavior occurs.

4 Proposed Work Plan

The work for this project is envisioned to proceed as follows:

1. **Linear System:** Consider error states that grow and shrink linearly with time, a reasonable approximation for many spacecraft problems. Attempt 2-3 learning strategies on the linear model.

   (a) Model Implementation: Write the models to implement a linear environment. Wrap these models in a manner suitable for reinforcement learning.

   (b) Model-Free Learning: Implement time-difference Q-learning for the system.

   (c) Model-Based Maximum Likelihood methods: Implement a maximum-likelihood learning method, such as *Dyna* (suggested by Kochenderfer [**?**])

2. **Nonlinear System:** Apply the learned linear models to a "true" nonlinear system and examine their performance. Attempt to apply learning techniques to the nonlinear system.