Optical Sensor Tasking using Monte Carlo Tree Search

by

Samuel J. Fedeler

B.S., North Carolina State University, 2018

M.S., University of Colorado at Boulder, 2021

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Smead Department of Aerospace Engineering Sciences

2023

Committee Members:

Prof. Marcus Holzinger, Chair

Prof. Daniel Scheeres

Prof. Zachary Sunberg

Prof. Nisar Ahmed

Dr. Richard S. Erwin

Dr. William Whitacre

Fedeler, Samuel J. (Ph.D., Aerospace Engineering)

Optical Sensor Tasking using Monte Carlo Tree Search

Thesis directed by Prof. Marcus Holzinger

As space object populations in the near-Earth and cislunar regimes exponentially grow, increased supervision of the near-Earth environment is critical to space sustainability. Limited observational resources must maintain knowledge of space objects in order to avoid collision events, verify maneuvers, and prevent adversarial activities. Observing agents must operate over a large decision space, considering what object to prioritize, when to observe an object, and what sensor to task for observation. The combinatoric nature of such a problem necessitates solutions that may be performed online and tend toward a globally optimal decision.

This thesis develops a framework for generic sensor tasking in the near-Earth environment, with a specific focus on the usage of optical sensor systems. The methodology presented poses sensor tasking as a sequential decision making problem and makes use of Monte Carlo Tree Search (MCTS), a planning algorithm that has been used widely in game-theoretic applications. MCTS is first applied in a centralized manner, with a controller transmitting tasking decisions to a set of observers operating towards a set goal. It is well-known that providing MCTS with effective means to explore the decision space can greatly increase convergence of the methodology; as such, this thesis develops extensive heuristics supported by dynamics and information theory to support initial exploration of the problem. The initial methodology is also supported by asymptotic analysis of MCTS using polynomial exploration.

Heuristics are first developed with goals of maintaining state estimates on a catalog of space objects and performing follow up observation on an admissible region of state space in which a new space object lies. The MCTS tasking methodology is then extended to consider scenarios in which a decision maker wishes to safely track agile space objects. An estimator is developed to better inform a tasker whether an object has recently maneuvered, and search tree exploration is augmented to consider object maneuver potential.

The MCTS methodology is then decentralized, allowing each observer to make decisions in isolation over a reduced decision space with limited communication. Further analysis is performed, considering nonsta-

tionary local reward distributions that arise as agents receive decision-making information from other agents. This methodology is extensible to many-agent coordination problems, with little additional local computational overhead as more observers are considered. A random graph formulation is applied to probabilistically guarantee communication between observing agents.

Finally, the tasking methods built throughout this thesis are applied to a physical scenario in which the VADeR observatory at the University of Colorado at Boulder is tasked to maintain tracks on the local geostationary environment. Space object covariances are maintained and minimized over a multi-night, autonomous observational campaign.

For Harold and Mary Grebe, and for Henry and Mary Fedeler.

Acknowledgements

First and foremost, I want to acknowledge my advisors, Marcus Holzinger and Will Whitacre. Their attentiveness and guidance has not only laid a strong foundation for this dissertation but has also given me a clear sense of direction. Working alongside such kind and caring individuals has made the research process so much more enjoyable.

I also want to thank my friends and fellow graduate students, as I've been lucky to learn and grow in a supportive environment with a bunch of incredible people. I came into the VADeR lab as it started a new chapter in Colorado, and I'm especially grateful to have done so alongside Shez Virani, Daniel Aguilar, and Sam Wishnek. I've also had great opportunities to be challenged by and collaborate with good friends, and I especially thank Damennick Henry and Jesse Greaves for acting as reviewers, brainstormers, and exceptional peers over these last few years.

Finally, it is imperative to recognize that support for this dissertation extends far beyond my time in Boulder. My grandparents, to whom I have dedicated this work, have instilled in me the value for education. Their encouragement also shaped my parents' pursuit of graduate degrees and inspired my mother, Carla, to become a lifelong educator. My mother taught me from an early age to learn for curiosity's sake, and perhaps my eventual career direction was spurred by Joe, the engineer, homebrewer, and 3D printer; I think them both for their support and the purpose I've found in learning. From an early age, my sister, Catherine, has also been a constant source of support and inspiration; I thank her for putting up with an annoying big brother. Lastly, I thank my partner, Maggie, for her love and understanding as I've approached the finish line of this dissertation. Your belief in me has been a constant source of motivation, and I am deeply grateful for you.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Sensor Tasking and Space Domain Awareness

Space domain awareness (SDA) may be defined as "actionable knowledge required to predict, avoid, deter, operate through, recover from, and attribute cause to the loss and degradation of space capabilities and services" [55]. To maintain and gain knowledge of the local space environment, ground and space-based observers must make timely observations, and critically, decisions on what to observe. The number of space objects (SOs) for which to maintain tracks is quickly increasing, especially in oft-used regions such as the geostationary belt. The European Space Agency maintains tracks on almost 50000 objects as of 2019 [31], and this figure is expected to greatly increase in the near term as constellations in low Earth orbit such as Starlink are deployed. The local space environment has also grown in an economic sense, and by NASA estimates, the space economy has expanded by 60 percent in the last decade, with a current value of approximately 400 billion US dollars. Continued growth is expected, and operational needs must be considered in the context of choke points and scarcity. At low Earth orbit (LEO), exponential growth in object populations greatly increases probability of collision events [4]. A growing concern in the geostationary region (GEO) is the limited availability of orbital slots necessary for satellites to be resolved by ground-based observers and radio bands to maintain custody [45]. Above geostationary orbit (XGEO), many lines of access to the near-lunar regime must transit near the Earth-Moon L1 Lagrange points. Such challenges motivate the careful maintenance of SO tracks in order to further scientific, commercial and governmental endeavors in space.

Space objects are most commonly tracked using radiometric and optical sensing systems. Operators from each sector with a stake in the space environment actively track space objects, and the most prominent

example of a sensing architecture is the United States Space Surveillance Network [97], which operates a portfolio of ground-based radar arrays and telescopes alongside a set of space-based sensors. In general, the global portfolio of observing assets has not kept pace with the growth in space objects. This problem must be addressed by tasking sensors efficiently in regards to objectives such as covariance minimization, acquisition of custody, or maneuver detection. The sensor tasking problem is combinatoric, with a decision space that grows as a function of time, the number of observers considered, and the number of SOs that are tracked.

In low Earth orbit, the sensor tasking problem is somewhat easier to address. A wealth of ground-based radar sensors have been deployed for LEO space object tracking, and phased array radars admit the capability of tracking multiple objects simultaneously. Sub-kilometer state uncertainties are achievable using commercially available data [83]. Largely as a result of power requirements, radar is challenged by more distant targets, and GEO space objects are most commonly tracked using optical sensors.

Also of interest when considering the sensor tasking problem is application to the cislunar regime of space. Relatively little literature has been produced on the subject, and the region is expected to be a growing frontier for space exploration in coming years [54, 14]. As volumes of space further from Earth are considered, dynamic complexities are introduced, and it is no longer feasible to neglect lunar and solar perturbing forces. Trajectories in the cislunar regime are often unstable, and many initial conditions are chaotic even when the circular restricted three-body simplification [87] is applied for analysis. Furthermore, traditional ground-based sensors, even optical sensors, become challenged by the distances of XGEO objects, and observation of an XGEO throughout the Earth-Moon synodic period is near-impossible. Even if an object is observed, a comparative lack of diversity in state information from measurements reduces the quality of orbit determination. These factors motivate the addition of space-based sensors to observing portfolios.

Methodologies driving sensor tasking are traditionally broken into tractable subproblems, in which the objective is to capture a single aspect of the overarching goal. Often, one may wish to generate new state estimates, expanding the set of SOs studied by searching for natural objects, orbiting satellites, or debris. Wide-ranging techniques for this objective exist in literature. Classical methodologies such as striping may be introduced as a pure search scheme. This strategy is shown to be effective for GEO maintenance and search

[2, 88], using either a targeted field through which SOs are allowed to drift or declination striping and multi-stripe raster scanning. Frueh further considers tasking for GEO search as an optimization, demonstrating significant improvements on to striping strategies [39]. Search strategies are useful for large populations near Earth, but largely assume two-body dynamics and become less applicable over long horizons in the cislunar regime. Little literature has considered the search process in the XGEO regime, with a variety of initial orbit determination methods considering associated observations [112, 118].

Alternatively, one may wish to maintain existing estimates, informing knowledge on a catalog of SOs. A variety of strategies have been proposed assuming a priori knowledge on state estimates and uncertainties. Erwin et al apply linear optimization to form a tasking solution and propose useful quantities for interpreting the value of a tasking decision [30]. This work is extended by Williams et al, using Lyapunov exponents to probe the stability of SO estimates [111]. Hill et al. further consider covariance as a tasking method and specifically illustrate the utility of covariance-based tasking for reduction of uncertainty in position, velocity, or semi-major axis [52]. A variety of approaches have also taken inspiration from the machine learning literature, with techniques such as stochastic gradient ascent [100], asynchronous actor-critic methods [69], and proximal policy optimization [91]. In each of these methods, the driving goal is determination of an optimal policy for decision making given a large set of candidate observations. Methods might consider custody maintenance from the perspective of hypothesis resolution [58], and especially powerful approach for purposes of anomaly and maneuver detection. Similarly, application of interacting multiple models [44] or incorporation of control policies into estimation [74] has proven useful to maneuver estimation alongside catalog maintenance.

It is worthwhile to briefly address methodologies that combine this major objectives. Such methodologies must necessarily incorporate multiple objective optimization (MOO) strategies such as NSGA-II [28]. While the research presented by Jaunzemis [58] is multi-objective, a priority weighting scheme is applied to scalarize the problem. Gehly et al. enforce catalog maintenance when expected information gain exceeds a desired threshold, otherwise performing search objectives and again scalarizing the full solution space [42]. Cai et al. apply NSGA-II alongside multi-target tracking for GEO search and maintenance, but in results consider subsets of the full Pareto front, demonstrating the challenges still present in selecting a policy from

front of non-dominated solutions.

Another open question in sensor tasking is how to best coordinate many decentralized observers in an optimal manner. A variety of methods in literature consider multiple sensing agents [100, 38, 18] but operate in the sense that a centralized planner commands each agent in the problem. To maximize resiliency, autonomy, and realism, it is critical to eventually decentralize the tasking architecture, especially in response to emerging events such as maneuvers, conjunctions, or anomalies. The subject has briefly been addressed by the SDA community [19], but the breadth of literature in the multi-agent reinforcement learning community has not previously been adapted to SDA sensor tasking problems.

The major goal of this thesis is to extend autonomous sensor tasking to address such open questions. The backbone of the methodology presented in this research is Monte Carlo Tree Search (MCTS) [63, 26], an online planning methodology applied as an extension of multi-armed bandit selection policies [5]. A detailed overview of MCTS and the problems it addresses is outlined in Chapter 2. The methodology, much like methods previously discussed that apply reinforcement learning techniques, aims to apply immediate decisions that are optimal over a long time horizon. MCTS may then be extended in a manner that enables decentralized SDA sensor tasking for catalog maintenance, follow-up observation, and maneuver detection. Such algorithms support autonomous decision making for decentralized, space-based observers as satellite populations expand in the GEO and XGEO regimes.

Thesis Statement: Monte Carlo Tree Search may be applied to optical sensor tasking to quickly reach near-optimal solutions over long time horizons, improve custody maintenance, and decentralize decision-making, fundamentally improving autonomous space domain awareness and observer efficiency.

1.1 Sensor Tasking for Catalog Maintenance

In order to maintain known populations of SOs, a planning methodology must be applied that maintains custody of each object and minimizes state uncertainties over long time horizons. The aim of this contribution is to introduce MCTS as a planning methodology that supports such goals, with specific results motivated by plans to establish long term presence in the XGEO regime [96]. Monte Carlo Tree Search utilizes stochastic exploration of potential decisions over a long time horizon or until the termination of a

problem to constrain the long term value of an immediate action. In this case, it is logical to define an action at an epoch as the SO an observer chooses to track. This decision space can quickly grow quite large; in a case with M observers, N objects, and T epochs, the number of possible decision sequences is $N^{M\times T}$ and subject to combinatorial explosion. In any sensor tasking solution, a means for approximating valuable decision sequences in this combinatoric space is critical to efficient use of observers.

It is most common to consider the catalog maintenance problem in an information sense, utilizing knowledge of the underlying dynamical system and the applied estimators. Classically, these features are applied by sequentially prioritizing covariance features [51, 30] or the immediate nonlinearity of the dynamics [111]. Such methods are often myopic, and more recent literature has recognized a critical feature of the catalog maintenance problem is when to observe a candidate object that is high value in such metrics. Often, for reasons as such as an object being occluded, nearing a threshold probability of detection, or achieving a future close approach, there is a decision trade space in time. These factors motivate more recent methodologies incorporating decision making over time horizons, in particular the use of the Markov decision process (MDP) formalism. Discussed further in Chapter 2, an MDP applies stochastic transitions between states, which fully encapsulate current information, over which an actor may make sequential decisions, receiving a reward at each epoch. The goal of an MDP is to maximize long-term reward. Most prominently, the reinforcement learning paradigm utilizes MDPs as the underlying problem setting, and such literature has been incorporated into the catalog maintenance problem [70, 91]. Alternatively, MCTS is commonly applied to solve MDPs in an online, anytime manner, with great success in partially observable [101] and continuous [67] environments. As such, it is an excellent candidate for the decision space presented in the catalog maintenance problem.

Contribution 1: Monte Carlo Tree Search is established as an optimally convergent planning methodology for multisensor tasking in cislunar space, solving the catalog maintenance objective of SDA sensor tasking.

1.2 Sensor Tasking for Follow-up Observation

When searching for new SOs, it is also important to note that detections made with optical sensors generally do not fully observe the object state; as a result of this "Too Short Arc" problem, an admissible

region (AR) [78] of unobservable ranges ρ and range rates $\dot{\rho}$ may be formed. This admissible region is a two-dimensional manifold of feasible pairs $(\rho, \dot{\rho})$ that may be projected into the six-dimensional state space. Note that this region may be uniformly distributed in range and range rate or probabilistic [116] if measurement uncertainty is incorporated. Gehly et al. leverage the AR methodology in tandem with Finite Set Statistics to approach the tracking problem, representing the admissible region as a Gaussian mixture to be ingested by a CPHD filter [42]. Methodologies for generating Gaussian mixture representations of admissible regions are introduced by DeMars and Jah [29]. AR pairs over longer observation intervals may be used for initial orbit determination [40], [89]. These methodologies are not typically used in an online manner, but rather consider large populations of admissible regions generated from detected tracklets over several observation campaigns. Admissible regions may also be utilized for follow-up observation, where a sensor may search over the AR or another extended projection in measurement space at some point after the full state becomes observable [80, 53].

Several other scenarios are common in which the follow-up observation problem becomes nontrivial. For example, an object may be hypothesized to have made a maneuver at some prior epoch leading to loss of custody. The region of state space the object can reach over a large time horizon, a reachable set, can then be projected into measurement space in a similar manner to an admissible region. Alternatively, one may desire to track an object that already has a large uncertainty. When projected into measurement space, this uncertainty is larger than the sensor field of view, and multiple observations must be taken to ensure the object is detected.

In any of these scenarios, it is desirable to optimally exhaust the feasible region in which an object lies. To do so, the follow-up observation problem may be treated as a sequential decision making process, and Monte Carlo Tree Search can be demonstrated successfully in this regime. Monte Carlo Tree Search ensures efficient generation of viable action trajectories, with the decision space considered as the binned subset of measurement space to exhaust during search.

It is also critical to apply efficient estimation schemes alongside a follow-up tasking methodology. Of particular interest is the impact of a missed or null detection in one region of measurement space on the probability density arbitrarily far away, an effect commonly seen in multi-target tracking filters [36]. A logical

extension of this effect, then, is determining how a null detection in a subset of the projected feasible set may affect knowledge on other regions within that volume. In particle filters, null detections are somewhat trivial to incorporate, but novel methods may be applied to Gaussian sum filters.

Contribution 2: Efficient techniques for tasking and estimation are formulated for constraining a prior state estimate with large uncertainties in a sensor field of view, supporting search and track instantiation needs in SDA.

1.3 Sensor Tasking for Maneuver Detection and Estimation

This challenges outlined insofar must be extended to further account for maneuvering SOs, and a variety of estimation techniques may feasibly be introduced to avoid filter divergence in maneuvering scenarios. Tuning methods for process noise such as state noise or dynamic model compensation can be used to increase estimator robustness [103], and given that maneuvers may occur on vastly different scales, any technique that is used must adapt to observed features such as measurement residuals that are impacted by maneuvers. A filter that incorporates control-theoretic considerations is especially useful, and Greaves recently demonstrated the utility of the U-OCBE for nonlinear problems [46]. The U-OCBE is shown to be especially useful for maneuver detection and classification, delivering information that can augment decision-making capabilities.

In prior contributions, the MCTS methodology is driven by information-theoretic heuristics for decision making such as Kullbeck-Liebler divergence or the trace of measurement innovation. The introduction of maneuvering trajectories obfuscates the decision making process in that goals become twofold. Covariance minimization is still a priority, but a successful tasking methodology must also ensure maneuvers are detected and estimated autonomously, without filter divergence. The U-OCBE is incorporated to inform knowledge on the occurrence and type of prior maneuvers [46], but in addition, it is desired to observe SOs near epochs when maneuvers will have tangible effects in the long term, especially for maneuvers departing the nominal trajectory. A wealth of literature has studied local effects of perturbations from periodic orbits in the Earth-Moon restricted problem, with Halo families probed in detail as early as the 1960s [17]. Largely, local analysis makes use of the state transition matrix, and recent stationkeeping literature has extended

this analysis to make use of the Cauchy-Green Stress Tensor (CGT) as a means to probe the distance an initial perturbation may stretch or compress from the nominal trajectory [48, 79]. This analysis offers utility for both stationkeeping and departure maneuvers, and may be applied alongside maneuver estimation to further inform maneuver potential.

Contribution 3: Methods are developed for adapting MCTS to decision problems considering agile targets, augmenting sensor tasking in nonlinear environments.

1.4 Decentralized Sensor Tasking

Largely, the methodologies presented are designed for single agent problems, but in practice, many applications necessitate coordination between multiple agents. MCTS and single-agent RL methods may be applied, treating other agents as part of the environment, but a robust methodology should account for the behavior of other agents in the problem. The multi-agent decision problem has thus been increasingly studied in recent years, with research largely built on the Markov game framework [57]. Settings are characterized as cooperative, in which agents share a common reward, competitive, in which the reward for one agent is exactly the loss of another, or a general-sum game, where little restrictions are placed on relationships[117]. A variety of approaches in each context have been discussed in literature, including multi-agent Q-learning [57], deep multi-agent RL [106], and policy gradient methods [35].

In the cooperative context of a multi-agent decision problem, it also becomes key to design an algorithm that allows for efficient communication. Communication-efficient methods have been proposed for actor-critic [68] and deep RL methods [60], and MCTS-based approaches have applied techniques such as coordination graphs [3, 23]. As new methods are developed, communication efficiency must also be considered, and while Monte Carlo approaches consider what agents should communicate with each other, there is a shortfall in considering how often said agents should communicate with each other. Agents may need to manage a limited communication budget, and communication methods may need to be robust to loss or denial of communication edges. This contribution aims to address the second case, considering robust communication at consistent intervals, outlining a process that could then be augmented by planning-aware communication methods.

The decentralized MCTS methodology presented in this contribution is then applied to the context of space domain awareness. The methods presented in this contribution extend prior studies, ensuring agents may operate in a decentralized manner with limited communication. Reduction of the action space locally to that of a single agent also ensures scalability of the methodology to the many-agent decision-making problem. Decentralized MCTS is demonstrated for geostationary and cislunar catalog maintenance and shown to be robust to maneuvers and large gaps in communication.

Contribution 4: MCTS is decentralized for many agents communicating over random graphs, improving SDA sensor tasking autonomy and robustness.

1.5 Contributions and Outline

The contributions in this thesis outline the applicability of MCTS to SDA sensor tasking, increasing sensor efficiency and autonomy. In Chapter 2, mathematical formalisms are established for MCTS problem settings and the sensors and dynamical systems used throughout the thesis are defined. Chapter 3 introduces two centralized methodologies for the catalog maintenance and follow-up observation SDA sensor tasking objectives. Chapter 4 extends the catalog maintenance objective to the problem of maneuvering space objects, and Chapter 5 further decentralizes this decision making problem. Finally, Chapter 6 presents application of the decentralized tasking methodology to the Vision, Autonomy, and Decision Research (VADeR) Observatory alongside methods developed for full autonomous operation of the observatory.

1.6 Thesis-Related Publications

The research presented in this thesis has appeared previously in a variety of journal articles and conference presentations. A full itemization of presented and submitted writing is outlined below.

1.6.1 Journal Articles

1. Fedeler, Samuel, Marcus Holzinger, and William Whitacre. "Sensor tasking in the cislunar regime using Monte Carlo Tree Search." *Advances in Space Research* 70.3 (2022): 792-811.

2. Fedeler, Samuel J., Marcus J. Holzinger, and William W. Whitacre. "Tasking and Estimation for Minimum-Time Space Object Search and Recovery." The Journal of the Astronautical Sciences 69.4 (2022): 1216-1249.

3. Fedeler, Samuel J., Marcus J. Holzinger, and William W. Whitacre. "Cislunar Space Object Tracking Considering Maneuver Estimation and Maneuver Utility." *Journal of Guidance, Control, and Dynamics* **submitted April 2023.**

4. Fedeler, Samuel J., Marcus J. Holzinger, and William W. Whitacre. "Decentralized Decision Making over Random Graphs for Space Domain Awareness." *Advances in Space Research* **submitted May 2023.**

1.6.2 Conference Papers

1. Fedeler, Samuel, and Marcus Holzinger. "Monte Carlo Tree Search Methods for Telescope Tasking." AIAA Scitech 2020 Forum. 2020.

2. Fedeler, Samuel J., Marcus J. Holzinger, and William Whitacre. "Optimality and Application of Tree Search Methods for POMDP-based Sensor Tasking." Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference. 2020.

3. Fedeler, Samuel J., Marcus J. Holzinger, and William W. Whitacre. "Optimally Convergent Minimum-Time Space Object Search and Recovery." 31st AAS/AIAA Space Flight Mechanics Meeting, 2021.

4. Fedeler, Samuel, Marcus Holzinger, and William Whitacre. "Spooky Coordinated Tasking and Estimation on Uninformative Priors." Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference. 2021.

5. Fedeler, Samuel, Marcus Holzinger, and William Whitacre. "Lunar Observer Efficacy for NRHO Target Tracking." International Astronautical Conference. 2022.

6. Fedeler, Samuel, Marcus Holzinger, and William Whitacre. "Decentralized Decision Making over Random Graphs." I 26th International Conference on Information Fusion. Accepted, 2023.

7. Fedeler, Samuel J., Marcus J. Holzinger, and William W. Whitacre. "Optimally Convergent Autonomous and Decentralized Tasking With Empirical Validation." Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference. Accepted, 2023.

Chapter 2

Preliminaries

A brief mathematical overview for concepts critical to this dissertation is now given. First, a conceptual overview of Markov Decision Processes and Monte Carlo Tree Search is provided. Next, an outline is provided of the estimators, dynamical systems, and measurement models utilized throughout this thesis.

2.1 Sequential Planning

This thesis relies heavily on concepts from the game theory literature, and the following section provides a review of planning methodologies that support the thesis contributions. In addition to this review, suggested readings include [21, 26, 63, 92, 101]. MCTS and any variants may be applied to sequential decision processes, and typically, a problem is described as a Markov Decision Process (MDP). In an MDP, the underlying states are fully observable, but in practice, this is often not the case. As such, it is also useful to discuss Partially Observable Markov Decision Processes (POMDPs), and this thesis explores contributions in both contexts. Finally, both MDPs and POMDPs assume that a single agent operates on the problem, but it is desired to find solutions for problems in which many agents are coordinated. The generalization of such cases is typically labeled a Multiagent MDP (MMDP) or Markov game.

2.1.1 Markov Decision Processes

A Markov Decision Process is formally described by the 5-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$, where

- $\mathcal{S}$: the probabilistic state of the environment.
- $\mathcal{A}$: the action space over which agent(s) act.

- $\mathcal{T} : \mathcal{S} \times \mathcal{A}$: a transition function between states over time conditioned on global sets of actions.
- R: a scalar reward function for the agent(s).
- $\gamma \in [0, 1]$: the discount factor over time, impacting the prioritization of short term rewards.

The problem is defined over a state space $\mathcal{S} \in \Re^n$; this is a representation of a discrete or continuous space in which the studied system may evolve, where n is the dimension of the state. Decisions are made over an action space $\mathcal{A} \in \Re^p$; again, this space may be either discrete or continuous, with dimension p. States evolve with transition probabilities $\mathcal{T} : \Re^n \times \Re^p \to [0, 1]$. In this dissertation, $\mathcal{T}$ often represents the propagation of space object states and uncertainties over time, but actions taken may also impact the evolution of estimated states. Finally, a reward $R : \Re^n \times \Re^p \to \Re^1$ applies a general objective function to determine the value of an associated change in state and action. The key aim of solving a MDP is the evaluation of a policy π that probabilistically describes the actions an agent should take at the current state to maximize the expected cumulative discounted reward

$$R_\infty = E\left[\sum_{t=0}^{\infty} \gamma^t R_{a_t}(\mathbf{s}_t)\right]. \tag{2.1}$$

This notion of long term reward considers both the immediate value of an action and, with discount, the future returns achieved from the state that one transitions into. To generate policies, a MDP solver often estimates the state-action value function $Q(s, a)$, or the expectation of the discounted cumulative reward in state s associated with taking action a, at each state. With good estimates on the state-action value function across the state space, one can then choose to proceed with actions over time that maximize the expected cumulative reward.

Traditionally, dynamic programming approaches are applied to solve MDPs, with the goal of leveraging the Bellman equation for optimality

$$V_*(\mathbf{s}_t) = \max_a \sum_{\mathbf{s}_{t+1}, r} p(\mathbf{s}_{t+1}, r|\mathbf{s}_t, a)\left[r + \gamma V_*(\mathbf{s}_{t+1})\right] \tag{2.2}$$

.

Two major examples include value iteration [9], and policy iteration [102], which evolve around evaluating refining a policy across state space. Policy iteration recursively evaluates the value function associated with a policy as

$$V(\mathbf{s}_t) = \sum_{\mathbf{s}_{t+1}, r} p(\mathbf{s}_{t+1}, r|\mathbf{s}_t, \pi(\mathbf{s}_t) \left[r + \gamma V(\mathbf{s}_{t+1})\right], \tag{2.3}$$

then modifies the policy applied using knowledge of the value functions, with

$$\pi(\mathbf{s}_t) = \underset{a}{\operatorname{argmax}} \sum_{\mathbf{s}_{t+1}, r} p(\mathbf{s}_{t+1}, r|\mathbf{s}_t, a) \left[r + \gamma V(\mathbf{s}_{t+1})\right]. \tag{2.4}$$

Value iteration instead combines these alternating processes, applying the Bellman equation directly as

$$V(\mathbf{s}_t) = \max_{a} \sum_{\mathbf{s}_{t+1}, r} p(\mathbf{s}_{t+1}, r|\mathbf{s}_t, a) \left[r + \gamma V(\mathbf{s}_{t+1})\right] \tag{2.5}$$

until convergence is achieved and values are stationary. These methods are important to note because of the similarities they share with concepts in MCTS. Like such methods, MCTS utilizes evaluation of future values to inform the value estimate at the current state $\mathbf{s}$, but a major differentiation is that MCTS is not exhaustive. While the exhaustive approach guarantees that an optimal policy while eventually be reached, it also subjects methods like policy and value iteration to curses of dimensionality and history. As the state space grows or becomes continuous, it very quickly becomes infeasible to evaluate value functions over the entire domain. Similarly, as the number of actions an agent may take grows, the branching factor of the problem increases, and the expressed recursions become impossible to evaluate. This motivates the use of Monte Carlo methods, in which learning evolves through evaluation of a series of episodes, from which the returns associated with an action may be considered. Such techniques begin to overcome the challenges associated with large state and action spaces, though additional complexities arise when the state space is no longer fully observable.

2.1.2 Partially-Observable Markov Decision Processes

A Partially-Observable MDP (POMDP) extends the MDP definition using the 7-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, R, \mathcal{O}, \mathcal{H}, \gamma)$. Note that the definitions for terms enumerated in the prior section remain the same, with extensions considering how the state may be observed.

- $\mathcal{S}$: the probabilistic state of the environment.
- $\mathcal{A}$: the action space over which agent(s) act.
- $\mathcal{T} : \mathcal{S} \times \mathcal{A}$: a transition function between states over time conditioned on global sets of actions.
- R: a scalar reward function for the agent(s).
- $\mathcal{O}$: the space over which the environment may be observed.
- $\mathcal{H}$ a probabilistic transformation to the measurement space given a state and action.
- $\gamma \in [0, 1]$: the discount factor over time, impacting the prioritization of short term rewards.

Again, the problem is defined over a state space $\mathcal{S}$, and decisions are made over an action space $\mathcal{A}$. The system is observed over the observation space $\mathcal{O}$, with probabilities defined by $\mathcal{H}$ acting on the current state and conditioned on an action. As with the state and action spaces, O may be discrete or continuous over the domain $\Re^m$. $H : \Re^n \to \Re^m$ is often a measurement function incorporating uncertainty. A major challenge that quickly arises in a POMDP is the application of the belief update, which has a structure analogous to that of Bayes' Theorem, with

$$b_{t+1}(\mathbf{s}_{t+1}) \propto O(\mathbf{y}|\mathbf{s}_{t+1}, a) \sum_{\mathbf{s}_t} p(\mathbf{s}_{t+1}|\mathbf{s}_t, a) b(\mathbf{s}_t). \tag{2.6}$$

The goal of a POMDP is the determination of an optimal policy π (sequence of actions $a_{1:N} \in A$) that maximizes value V given an initial probabilistic belief in states b_0, rather than some known initial state. Additionally, consider a reward that is now a function of the action applied and uncertain prior states. The expectation of such a reward must be considered over the prior and posterior states as

$$E[R(b_t, a)] = \sum_{\mathbf{s} \in \mathcal{S}} b(\mathbf{s}_t) R_{a_t}(\mathbf{s}_t) \tag{2.7}$$

Note use of the expectation operator $E[]$, which describes the probabilistic average of the random variable to which it is applied, in this case, the prior belief. It is logical to maximize the expectation of cumulative reward, with

$$V^{\pi}(b_0) = \sum_{t=0}^{\infty} \gamma^t E[R(b_t, a | b_t, \pi)]. \tag{2.8}$$

Because of the computational challenges inherent to Equations 2.6 and 2.7, a variety of approaches need be considered for tractability. Often, the structure presented leads POMDPs to be termed belief MDPs, where the problem is described as a MDP over belief states and the full state space is the continuous set of beliefs. For many systems, sample-based approaches must be utilized to perform belief updates, and particle filters are often applied. In much of the research presented in this thesis, SO states are conveniently treated as Gaussian, and catalogs are then represented as an ensemble of Gaussian states. In such scenarios, the belief update may be performed using methodologies such as Kalman filtering [94] and nonlinear extensions [59].

General solutions to POMDPs are demonstrated to be PSPACE-complete in the worst case, and infinite-horizon solutions are shown to be uncomputable [62]. As such, common solutions to POMDPs tend to form approximations over the belief space, such as QMDP[72]; the QMDP approximation assumes perfect information in the state after the immediate timestep, greatly reducing the complexities of any value iteration. Another common offline method, point-based value iteration [82], iteratively performs belief backups with a set of belief points to improve sets of alpha vectors, over which value may then be interpolated over belief. These methods remain challenged by a high-dimensional state space, in which case online methods are often more trivial to utilize.

Here, forward search methods like POMCP [92] have seen great success. Further discussion on useful techniques for application of online methods shall be given in Section 2.1.4.

2.1.3 Markov Games

This section provides a brief overview of a problem formulation for multi-agent planning processes. Formally, many agents operate over a concept described as a Markov game [71, 57]. This problem setting may be defined by the set of elements $\{K, \mathcal{S}, \{\mathcal{A}^i\}_{i\in\{1,\dots,K\}}, \mathcal{P}, \{R^i\}_{i\in\{1,\dots,K\}}, \gamma_i\}$, where

- K: the number of agents considered in the problem.
- $\mathcal{S}$: the probabilistic state of the environment shared across agents.
- $\{\mathcal{A}^i\}$: the action space for agent i
- $\mathcal{P} : \mathcal{S} \times \mathcal{A}$: a transition function between states over time conditioned on global sets of actions.
- R^i: a scalar reward function for the ith agent.
- $\gamma_i \in [0, 1]$: the discount factor over time for agent i.

Let the scalar reward for each agent be considered in the context of pure cooperation toward a global objective g, a scenario typically referred to as a Markov potential game. Agents share a common reward function, and R^i may be considered in the context of local utility as the difference in global utility achieved by agent i conditioned on the actions performed by each other agent. This local utility may be considered in practice by evaluating the impact if no action $\mathbf{a}^i$ is performed. Note that $\mathbf{a}$ is now denoted a vector of dimension K with each index corresponding to the action for the associated agent.

$$R^i(\mathbf{s}_t, \mathbf{a}) = g(\mathbf{s}_t, \mathbf{a}) - g(\mathbf{s}, \mathbf{a}^{-i}). \tag{2.9}$$

Each agent wishes to maximize its own long-term reward, approximating a policy π^i. Because all agents jointly influence the global utility, the value function of each agent $V^i : \mathcal{S} \to R^i$ is a function of the joint policy $\pi(\mathbf{a}|\mathbf{s}_t) = \prod_{i\in\{1,\dots,K\}} \pi^i(\mathbf{a}^i|\mathbf{s}_t)$. Specifically,

$$V^i_{\pi^i,\pi^{-i}}(s) = E\left[\sum_{t=0}^{\infty} \gamma^t R^i(\mathbf{s}_t, \mathbf{a}, \mathbf{s}_{t+1}) \,\middle|\, a^i_t \sim \pi^i(\cdot|s_t), s_0 = s\right] \tag{2.10}$$

A key challenge in multi-agent decision problems is learning through challenges in communication. Often, such communication occurs at set intervals [12] or over coordination graphs [23], but in many cases, while the decision space is split for each agent, the implementation used remains centralized [3]. Decentralization also has the potential to greatly increase computation speeds for many-agent decision problems, in that each agent takes a greatly reduced view of the decision space of all other agents. This is especially relevant to MCTS if rollout models that are expensive in evaluation of the action space.

2.1.4 Monte Carlo Tree Search

Monte Carlo evaluation allows random actions to be simulated until a valuable result is reached. The following concepts are critical to the understanding of the MCTS methodology. Algorithms 1 and 2 can be used as a reference as concepts are discussed. The MCTS methodology may be applied to each of the discussed problem settings, but Algorithm 1 may be applied to MDPs with large action spaces and Algorithm 2 specifically takes focus on the POMDP setting discussed in Section 2.1.2. Further discussion of a multi-agent MCTS algorithm is presented in Chapter 5.

(1) **Nodes** - Following general data structures terminology we refer to an arbitrary index in the search tree as a node. The search tree is initialized by a root node. Any node may have zero to many child nodes, and a node with no children is referred to as a leaf node. Excluding the root node, all nodes must be associated with a parent node. Nodes are characterized as decision nodes, where a new action is sampled, and observation nodes, where an observation associated with the parent action is generated.

(2) **Rollout-based planning** - Generally, MCTS is applied over a set depth or until the problem at hand is resolved to a terminal state. To explore a large decision space, a methodology to select new actions must be determined. A rollout heuristic is defined as the means to generate a new set of actions from a leaf node in a search tree to a terminating depth. The rollout heuristic can take a variety of forms; fully random sequences could be chosen, or system knowledge can be applied to inform the relative value of actions. If a new action is not needed, a child node is selected and tree

search is recursively applied from that node.

(3) **Backpropagation** - Backpropagation is defined as the means by which immediate rewards simulated by leaf nodes in the search tree impact the estimated value function at parent nodes. Given a rollout or simulation routine that returns the discounted cumulative reward R_i sampled for a sequence of actions, one must update the estimated state-action value at a node $Q(h, a)$. Given the total number of visits to the node $N(a)$, the empirical mean may be updated as

$$Q(h, a) = Q(h, a) + \frac{R_i - Q(h, a)}{N(a)}. \tag{2.11}$$

Other statistical measures may also be incorporated, especially if there are concerns about the variance of rewards.

(4) **Selection** - If a new action is not generated, one must determine what previously sampled action to take. Generally, the selection method must balance more detailed exploration of simulated actions with high expected value with further exploration of undersampled actions. As such, a deterministic score function sc is applied for selection such that the child node i maximizing

$$sc(i) = Q(h, a_i) + c\sqrt{\frac{f(N))}{N(i)}} \tag{2.12}$$

is selected. $Q(h, a_i)$ describes the estimated value at the child node i, and $N(i)$ represents the number of times the child node was previously selected. $f(N)$ is an arbitrary non-decreasing map from $\Re^1$ to $\Re^1$ utilizing the number of visits to the root. This second term is derived from multi-armed bandit literature, and can be related to a confidence interval for the true value of an action [5]. Logically, as an action is explored in detail, the variance in the estimated value is decreased, thereby reducing this term. Generally, the natural logarithm is applied as $f()$, but other methods such as polynomial exploration have recently been considered [6, 32, 90].

(5) **Progressive widening** -Large state and action spaces can lead to curses of dimensionality in decision processes. When state and observation spaces are large or continuous, curses of history also occur. As actions lead to transitions described by generative models, without handling this behavior search trees will become infinitely wide after a single transition. An arbitrary action will lead to a

different representation of belief for each associated observation that is sampled. As such, in order to control the breadth of the search tree, one must artificially limit the number of actions explored, as well as the number of observations generated and associated with each sampled action. This so-called arm-increasing rule or progressive widening is analyzed in [110]. Widening is applied for MCTS by [25] with success; this is the first example of double progressive widening, in which the search tree breadth is slowly widened for both generation of new action sequences and state transition or observation generation. Generally, whether progressive widening is allowed is determined by a rule as a function of visits to the parent node i

$$|i| \leq N(i)^{\alpha_d} \tag{2.13}$$

such that the number of child nodes are upper bounded by a power law $\alpha_d \in (0, 1)$. Here, $|\ .\ |$ is utilized to describe the number of children at an arbitrary node.

With these concepts in mind, the tree search routine from a root node is described as follows. First, an action is determined using progressive widening. If the tree is allowed to widen, a new action is sampled; otherwise, a previous action is chosen that maximizes the score function outlined in Equation 2.12. Next, double progressive widening is applied using Equation 2.13 to determine whether new transitions or observations should be generated (the specifics depending on whether the problem is fully observable). If a new observation node is generated, the rollout model is applied; otherwise, a previous transition is chosen. If the problem is fully observable, each transition is given equal selection probability; otherwise, previous observation nodes are selected according to measurement likelihoods. The simulation process is then recursively completed from the selected child node. Finally, cumulative rewards from the rollout or recursive simulation are utilized to update expected reward, and total cumulative reward for the search iteration is returned.

Algorithm 1 The recursive simulation routine for the MCTS algorithm for MDPs, returning an updated history and reward.

procedure SIMULATE(s, h, d)
  if $d = 0$ **then**
    return $\{h,\ 0\}$
  $a \leftarrow$ PROGRESSIVEWIDEN(h)
  $s' \leftarrow \mathrm{T}(s,\ a)$
  $r \leftarrow \mathrm{R}(s',\ s)$
  if $a \notin C(h)$ **then**
    $C(h) = C(h) \bigcup \{a,\ s',\ r\}$
    $\{ha, r_{t+}\} \leftarrow$ ROLLOUT(s', ha, d-1)
    $R_i \leftarrow r + \gamma r_{t+}$
  else
    $\{ha, r_{t+}\} \leftarrow$ SIMULATE(s', ha, d-1)
    $R_i \leftarrow r + \gamma r_{t+}$

  $N(h) \leftarrow N(h) + 1$
  $N(ha) \leftarrow N(ha) + 1$
  $Q(ha) \leftarrow Q(ha) + \frac{R_i - Q(ha)}{N(ha)}$

  return $\{h, R_i\}$

procedure PROGRESSIVEWIDEN(h)
  if $C(h) \leq kN(h)^{\alpha}$ **then**
    $a \leftarrow$ DRAW(h)
  else
    $a \leftarrow \underset{a \in C(h)}{\operatorname{argmax}} Q(ha) + c\sqrt{\frac{f(N(h))}{N(ha)}}$

  return a

Algorithm 2 The recursive simulation routine for the MCTS algorithm for POMDPs, returning an updated history and reward.

procedure SIMULATE(b, h, d)

  if $d = 0$ **then**

    return $\{h,\ 0\}$

  $a \leftarrow$ PROGRESSIVEWIDEN(h)

  $b' \leftarrow \mathrm{T}(b,\ a)$

  $y \leftarrow \mathrm{H}(b', a)$

  $r \leftarrow \mathrm{R}(b',\ b)$

  if $|C(ha)| \leq N(ha)^{\alpha_y}$ **then**

    $W(hay) \leftarrow Z(y \mid b,\ a)$

  else

    $\{b',\ y,\ r\} \leftarrow$ select $C(ha)$ w.p. $\frac{W(hay)}{\sum_y W(hay)}$

  if $y \notin C(ha)$ **then**

    $C(ha) = C(ha) \bigcup \{b',\ y,\ r,\ W(hay)\}$

    $\{hay, r_{t+}\} \leftarrow$ ROLLOUT($b',\ hay,\ d\text{-}1$)

    $R_i \leftarrow r + \gamma r_{t+}$

  else

    $\{hay, r_{t+}\} \leftarrow$ SIMULATE($b',\ hay,\ d\text{-}1$)

    $R_i \leftarrow r + \gamma r_{t+}$

  $N(h) \leftarrow N(h) + 1$

  $N(ha) \leftarrow N(ha) + 1$

  $Q(ha) \leftarrow Q(ha) + \frac{R_i - Q(ha)}{N(ha)}$

  return $\{h, R_i\}$

2.2 Dynamical Systems and Estimators

This section provides a brief overview of the estimators and dynamical systems utilized within Monte Carlo Tree Search. First, an overview of the Unscented Kalman Filter is provided. Next, we briefly review Keplerian and Circular-Restricted Three Body Problem (CR3BP) dynamics.

2.2.1 Unscented State Estimation

This thesis largely makes use of the Unscented Kalman Filter (UKF) [59] within the tree search loop, with Gaussian or Gaussian mixture assumptions on catalog space objects. The UKF is an extension of the Kalman filter [103] that applies a sigma point transformation to the covariance. It is first useful to define the sigma point transformations used within the UKF. Consider some multivariate Gaussian random variable defined by a mean $\hat{\mathbf{x}}_{k-1|k-1}$ and covariance $P_{k-1|k-1}$. Further, a square root decomposition such as the Cholesky decomposition may be performed on $P_{k-1|k-1}$, obtaining the matrix square root A. The unscented transform, using the weighting scheme defined by Wan and van der Merwe [108], may then be applied to express such a random variable as a series of sigma points, subject to hyperparameters (α, β, κ) and state dimension n.

$$\chi_{k-1|k-1} = [\mathbf{s}_0, ...\mathbf{s}_{2n}] \tag{2.14}$$

$$\mathbf{s}_0 = \hat{\mathbf{x}}_{k-1|k-1} \tag{2.15}$$

$$W_0^{\mu} = \frac{\alpha^2(\kappa + n) - n}{\alpha^2(\kappa + n)} \tag{2.16}$$

$$W_0^{P} = W_0^{\mu} + 1 - \alpha^2 + \beta \tag{2.17}$$

$$\mathbf{s}_j = \hat{\mathbf{x}}_{k-1|k-1} + \alpha\sqrt{\kappa}A_j \tag{2.18}$$

$$\mathbf{s}_{n+j} = \hat{\mathbf{x}}_{k-1|k-1} - \alpha\sqrt{\kappa}A_j \tag{2.19}$$

$$W_j^{\mu} = W_j^{P} = \frac{1}{2\alpha^2(\kappa + n)}, \quad j \in 1, ..., 2n. \tag{2.20}$$

The resultant set of sigma points allows for convenient, gradient-free transformations both through a dynamical system and through any nonlinear measurement function. Note that different weights are utilized

for the central sigma point for mean and covariance. Then, to apply a prediction step, let some dynamical system be defined as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t; \mathbf{p}) \tag{2.21}$$

and numerically integrated such that

$$\mathbf{x}(t) = \mathbf{F}(\mathbf{x}_0, t, t_0, \mathbf{p}). \tag{2.22}$$

Each sigma point may then be propagated forward in time applying the dynamical system as

$$\chi_{k|k-1} = \mathbf{F}(\chi_{k-1}, t_k, t_{k-1}, \mathbf{p}) \tag{2.23}$$

The resultant set of sigma points is then desampled (optionally with process noise Q_k) as

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2n} W_i^{\mu} \chi_{k|k-1,i} \tag{2.24}$$

$$P_{k|k-1} = \sum_{i=0}^{2n} W_i^{P} \left(\chi_{k|k-1,i} - \hat{\mathbf{x}}_{k|k-1}\right) \left(\chi_{k|k-1,i} - \hat{\mathbf{x}}_{k|k-1}\right)^T + Q_k. \tag{2.25}$$

The sigma points may further be transformed into measurement space via some measurement function

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, t; \mathbf{p}), \tag{2.26}$$

resulting in a set of sigma points $\mathcal{Y}_{k|k-1}$ with dimension $k \times 2n$, where k is the dimension of the measurement. The mean and innovation may be evaluated as

$$\hat{\mathbf{y}} = \sum_{i=1}^{2n} W_i^{\mu} \mathcal{Y}_{k|k-1,i} \tag{2.27}$$

$$S_{k|k-1} = \sum_{i=0}^{2n} W_i^{P} \left(\mathcal{Y}_{k|k-1,i} - \hat{\mathbf{y}}\right) \left(\mathcal{Y}_{k|k-1,i} - \hat{\mathbf{y}}\right)^T + R_k. \tag{2.28}$$

The innovation represents the nonlinear projection of state uncertainty into measurement space with measurement uncertainty incorporated. The cross-covariance C_{xy} is also needed, and may be evaluated from the sigma points as

$$C_{xy} = \sum_{i=0}^{2n} W_i^c \left(\chi_{k|k-1,i} - \hat{\mathbf{x}}_{k|k-1}\right)\left(\mathcal{Y}_{k|k-1,i} - \hat{\mathbf{y}}\right)^T \tag{2.29}$$

The cross-covariance acts as a means to transform information between spaces, an insight that is also leveraged in Chapter 4 for development of the U-OCBE smoother. Here, it may be used to transform measurement information into the full state space in a somewhat analogous manner to the Kalman gain. The resultant form may then be used to directly apply the unscented measurement update.

$$K_k = C_{xy} S_{k|k-1}^{-1} \tag{2.30}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k(\mathbf{y}_k - \hat{\mathbf{y}}) \tag{2.31}$$

$$P_{k|k} = P_{k|k-1} - K_k S_{k|k-1} K_k^T = P_{k|k-1} - C_{xy} S_{k|k-1}^{-1} C_{xy}^T \tag{2.32}$$

The resultant filter is applied throughout this thesis in order to maintain state estimates on ensembles of SOs. This procedure may also be utilized for Gaussian mixands in a Gaussian sum filter, as in Chapter 3.2.

2.2.2 Keplerian and Cislunar Dynamics

This thesis considers the use of two dynamical systems. While forces from solar radiation, atmospheric drag, oblateness, and distant bodies are significant and should be accounted for in a full model, efficient and fast propagation schemes are desired within the tree search process, and transforming uncertainties for a large catalog of space objects between epochs is burdensome and often a bottleneck. As such, for near-Earth objects up to the GEO regime, we choose to limit tree search to the use of two-body dynamics, with

$$\ddot{\mathbf{r}} = -\frac{\mu \mathbf{r}}{r^3}. \tag{2.33}$$

Letting the full state vector be decomposed into position and velocity components, the dynamical system may be expressed as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t; \mu) = \begin{bmatrix} \mathbf{v} \\ \\ \ddot{\mathbf{r}} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{r} \\ \\ \mathbf{v} \end{bmatrix}. \tag{2.34}$$

For applications such as the demonstration presented in Chapter 6, more complex dynamical models may be applied outside the loop to better approximate reality. Otherwise, process noise levels may be chosen on the order of the dominant unmodeled forces in the problem.

A large number of the contributions presented in this thesis focus on application to cislunar space. In these cases, non-dimensional CR3BP dynamics are utilized [87]. In the CR3BP, the third body (the object state of interest) is assumed to have comparatively infinitesimal mass, while the primary bodies – in this case, the Earth and the moon – are assumed to follow circular orbits about the barycenter of the system. Via a transformation into the rotating frame, the non dimensional equations of motion for the third body may be determined analytically. Note that the primary bodies are placed along the x axis at $x_1 = -\mu$ and $x_2 = 1-\mu$, where μ is the gravitational parameter $\mu = \frac{m_2}{m_1+m_2} \leq \frac{1}{2}$. For the Earth-moon system, $\mu \approx 0.0122$.

$$x'' - 2y' - x = -(1-\mu)\frac{x-x_1}{r_1^3} - \mu\frac{x-x_2}{r_2^3} \tag{2.35}$$

$$y'' + 2x' - y = -\left(\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3}\right) y \tag{2.36}$$

$$z'' = -\left(\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3}\right) z \tag{2.37}$$

Applying these equations of motion, the dynamics for the state vector in the rotating frame may be expressed as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t; \mu) = \begin{bmatrix} x' \\ y' \\ z' \\ 2y' + x - (1-\mu)\frac{x-x_1}{r_1^3} - \mu\frac{x-x_2}{r_2^3} \\ -2x' + y - \left(\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3}\right) y \\ -\left(\frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3}\right) z \end{bmatrix}. \tag{2.38}$$

2.3 Optical Sensors in Space Situational Awareness

It is critical to provide an overview of the agents utilized in this thesis in addition to the decision-theoretic methods that are applied. In space situational awareness activities it is common to see the use of both electro-optical sensors and ground-based radar. Radar, especially passive radar, is quite useful for detection and tracking of space objects in low Earth orbit, but the methodology is challenged by objects that lie further from the Earth. At geostationary ranges and on, traditional radar becomes impractical. Optical sensors offer advantages in that they can cover wide swathes of measurement space comparatively quickly and can make detections at significantly greater range. Optical sensors placed into orbit can further avoid losses of efficiency due to weather and atmospheric transmission, making space-based sensors excellent candidates for long-term space surveillance. This section provides an overview of optical sensors as a whole. Ground and space-based optical sensors are then presented and compared. Finally, methods of utilization of optical sensors for initial orbit determination are outlined.

2.3.1 Optical Observations

Optical sensors make angular detections on the celestial sphere, and during long exposures, angular rates may also be determined. Right ascension and declination measurements may be evaluated as a function of the relative position vector ρ, with

$$\alpha = \arctan\left(\frac{\rho_y}{\rho_x}\right) \tag{2.39}$$

$$\delta = \arcsin\left(\frac{\rho_z}{|\rho|}\right) \tag{2.40}$$

and

$$\rho = \mathbf{r} - \mathbf{o} \tag{2.41}$$

in an inertial frame. While the inertial specification is not necessarily needed for angular measurements, it is a more useful assumption when considering angular rate measurements. Recognizing that the angular measurements are explicitly a function of the relative position vector, it is convenient to apply the measurement jacobian $H = \frac{d\mathbf{y}}{d\rho}$, where $\mathbf{y} = [\alpha \quad \delta]$. Then,

$$[\dot{\alpha} \quad \dot{\delta}] = \frac{d\mathbf{y}}{d\rho}\frac{d\rho}{dt} = H\frac{d\rho}{dt} \tag{2.42}$$

$$H = \begin{bmatrix} -\frac{\rho_y}{|\rho_{xy}|^2} & \frac{\rho_x}{|\rho_{xy}|^2} & 0 \\ -\frac{\rho_x\rho_z}{|\rho|^2|\rho_{xy}|} & -\frac{\rho_y\rho_z}{|\rho|^2|\rho_{xy}|} & \frac{|\rho_{xy}|}{|\rho|^2} \end{bmatrix} \tag{2.43}$$

$$|\rho_{xy}| = (\rho_x^2 + \rho_y^2)^{1/2} \tag{2.44}$$

and $\dot{\rho}$ is easily computed given knowledge of the observer trajectory and the object state estimate.

Measurement models may incorporate explicit computation of uncertainty and probability of detection, with further discussion on the subject in [77] and [24]. When determining measurement uncertainties, two theoretical limits must be taken into consideration. First, the spatial resolution of a sensor may be determined via computation of the instantaneous field of view at a pixel level. This is trivially determined as

$$IFOV = 2\arctan\left(\frac{p}{2fd}\right) \tag{2.45}$$

where p is pixel pitch, f is the f-number, and d is the aperture diameter. In order to account for pixel uncertainty, an uncertainty floor is set at half the instantaneous field of view $1 - \sigma$. Additionally, one

must consider diffraction limits on imaging uncertainty. The Rayleigh criterion for a circular aperture is well-known as

$$\sin(\theta_R) \approx \theta_R = 1.22\frac{\lambda}{d}. \tag{2.46}$$

When the diffraction limit exceeds pixel resolution, it must be incorporated into the uncertainty model.

One must also consider expected uncertainties on angular rates for an optical observer. Over a long exposure, a target object is generally resolved as a streak in the observer field of view, assuming approximate straight-line motion over a short arc from the perspective of the observer. Alternatively, the object may be tracked, and streaking motion may be extracted from stars that are stationary in the celestial sphere. Each endpoint of that streak is considered an independent angular measurement with accurate timing. Note that there is ambiguity in the direction of the streak, but that said ambiguity may be quickly resolved with prior knowledge of the space object or may be used to generate two different angular rate measurements. Let measurement 1 have Gaussian distribution $\mathbf{y}_2 \approx \mathcal{N}(\mu_1, R_\theta)$ at time $t_1 = 0$, and let measurement 2 have Gaussian distribution $\mathbf{y}_2 \approx \mathcal{N}(\mu_2, R_\theta)$ at time $t_2 = \Delta t$. The expected distribution of the angular rates may be considered a transformation of Gaussian random variables, where

$$\dot{\mathbf{y}} = \frac{\mathbf{y}_2 - \mathbf{y}_1}{\Delta t}. \tag{2.47}$$

The expected mean is simply

$$E[\dot{\mathbf{y}}] = \dot{\hat{y}} = \frac{\mu_2 - \mu_1}{\Delta t} \tag{2.48}$$

and the covariance of this transformed random variable is also trivially computed. Recognizing that the two measurements are independent, we find

$$cov(\dot{\mathbf{y}}) = E[(\dot{\mathbf{y}} - \dot{\hat{y}})(\dot{\mathbf{y}} - \dot{\hat{y}})^T] = \frac{2R_\theta}{\Delta t^2}. \tag{2.49}$$

It is also important to show the resultant rates are uncorrelated with the angular measurements. With the covariance between two random variables defined as

$$cov(X, Y) = E[XY] - E[X]E[Y] \tag{2.50}$$

For random variables $\mathbf{y}$ and $\dot{\mathbf{y}}$,

$$E[\mathbf{y}\dot{\mathbf{y}}^T] = E[\dot{\mathbf{y}}\mathbf{y}^T]^T \tag{2.51}$$

$$= \frac{1}{2\Delta t} E[(\mathbf{y}_1 + \mathbf{y}_2)(\mathbf{y}_2 - \mathbf{y}_1)^T] \tag{2.52}$$

$$= \frac{1}{2\Delta t} (E[\mathbf{y}_1\mathbf{y}_2^T] + E[\mathbf{y}_2\mathbf{y}_2^T] - E[\mathbf{y}_1\mathbf{y}_1^T] - E[\mathbf{y}_2\mathbf{y}_1^T]) \tag{2.53}$$

$$= \frac{1}{2\Delta t} (R_\theta - R_\theta + E[y_1]E[y_2]^T - E[y_2]E[y_1]^T) \tag{2.54}$$

$$= \frac{1}{2\Delta t} (\mu_1\mu_2^T - \mu_2\mu_1^T) \tag{2.55}$$

and

$$E[\mathbf{y}]E[\dot{\mathbf{y}}]^T = \frac{\mu_1 + \mu_2}{2} \frac{(\mu_2 - \mu_1)^T}{\Delta t} \tag{2.56}$$

$$= \frac{1}{2\Delta t} (\mu_1\mu_2^T + R_\theta - R_\theta - \mu_2\mu_1^T) \tag{2.57}$$

Therefore, $cov(\mathbf{y}, \dot{\mathbf{y}}) = 0$ and the measurements are uncorrelated. A full observation of optical angles and angular rates may then be expressed as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y} \\ \dot{\mathbf{y}} \end{bmatrix}, \quad R = \begin{bmatrix} R_\theta & 0 \\ 0 & \frac{2R_\theta}{\Delta t^2} \end{bmatrix} \tag{2.58}$$

A variety of constraints may be applied for both space-based and ground-based sensors. Daylight constraints are imposed on ground-based sensors, defined when the angle between incoming solar rays and the surface normal is greater than 90 degrees. Occlusion and eclipse constraints are also defined. For any object, if the Earth or moon passes in front of either an observer or oncoming solar rays, that object cannot be detected. Finally, in this thesis, pointing constraints are considered for the Earth, Moon, and Sun, and detections are not made near the limb of such bright sources. A visualization of these constraints is given in Figure 2.1.

It is also critical to note that the full state is not immediately observable with a single optical sensor. During a single pass of a target, the observations are effectively locally linear in the observer field of view leading to what is colloquially defined as a Too Short Arc (TSA), tracklet or uncorrelated track [78]. Data from a TSA are assumed to be associated as they typically form a great circle. As this is expected to be

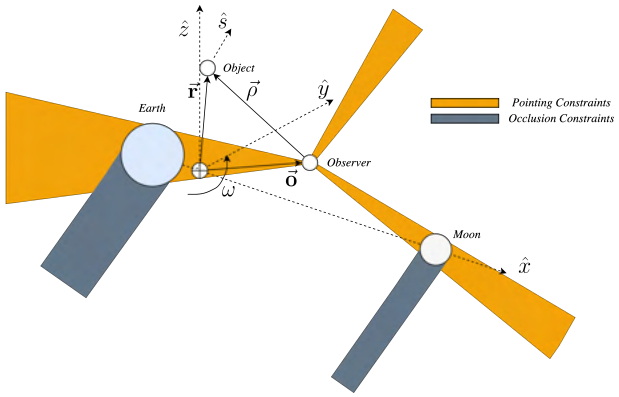


Figure 2.1: Unobservable regions of state space in the three body tasking problem.

resolved approximately as a straight line in the sensor field of view, the observation is effectively realized as a four-dimensional subset of the target state known as the attributable vector. The unobservable subset describes the target range and target range-rates, which may be dynamically constrained as an "admissible region" in which range and range rate might lie. Derivations for admissible region boundaries and methods to instantiate estimates over admissible regions are provided in Appendix 2; these methods are relevant for the follow-up observation techniques presented in Chapter 3.

2.3.2 Optical Observers

An optical observer may be utilized in a variety of contexts, with benefits for observation of space objects in a variety of regimes. Several candidates for observational platforms are outlined below and utilized throughout this thesis.

2.3.2.1 Earth Ground-based Observers

Earth ground-based observers are most commonly used for space situational awareness today with a variety of examples visible in government, commercial, and academic activities. As a prime example, the Vision, Autonomy, and Decision Research (VADeR) Observatory at the University of Colorado at Boulder shall be discussed later in this thesis. Ground-based instruments typically range from relatively small optics on the order of 0.1 meters in aperture to exquisite, diffraction-limited instruments with multi-meter optical apertures. These instruments are weather dependent and largely may only operate after astronomical

twilight, but are the most cost-effective means of placing optical observers into operation. Additionally, ground-based observers offer great utility for objects at geostationary orbit and below. The closeness of such objects leads to very useful positional information, and often, objects are moving at very fast angular rates. This is not the case for geostationary objects, but since geostationary objects are fixed in an Earth-fixed frame, it is somewhat straightforward for ground-based observers to maintain knowledge on the local geostationary environment.

Ground-based observers may be assumed to further struggle with detection and tracking of space objects in the XGEO regime, especially in the near-lunar environment. Much less angular rate information is available for such objects from Earth, and further, there is little diversity in viewing angles for a ground-based observer. This can lead to directions in which positional uncertainty is quite large, a problematic issue for comparatively nonlinear orbits. Furthermore, detection of objects in the cislunar regime is a nontrivial issue for ground-based observers because of increased background noise near the lunar limb. Orbits are then impossible to observe near perilune, and objects in low lunar orbit are undetectable for ground-based observers. These factors motivate the inclusion of space-based observers throughout this thesis, and such observers are demonstrated to aid in resolving the observational challenges of objects in the XGEO regime.

2.3.2.2 Keplerian Observers

Space-based observers with semi-major axes at or below geostationary orbit are expected to become increasingly common in the near future, especially as the local catalog of space objects grows exponentially. Unlike ground-based observers, any space-based observer may remain online at all times, so long as the solar phase angle to a target object is sufficient to detect the object. While such observers aren't often considered in this thesis, it is worthwhile to briefly discuss useful parameterizations for observers.

Geostationary space-based observers are useful to many space surveillance cases. Because they are fixed in the Earth-centered Earth-fixed frame, communication and coordination with geostationary observers from ground stations is comparatively trivial. In addition, geostationary observers maintain fixed relative states to other geostationary objects in the Hill frame, so visibility of geostationary objects of interest may be guaranteed. Finally, if a geostationary observer is used to track space objects with smaller semi-major axes,

the line of sight information gained is significantly different from that achieved by ground-based observers, and the combination of observers may be used to bound the state of target objects with much greater certainty.

Another useful Keplerian orbit for space situational awareness is the sun-synchronous orbit, in which the oblateness of the Earth is utilized to match orbital precession with the movement of Earth around the sun. This results in the ground track of the orbit passing set points on Earth at the same time every day, a useful feature for coordinating with ground-based sensors. Generally, sun-synchronous observers operate in very inclined orbits with comparatively short periods. This structure results in a variety of viewing angles for a target space object that can be used to quickly reduce state uncertainty. While no results using sun-synchronous observers are presented in this thesis, further discussion may be found in [33].

2.3.2.3 Cislunar Periodic and Quasiperiodic Observers

Much of this thesis considers the application of space-based observers to the emerging cislunar regime. These observers are of special interest due to the expected growth in space objects near the moon, and such objects are especially difficult to track because of visibility constraints and nonlinearity of the underlying orbits. Any cislunar observer must balance accessibility to regions of interest across a variety of solar phase angles alongside the utility of any observations it makes on target space objects. A rich discussion of accessibility is presented by Vendl and Holzinger [105], and this thesis attempts to further probe the question of utility, especially as a catalog of objects is considered. Largely, this thesis focuses on the use of periodic and quasiperiodic Halo and Lyapunov orbits, further analyzing the orbits shown by Vendl to be promising for cislunar space surveillance.

Lyapunov orbits are interesting candidates for a variety of reasons, and Vendl demonstrates that a phased L1 Lyapunov orbit that is 1:1 resonant with the Earth-Moon synodic period offers full accessibility into the near-lunar region [105]. Lyapunov orbits are also interesting in that they traverse further from the moon than Halos, improving observation quality for near-Earth space objects and objects orbiting about the L3, L4, and L5 Lagrange points. During the close lunar approach, Lyapunov orbits can achieve a great variety of viewing angles for near-lunar objects, but for much of their orbits, Lyapunov observers are plagued

by very slow moving line of sight relative to target space objects. This is especially the case for an orbit such as the 1:1 resonant case with a period on the order of a month.

Compared to a 1:1 resonant Lyapunov orbit, Halo orbits generally traverse a much smaller region of state space over a shorter period. Such orbits offer very useful positional resolution for other Halo orbits and low lunar orbit, and because they move comparatively quickly, Halo orbits grant a diversity of viewing angles on target space objects. Unlike Lyapunov orbits, they also offer out of plane viewing angles. Vendl demonstrates that accessibility across a synodic period can be more challenging for Halo orbits, especially for objects that are distant from the moon.

2.3.2.4 Lunar Surface Observers

Lastly, it is worth considering the placement of optical sensors on the lunar surface. With the growth in interest in lunar surface operations, such configurations may be cost effective compared to space-based sensors. These sensors are stationary in the Earth-Moon rotating frame, but lunar observers offer great utility, especially for objects following near-lunar periodic orbits. Any other optical sensor struggles to track such objects during close lunar approaches, but lunar surface observers, especially those placed at the lunar poles, offer access to space objects during close lunar approaches. These approaches are the most dynamically sensitive subsets of an orbit, and at these times, small maneuvers are most impactful.

Chapter 3

Optical Sensor Tasking using Monte Carlo Tree Search

This chapter presents several initial applications of Monte Carlo Tree Search to optical sensor tasking. First, analysis of the utilized MCTS algorithms is summarized. The presented algorithm is novel in that it utilizes polynomial exploration alongside large state and action spaces, features that are enabled by the state representations discussed in the prior chapter. A full derivation of the methods utilized is presented in Appendix A. In addition to theoretic analysis of the methods, a numerical analysis of returns is presented for the catalog maintenance objective. Techniques and results are then presented for two sensor tasking objectives. For the catalog maintenance objective, a variety of methods are developed that inform the expansion and backpropagation phase of MCTS; the techniques used consider information-theoretic metrics that may be derived from state estimates and observer knowledge. A result set is considered studying future sensor tasking needs in cislunar space. A variety of methods to be used in the expansion phase of MCTS are next developed for follow-up sensor tasking. This problem is then considered from an estimation perspective, and a methodology for incorporating negative information from null detections is derived. These techniques are then united for a scenario in which follow-up observation is desired for an initial detection that is posited to be a bounded SO.

3.1 Analysis of Monte Carlo Tree Search with Double Progressive Widening and Polynomial Exploration

This section provides detailed insight into theoretic and numerical convergence of the MCTS methodology. The full derivation is presented in Appendix A, and this section outlines major results of the proof.

The proof by induction is presented determining an upper bound for the error in the expected optimal value as a function of node visits and search tree depth. Generally, the induction techniques utilzed by Kocsis [64] and Auger [6] are followed. Similar analyses were also presented by Shah et al. [90]. The major contribution of this section lies in the extension of analysis to partially observable settings.

3.1.1 Asymptotic Analysis: Preliminaries

This proof makes the assumption that the true value function V^* is a random variable such that $V^* \in [0, 1]$; therefore the estimation error at an arbitrary node $V - V^*$ may be bounded as $V - V^* \in [-1, 1]$.. Several definitions are necessary for this analysis.

Definition 1 (Exponentially Sure in n)**.** *Some property P depending on integer N is exponentially sure in N (e.s) if there exist positive constants C, h, η such that the probability P holds is at least*

$$p(P) \geq 1 - C\exp(-hN^{\eta}) \tag{3.1}$$

Definition 2 (Consistency)**.** *There exist coefficients $C_d > 0,\ \mu_d > 0$ such that for all nodes at integer depth d,*

$$|V(z) - V^*(z)| \leq C_d N(z)^{-\mu_d} \tag{3.2}$$

exponentially surely in $N(z)$. That is, the difference between the estimated value function and true value exponentially decreases to zero as nodes are visited.

Definition 3 (Regularity)**.** *For any $\Delta > 0$, we assume there exist $\theta > 0$ and $p > 1$ throughout the simulation such that the probability the value function of a sampled action i differs from the Bellman optimal value is bounded by Δ is*

$$p(V(i) \geq V^*(z) - \Delta) \geq \min(1, \theta\Delta^p) \tag{3.3}$$

To further clarify the inductive proof, nodes are split into decision nodes, where an action is selected, and observation nodes, where a measurement is applied and a belief update is performed. First, the bounds on the error of the estimate of the value function for the transition from a decision node to an observation node must be considered. To do so, it is key to consider that for a transition on observation nodes in the POMDP framework, it is assumed that child nodes shall be selected according to observation likelihood given current belief.

3.1.2 Observation nodes are selected according to observation likelihood

We first wish to establish that observation nodes shall be selected in proportion to observation likelihood ω_i within the tree search process. Deterministic sampling is applied, and the number of visits to a child node i may be bounded as

$$N(i) \geq \frac{\omega_i N^2}{N + |w| - 1}, \tag{3.4}$$

$$N(i) \leq \omega_i N + 1. \tag{3.5}$$

$|w|$ is the current number of observation nodes, and assuming double progressive widening is applied with coefficient α_o at the observation node depth,

$$|w| = \lfloor N^{\alpha_o} \rfloor. \tag{3.6}$$

These bounds hold for all nodes except those recently generated, in which case the number of visits to the child node is explicitly

$$N(i) = N - \left\lceil \lfloor N^{\alpha_o} \rfloor^{\frac{1}{\alpha_o}} \right\rceil.$$

It may be assumed that $N(i)$ is large relative to the expected number of visits $\omega_i N$, such that the computed bounds hold for all nodes. These bounds are then utilized to evaluate the consistency of the observation node transition. Further justification of this result is given in Appendix A.1.

3.1.3 Consistency of Observation Nodes

Using this result, we next wish to form bounds on the estimation error over the transition from decision nodes to observation nodes. Note that this section applies Hoeffding's inequality [16] to upper bound the

probability the sum of bounded random realizations diverges from the expected sum. For a set of bounded random variables X, Hoeffding's inequality is generally defined as

$$p(|S_n - E[S_n]| \leq t) \geq 1 - 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right), \tag{3.7}$$

$$S_n = \sum_{i=1}^{n} X_i, \quad X_i \in [a_i, b_i].$$

The convergence rates of these bounds are established as a function of child nodes in Lemma 1.

Lemma 1. *Let the asymptotic convergence rate of estimation error in child observation nodes at depth d be defined as μ_d. The asymptotic convergence rate at the parent node at depth $d - \frac{1}{2}$ may be expressed as*

$$\mu_{d-\frac{1}{2}} = \frac{\mu_d}{1 + 3\mu_d} \tag{3.8}$$

Proof. Using likelihood-based selection, a proof of this result is outlined in Appendix A.2. This result holds using the expectation to an arbitrary level of variance. Alternatively, convergence rates for a hard upper bound are expressed in Appendix A.3. □

This result may then be passed to a parent node, allowing for recursive evaluation across observation nodes.

3.1.4 Consistency of Decision Nodes

Supposing there exists some constant for consistency across observation nodes $\mu_{d-\frac{1}{2}}$, we now wish to determine progressive widening coefficients for decision nodes α_d and a recursion for convergence factor μ_{d-1}. The recursion is established in Lemma 2.

Lemma 2. *Let the asymptotic convergence rate of estimation error in child decision nodes at depth $d - \frac{1}{2}$ be defined as $\mu_{d-\frac{1}{2}}$. Then, the asymptotic convergence rate at the parent observation node at depth $d - 1$ may be expressed as*

$$\mu_{d-1} = \frac{4\alpha_d(1 - \alpha_d)}{4 + \alpha_d(1 - \alpha_d)} \tag{3.9}$$

with

$$\alpha_d = \frac{2\mu_{d-\frac{1}{2}}}{1+4\mu_{d-\frac{1}{2}}} \leq \frac{2}{5}. \tag{3.10}$$

Proof. A similar process is applied as compared to that of the solution to Lemma 1. An asymptotic analysis is performed in Appendix A.4. Note that this result is more conservative than that of [6]. □

3.1.5 Proof Conclusion

To analyze convergence as a function of a maximal depth d_{max}, we initialize the widening coefficient at an observation node as $\alpha_o(d_{max}) = 1$. Then, using Equation 3.8, $\mu_{d_{max-\frac{1}{2}}} = \frac{1}{4}$. It follows using Equation 3.9 that $\alpha_d = \frac{1}{4}$ and $\mu_{d_{max-1}} = \frac{12}{67}$. Recursively, parent nodes may be analyzed until the root node is reached, and polynomial asymptotic bounds on value function estimation error convergence are obtained.

3.1.6 Numerical Evaluation

The developed methodology may also be analyzed numerically by considering the expected return of the output policy as a function of iterations through the search tree. To do so, a simplified version of the simulations outlined in Section 3.2.2 is developed so that a consistent return may be expected.

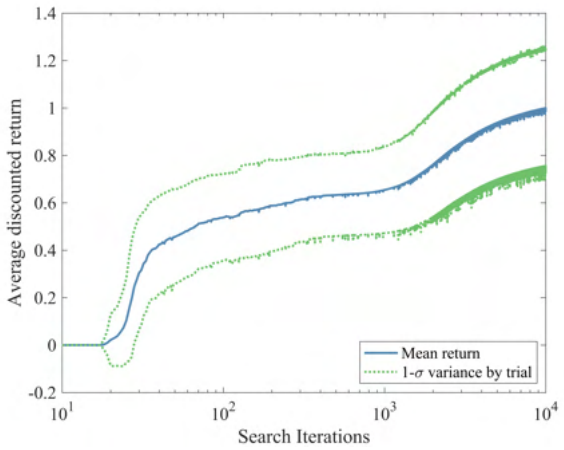


Figure 3.1: Mean discounted reward as a function of MCTS iterations.

In Figure 3.1, the average expected return over a large set of simulations is plotted against the number of search iterations allowed. Expected returns follow a structure consistent with MCTS algorithms for other applications. It is clear that the tasking solution tends to converge toward a maxima, and further Figures help support this point. First, Figure 3.2 demonstrates the average return of a search trajectory at each iteration. Figure 3.3 demonstrates the immediate value of the action associated with the estimated optimal trajectory.

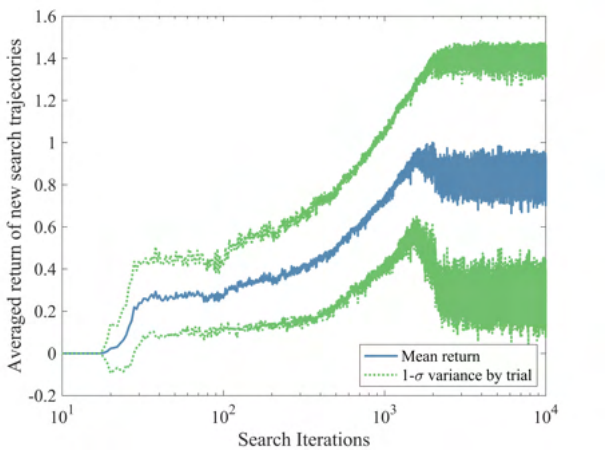


Figure 3.2: Expected return for each search iteration.

Figure 3.2 helps visualize the balance between exploration of branching policies and exploitation of planning results found to be valuable in the sense of the associated reward. Here, we observe that the MCTS algorithm begins to explore actions that are more valuable within 1-2000 search iterations. The slight dip in average reward with high variance intuitively corresponds to the further exploration of solutions around the found maxima, and returning to Figure 3.1, one may observe that this allows the algorithm to extract further value in an updated decision sequence.

There are two points of interest that may be gained from Figure 3.3. First, a dip in the immediate reward found occurs as discounted reward is gained, which further supports the point that there is value to be exploited and that there is an inefficiency in taking a myopic view in planning for this problem. Second, having some sense of the immediate reward in this problem gives a better idea of the upper bound in discounted reward, with the assumption that the immediate reward remains somewhat constant between

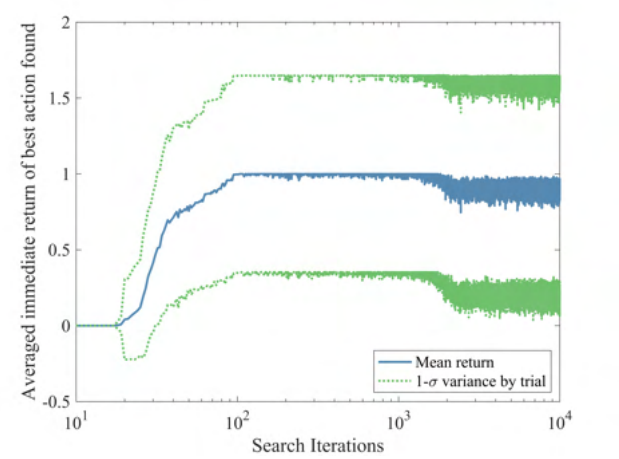


Figure 3.3: Mean discounted reward as a function of MCTS iterations.

observations. If this is the case, we also may apply the set search depths and discounts for this numerical evaluation (depth $d = 10$ and discount $\gamma = 0.9$, respectively). With this knowledge and an unnormalized expected immediate reward of $R = 100000$, the discounted reward may be approximated as

$$R_\gamma = \sum_{i=0}^{9} R \approx 650000 \tag{3.11}$$

The average discounted return after 10000 search iterations is about 570000, or around 88 percent of this approximation. With the return appearing to continue to grow, this plot helps demonstrate the assumption that the evaluated tasking solution is nearing a maximal value over the horizon is fair.

3.2 Sensor Tasking for Catalog Maintenance

It is now left to consider how the demonstrably useful MCTS methods may be applied to specific problems. We first revisit the problem of tasking sensors to maintain an uncertain catalog of space objects, and it is worth briefly revisiting the problem setting. As discussed in Section 1.1, it is useful to characterize decisions utilizing information-theoretic measures of the studied SOs [30, 47]. As is demonstrated in the prior section, such methods may be extended to avoid a myopic view of the tasking problem, and the (PO)MDP framework is a convenient formalism for this process. This framework has been utilized with great success, applying offline reinforcement learning methodologies [70, 91].

This section considers application of the online, anytime methods presented in Section 3.1 to the cata-

log maintenance tasking problem and takes specific focus in considering how MCTS may be bootstrapped for efficient exploration of the decision space. Shah demonstrates theoretic improvements to MCTS convergence when the expansion phase is informed by supervised learning or expert knowledge [90], and the ensuing section outlines how expert knowledge may be applied to this problem.

3.2.1 Rollouts and Rewards for Catalog Maintenance

Methods for applying expert knowledge may be applied in two separate segments during MCTS simulation. First, a rollout policy may be developed that is applied when new leaf nodes are generated in a search tree. Alternatively, a reward may be applied that is based on information-theoretic quantities, leading MCTS to target actions with information-theoretic value through the observed returns instead of the expansion phase. Throughout this thesis, the results presented utilize a mix of these concepts.

3.2.1.1 Covariance-based Rollout Heuristics

We first present a variety of rollout heuristics that determine action sampling as the search tree is generated. In each method, stochastic sampling is applied to select new actions, with an action i given a relative sampling weight ω_i. For catalog maintenance, action i for observer j refers to the object that this observer chooses to observe at a certain epoch. At a base level, action sampling may be performed in a completely random manner, but this form of expansion is incredibly inefficient for the catalog maintenance problem. Often, such actions would lead to zero or negative reward if an observer cannot detect the tasked object, and it is relatively straightforward to ensure actions expected to be useful are explored.

Action sampling may be performed purely in state space, and it is first useful to scalarize the covariance of a target estimate with

$$\omega_i = \mathrm{tr}\left(P\right) = \alpha \mathrm{tr}\left(P_r\right) + \beta \mathrm{tr}\left(P_v\right). \tag{3.12}$$

In a non-dimensional frame, the first weighting scheme is beneficial in that one may expect it would lead to approximately consistent target uncertainties across the studied catalog, with equal weight given to position and velocity information. Alternatively, normalizing weights may be applied to the positional

and velocity subsets of the full covariance to ensure objects in a dimensional frame with large positional uncertainties are not the sole candidates to prioritize. It is also important to note that this weighting scheme may be considered agnostic to the observer, in that no consideration is given to specific observer geometries. Improvements to this first measure may quickly be made by taking an observer into consideration. A logical first extension is the projection of target uncertainties into measurement space, such that

$$\omega_i = \text{tr}\left(HPH^T\right). \tag{3.13}$$

This transformation is useful in several ways, and for optical sensors, the resultant weight effectively describes the sum of angular variance of the projected SO. The resultant covariance in measurement space may also be compared to measurement uncertainty to gain insight into the potential benefits of observing the target. If the target is distant, or if there is already significant information on the observable subset of target uncertainty, this methodology lowers the associated weight accordingly. Additionally, when applying this transformation, one may easily apply observing constraints, setting weights to zero if any constraint is not satisfied.

Finally, one may consider the fulll effects of processing observations in an information-theoretic manner. Quantities like Kullbeck-Liebler divergence, Shannon entropy, and mutual information may easily be computed. Generally, it is more useful to consider target uncertainties in developing sampling heuristics, rather than to use a pure Fischer information gain approach, where

$$\omega_i = \det\left(H^T R^{-1} H\right). \tag{3.14}$$

This method often leads to scenarios in which observers repeatedly observe a target that maximizes information gain neglecting the tracking needs for other objects. Softmax models may also be considered for each of these sampling techniques; these transformations effectively increase prioritization of targets with high sampling rates as

$$\omega_i = \frac{e^{\omega_i}}{\sum_j e^{\omega_j}}. \tag{3.15}$$

Finally, it is important to note that these weighting schemes can significantly challenge an observer in terms of attitude control input necessary to achieve tasked actions. As such, it is often useful to apply

slew penalties to tasking actions. In practice, rather than reweighting a potential action by distance from the current observation, we enforce zero weight for actions more distant in measurement space than what is achievable given observer slew rates and allocated slew and settle times. Similarly, the rollout methods utilized often accommodate seeing constraints described in the observational models in Section 2.3. If an object is not visible to an observer, the associated action is then given a weight of zero.

3.2.1.2 Rewarding Catalog Maintenance Actions

Reward functions for this problem may be applied in a very similar manner to the expansion phase, but often apply the dynamical evolution of the full state between epochs. Several information-theoretic rewards may be considered, including change in covariance traces, Kullbeck-Liebler divergence, Fischer information, and differential entropy. Given that the vast majority of the catalog is not observed at a given epoch, it is useful to reduce such quantities to consideration of the observed SOs. A covariance trace-based reward may then be expressed as

$$r_i = \sum_{j=0}^{|a_i|} \text{tr}\left(P_{j,i}^{-} - P_{j,i}^{+}\right) \tag{3.16}$$

With Gaussian assumptions on states, the Kullbeck-Liebler divergence may similarly be expressed as

$$r_i = \sum_{j=0}^{|a_i|} \frac{1}{2}\left(\text{tr}\left(P_{j,i}^{+-1}P_{j,i}^{-}\right) + \left(\hat{\mathbf{x}}_{j,i}^{+} - \hat{\mathbf{x}}_{j,i}^{-}\right)^T P_{j,i}^{+-1}\left(\hat{\mathbf{x}}_{j,i}^{+} - \hat{\mathbf{x}}_{j,i}^{-}\right) + \log\left(\frac{|P_{j,i}^{+}|}{|P_{j,i}^{-}|}\right) - n\right). \tag{3.17}$$

Methods that consider the tracked catalog as a whole may also be considered, and such methods are generally computationally trivial. For example, a consistent reward may be given if each object is studied within a recent decision history, or rewards may be given in proportion to the unique objects visited. Such reward schemes can better encourage revisiting objects, a goal that is less critical in covariance minimization scenarios, but may be impactful for objects with maneuver potential.

3.2.2 Application to cislunar space

The discussed rollout and reward methods are now directly to the problem of optical tracking. Note that this contribution was published as "Sensor tasking in the cislunar regime using Monte Carlo Tree Search," as outlined in Section 1.6. MCTS within this context is modeled as a POMDP, with

- $\mathcal{S}$: an ensemble of N space object states modeled as multivariate Gaussian random variables, along with completely known observer states.
- $\mathcal{A}$: Each observer may choose to observe a single space object, leading to an action space with dimension $M \times N$, where M is the number of observers and N is the number of space objects.
- $\mathcal{T} : \mathcal{S} \times \mathcal{A}$: a transition function between states over time conditioned on global sets of actions. In this case, Kalman measurement updates are performed on each tasked action if an object is detected, and space object states dynamically evolve under CR3BP dynamics.
- R: reduction in covariance traces.
- $\mathcal{O}$: the space over which the environment may be observed, in this case, right ascension, declination, and associated rates.
- $\mathcal{H}$ the transformation of spacecraft states onto the celestial sphere in the field of regard of the applied observer.
- $\gamma \in [0, 1]$: the discount factor over time, impacting the prioritization of short term rewards.

The presented application largely makes use of the CR3BP [87], discussed in detail in Section 2.2.2. Propagation of uncertainty and measurement updates are performed using the Unscented Kalman Filter, discussed in Section 2.2.1. The unscented transform or another filter that captures higher order statistics is necessary for this application given the nonlinear motion prevalent even in simplified versions of the three body problem.

When considering candidate objects in the cislunar regime, it is critical to characterize the expansive breadth of potential trajectories. Regions of chaotic motion exist, but few to no feasible use cases exist for such trajectories in scientific or commercial missions. Periodic trajectories generally arise about the Lagrange points, which are stationary points in the rotating frame. This study is performed using a variety of these periodic trajectories. In addition to periodic orbits, a variety of highly elliptic transfers from lower orbits are also incorporated. An itemization of the full studied catalog is provided in Table 3.1. Of particular interest are objects in the near-lunar environment, specifically SOs in Halo, Lyapunov, and Distant Retrograde orbits.

Orbit Type	Object Count
Earth-Moon 3-1 Resonant	20
L1/2/3 Axial	60
Distant Retrograde Orbit	20
Long Period L4/5	40
Lyapunov L1/2/3	60
L1/2/3 Northern Halo	60
L1/2/3 Southern Halo	60
Short Period L4/5	40
Vertical L1/2/3	60
Highly Elliptic Orbit	80
Total	**500**

Table 3.1: Periodic orbits utilized in tasking simulations.

Two separate cases are considered, performed with an Earth-based telescope at Haleakalā, Maui, USA that has capabilities consistent with a large dedicated SDA instrument. In each case, observers are assumed to orbit on quasiperiodic tori in the circular-restricted three body problem, with trajectories computed using the PDE(DFT) methodology [7].

First, a planar trajectory about the L1 Lagrange point is considered. A Lyapunov orbit is relatively stable, requiring a comparatively small amount of station keeping thrust, and offers observability throughout the region of interest in space. Following the work of Vendl and Holzinger [105], an orbit with a period approximately equivalent to the Earth-Moon synodic period is selected, and the initial phase of the orbit is chosen to be aligned with the solar phase angle, such that the observing region directly about the moon remains well-illuminated from the perspective of the observer throughout the period of study.

Alternatively, an observer in a quasiperiodic L2 Northern Halo orbit is considered. Halo orbits are of great interest to scientific missions, and have been considered as candidate orbits for NASA's Gateway mission. In addition, Halo orbits have seen broad use about the Sun-Earth Lagrange points, for missions such as ISEE-3 and SOHO. These orbits are comparably unstable, requiring further station keeping, but as the orbits are three dimensional and offer shorter periods, observers gain a more diverse set of viewing geometries over the course of long-period observation. Again considering the work of Vendl [105], an approximate 3:1 synodic resonance is selected. Vendl finds little correlation between observation accessibility and initial phase of the observer, likely because this orbit has such a short period, but still remains relatively close to the

moon. Because of the short period and positionally small orbit, variations in the line of sight vector occur at a higher rate, but there is not as much significant change in the line of sight vector to points of interest like Earth or the Lagrange points, as in the case of the Lyapunov observer.

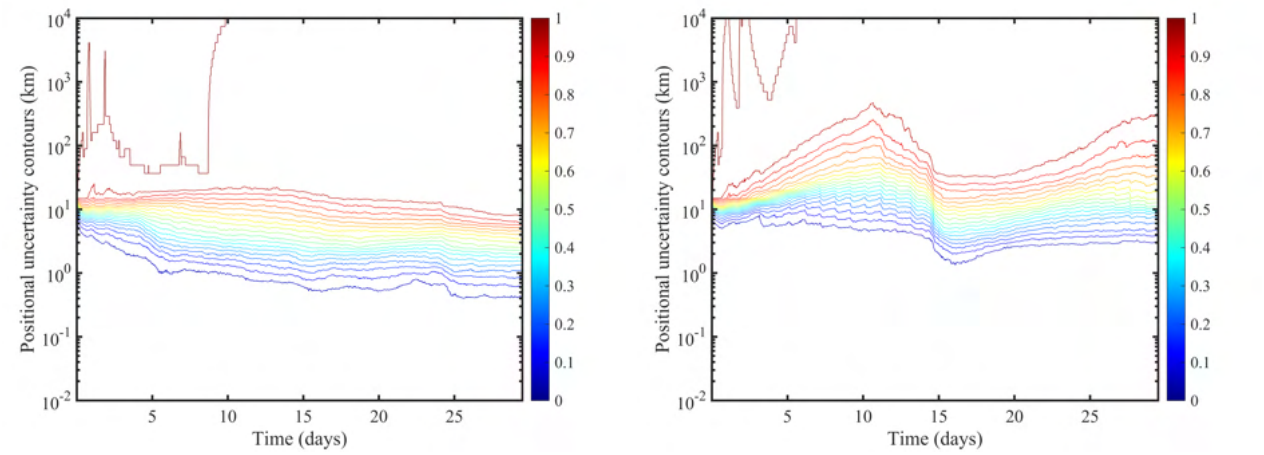
Specification	**Ground-based Sensor**	**Space-based Sensor**
Aperture (m)	3.67	0.3
f-number	200	7
Pixel pitch (μm)	1	1
QE	0.9	0.9
Read noise (pix/s)	5	2.0
Optical Transmission	0.756	0.756
Atmospheric Transmission	0.7	1.0

Table 3.2: Space and ground-based sensor specifications for large-scale catalog maintenance.

These results utilize the optical observation models outlined in Section 2.3, with a variety of constraints imposed for both space-based and ground-based sensors. Models for each sensor are provided in Table 3.2, and largely utilize models from commercial off the shelf sensors. Tasking is applied using the projection of state uncertainty into measurement space (Equation 3.13) as a rollout heuristic for search tree expansion, and change in covariance trace (Equation 3.16) is applied as a reward. Rewards are stochastic in that an observation is received if a random number exceeds the theoretic probability of detection of the target object.

Outside of the MCTS loop, tasking performance may be evaluated in a variety of ways. The first critical consideration is the evolving behavior of estimate uncertainties. To study this, one may track features of the covariance ellipsoid over time for each object. Position and velocity covariance traces are well-known tools to do so, and are often preferred over properties like the determinant, as they provide more insight into the relative scale of position and velocity errors. Also relevant to the specific problem of tasking in the cislunar regime is the question of whether custody is maintained for objects. An effective tasking solution ensures that the area projection of uncertainties into measurement space is not larger than the sensor field of view. If this is the case, a search over the projected space would likely be required to regain a track on the object in question. To track custody, the trace of the area projection of object uncertainties into measurement space for each space-based observer is also recorded. Finally, it is important to consider the utilization of each sensor. The tasking methodology ensures constant use of observers when feasible, but

the diversity of objects tracked is also relevant. Considering the number of unique objects observed over a rolling period helps elucidate modes of operation for an observer. One may then address whether an observer becomes stuck on objects with large uncertainties, whether the observer suite is nearing steady-state lower bounds on estimates, and in comparing the tasking histories of each observer, gain insight into the comparative effectiveness of various sensors.

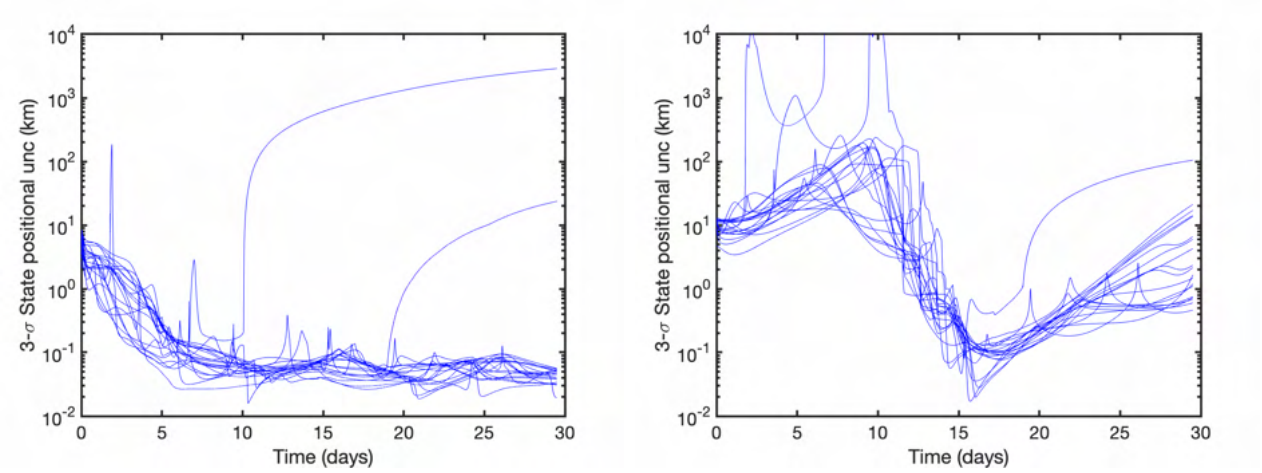


(a) Evolving uncertainty traces using a 3:1 L2 Northern Halo observer.

(b) Evolving uncertainty traces using a 1:1 L1 Lyapunov observer.

Figure 3.4: Positional uncertainties across a 500 object catalog. Contours are outlined as a percentage of the full catalog with 3-σ positional uncertainties below the contour line.

First, in Figure 3.4, evolution of uncertainty traces for the full 500 object catalog is considered. Median uncertainties on the order of 10 kilometers 3-σ are observed for each case, with some key differences. The halo observer achieves slightly improved performance, with especially notable improvements as target objects near the lunar region of cislunar space. Relative motion and observer-target distance are important considerations, something that is made apparent in the large reduction of uncertainty during the Lyapunov observer close lunar approach at approximately 15 days. This behavior is most apparent when studying specific orbit families that evolve near the moon. To best elucidate this point, best case uncertainties are first presented. These uncertainties are computed with the assumption that an observer shall measure an object at each opportunity it has to do so. Therefore, these uncertainties act as a lowest achievable bound for comparison, from which evolving structures may be discerned.



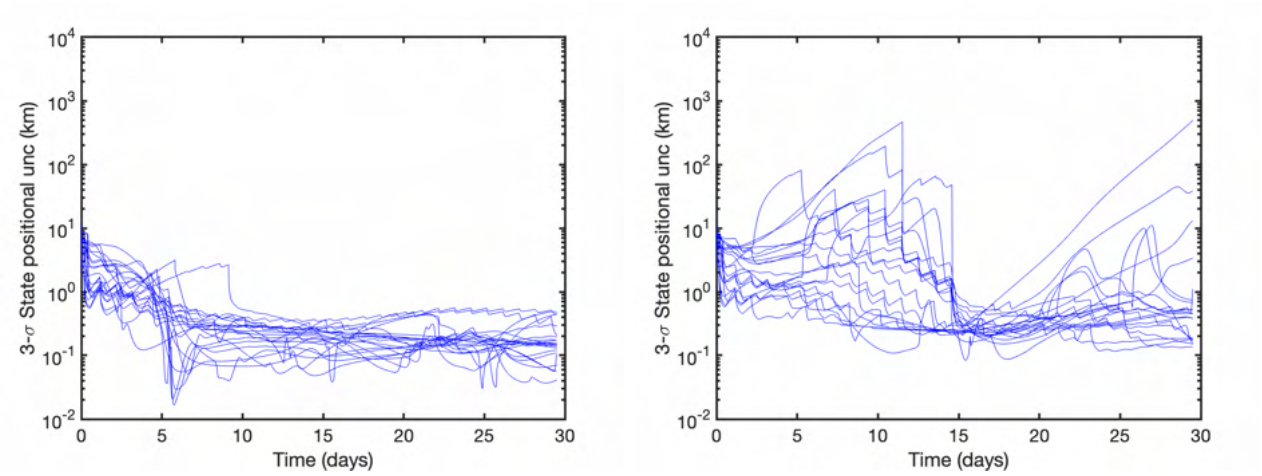
(a) Best case uncertainty traces using a 3:1 L2 Northern Halo observer.

(b) Best case uncertainty traces using a 1:1 L1 Lyapunov observer.

Figure 3.5: Best case uncertainties for the L2 Northern Halo family using a variety of observers.

In Figure 3.5, these lower bounds are illustrated for the L2 Northern Halo family of target objects. We generally note best case uncertainty bounds on the order of 100 meters for the family using the Halo observer, but two objects, evolving on the smallest orbits in the family, do not have convergent uncertainty bounds. These objects actually are given trajectories that would lead to near lunar impacts, challenging observation during close lunar approaches and introducing large amounts of dynamical uncertainty during these periods. Interestingly, the Lyapunov observer achieves better performance during its close approach at 15 days, but is otherwise challenged, with best case uncertainties observed generally an order of magnitude greater than the Halo case. This is a combination of less positional information being available as target objects grow more distant, and less rate information being available as there is less relative motion between the observer and the target. When the Lyapunov observer meets both of these criteria for making high quality observations during the close approach, the resulting estimates are also quite strong, but otherwise, the observer struggles to maintain estimates over the family.

Similar structures are observed in best case uncertainties for the L1 Lyapunov family, seen in Figure 3.6. Here, we note additional support from the ground-based observers, apparent in the jagged evolution of uncertainty traces as nighttime observations become available. Interestingly, many of the challenges visible in the L1 Lyapunov observer - Northern Halo target case still arise when considering observations of the



(a) Best case uncertainty traces using a 3:1 L2 Northern Halo observer.

(b) Best case uncertainty traces using a 1:1 L1 Lyapunov observer.

Figure 3.6: Best case uncertainties for the L1 Lyapunov family using a variety of observers.

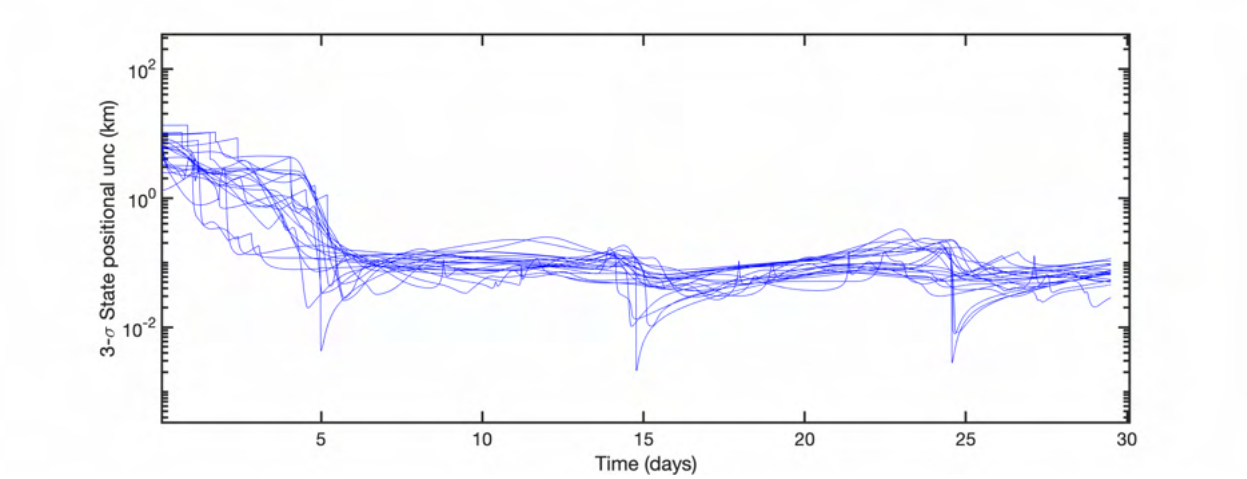


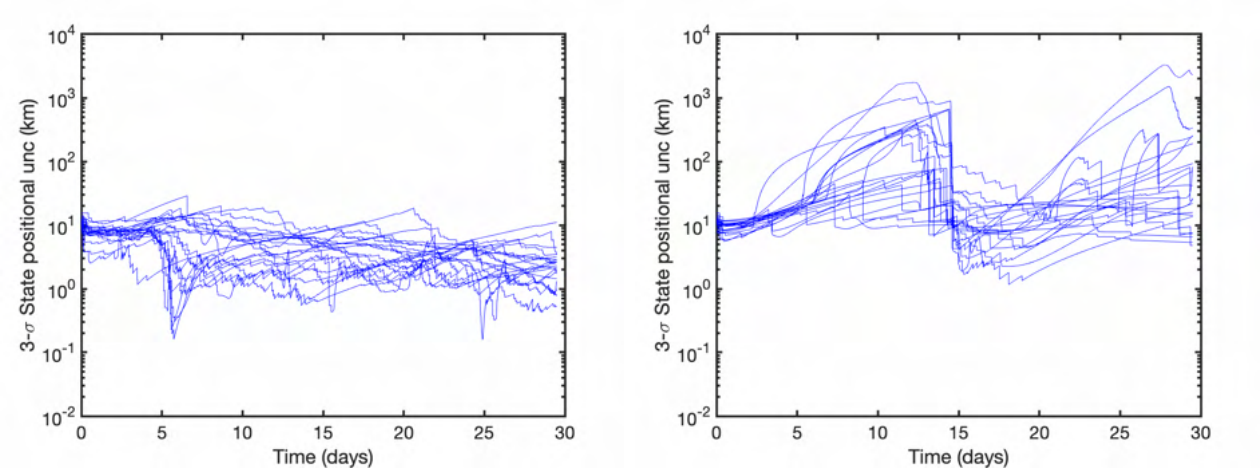
Figure 3.7: Best case uncertainties for the L1 Northern Halo family using a L2 Northern Halo observer.

L1 Lyapunov family with the L1 Lyapunov observer. The observer is still plagued by larger distances on average, and slow relative motion outside of the observer close lunar approach lead uncertainties to grow comparatively large.

The importance of relative motion for observation is further elucidated by considering best case uncertainties when the L2 Northern Halo observer studies the L1 Northern Halo family. This scenario, visualized in Figure 3.7, is characterized by three observer close lunar approaches at approximately 5, 15, and 25 days into the simulation. During these close approaches, the observer moves closer on average to the target family, and large reductions in object uncertainties are achieved.

A good way to understand this influence notionally is to consider what happens when measurement information is projected into state space. From a single measurement with rate information or short sequence of measurements, a four-dimensional subset of state space is immediately observable, and range and range rate may be considered an unobservable subset. The covariance ellipsoid may not be reduced in the range and range rate directions, and further significant reduction in the volume of the covariance ellipsoid may only be achieved as it rotates relative to the observer, such that range and range rate become observable. Dynamics notwithstanding, this rotation is a function of angular rates relative to the observer, and the more quickly this rotation occurs, the greater the reduction in uncertainty.

With these structures in mind, tasking results for specific families are now considered. First, Figure



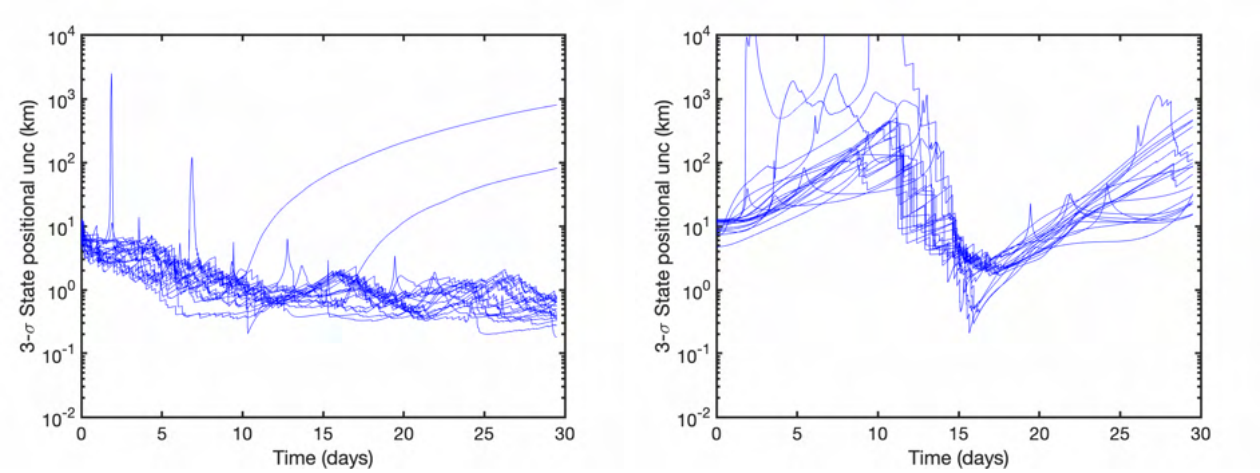
(a) Tasking uncertainty traces using a 3:1 L2 Northern Halo observer.

(b) Tasking uncertainty traces using a 1:1 L1 Lyapunov observer.

Figure 3.8: Tasking uncertainties for the L1 Lyapunov family using a variety of observers.

3.8 demonstrates results for the L1 Lyapunov family of target orbits.

The structures observed are consistent with the best case results expected. Again, a large decrease in uncertainties is seen as the L1 Lyapunov observer experiences a close approach, but outside of this short period, the L2 Northern Halo observer achieves greater performance. Mean positional uncertainties are on the order of 5 kilometers using the L2 Northern Halo observer, generally an order of magnitude better than uncertainties using the L1 Lyapunov observer.



(a) Tasking uncertainty traces using a 3:1 L2 Northern Halo observer.

(b) Tasking uncertainty traces using a 1:1 L1 Lyapunov observer.

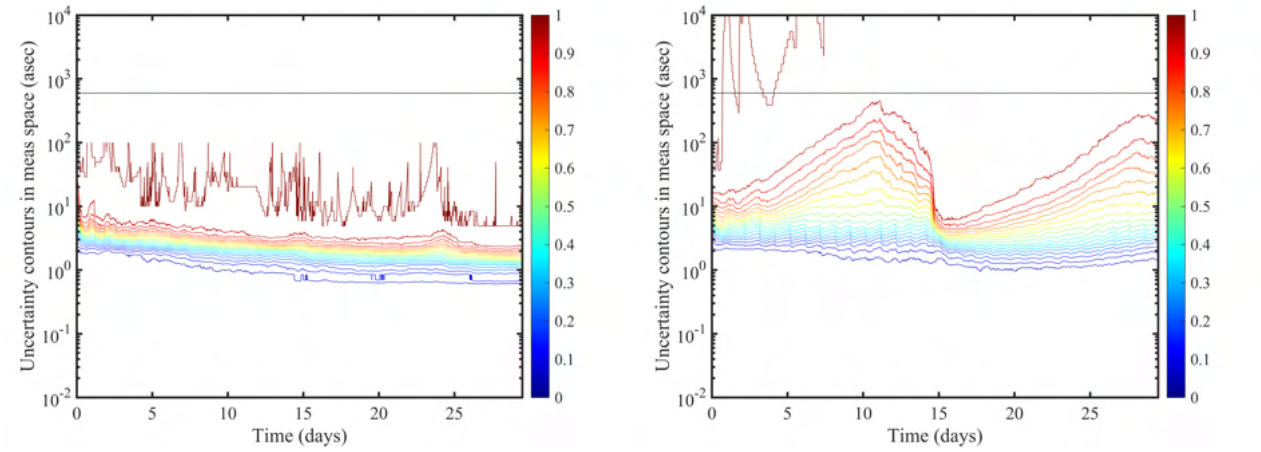
Figure 3.9: Tasking uncertainties for the L2 Northern Halo family using a variety of observers.

Figure 3.9 considers tasking uncertainty traces for this set objects following trajectories in the L2 Northern Halo family. The apparent structures are much like those visible in best case results. Here, it is notable that the smallest orbits in the family, found not to be trackable throughout the simulation in best case results, quickly diverge in the tasking simulation, with fast apparent growth in positional uncertainty during close approaches 2 and 9 days into the simulation. Uncertainties for the objects in the L2 Northern Halo observer case diverge at the same time as is found in the best case study. In general, greater performance is achieved using the L2 Northern Halo observer, with uncertainties for the majority of objects maintained to approximately 5 kilometers. The L1 Lyapunov observer, on the other hand, is comparatively inconsistent outside of the close approach, as is the general trend.

Next, an important factor of interest is whether custody is maintained for each object. This is the case if the projection of state uncertainty into measurement space remains smaller than the sensor field of view throughout the simulation period. In Figure 3.10, this projection is visualized for each space-based observer. Again, contours are provided describing the percentage of the catalog with projected uncertainties below the threshold. Note that the projection may also be made into the field of view of ground-based observers, but since performance of each space-based observer is of interest, the quality of custody maintenance is best expressed in terms of the space-based observers. A reasonable upper bound of 10 arc minutes 3 - σ is applied as a point at which maintaining custody is challenging; this bound is plotted against the square root of the trace of angular uncertainty.

Outside of spikes during close approaches, uncertainty projections remain below approximately 10 arcseconds throughout simulation for the L2 Northern Halo observer. This is also the case for most objects in the L1 Lyapunov observer case, but for a subset of near-lunar objects, some correlation remains between uncertainty projections and distance between the observer and objects. Further study is needed to explicitly determine the source of this behavior, but this structure may be explained by challenges in observability in addition to previously discussed difficulties in gaining information.

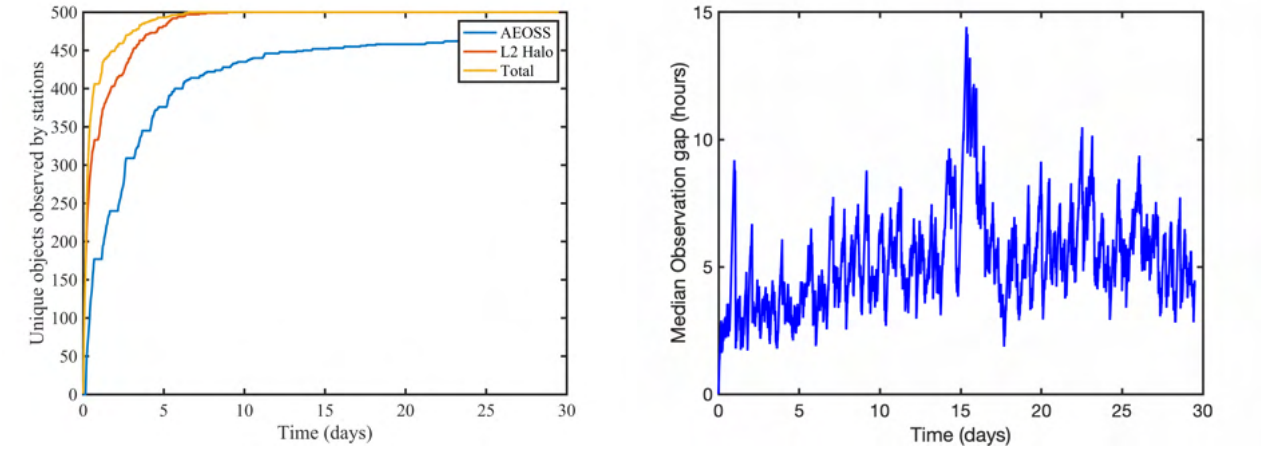
Finally, it is quite useful to consider not just how uncertainties evolve, but how often objects must be observed to maintain estimates. In addition to uncertainties, decisions are tracked throughout the simulation, and from this information, the time since an object was most recently observed may also be extracted. This



(a) Uncertainties projected into the measurement space of a L2 Northern Halo observer.

(b) Uncertainties projected into the measurement space of a L1 Lyapunov observer.

Figure 3.10: Uncertainty spread contours in measurement space over time.



(a) Unique objects observed by each sensor over time, as well as total objects detected in the catalog.

(b) Median observation gaps for the L2 Northern Halo family with a L2 Northern Halo observer.

Figure 3.11: Information on observation frequencies with a L2 Northern Halo observer.

analysis is visualized for the L2 Northern Halo observer, and Figure 3.11 illustrates trends in observation of the full catalog and the L2 Northern Halo family. All objects are detected within approximately 5 days of simulation, and objects local to the space-based observer are largely observed multiple times a day. Interestingly, the observation gap increases near new moon at approximately 15 days into simulation, where there is little contribution from the ground-based sensor.

One may addtionally consider whether there is any correlation between observation gaps and the

stability of the trajectory a target follows. The lack of such a correlation indicates that more critical indicators for observation frequency are observational challenges inherent to a trajectory and how quickly relative geometries change. Interestingly, within the studied catalog, changes in observer-target geometries tend to occur at a higher rate for more unstable trajectories; as those trajectories diverge further from stable points in the Earth-Moon system, the objects studied are more mobile in the rotating frame. Figure 3.12 presents median observation gaps simulated for each object across the catalog of periodic trajectories, using the L2 Northern Halo and ground-based observers.

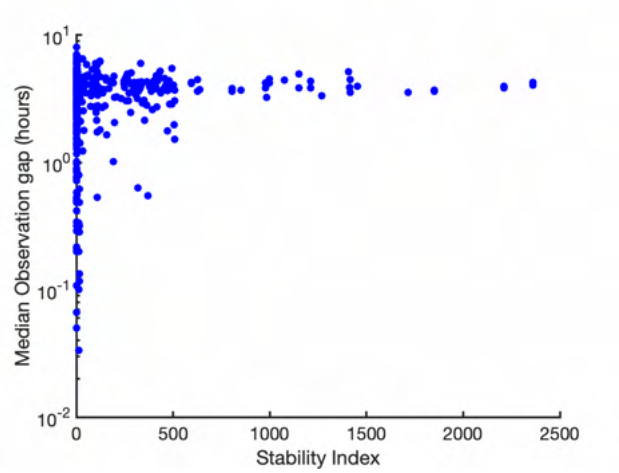


Figure 3.12: Median time between observations using a L2 Northern Halo and ground-based observers.

Interestingly, there is little intuitive structure to be extracted from this result, and the vast majority of objects requiring comparatively high-frequency observations follow relatively stable trajectories. These objects tend to originate from two classes. First, small orbits about stable points, especially the L1 and L2 Lagrange point, are never accessible from the ground-based observer. For these cases, the L2 Northern Halo observer tends to take frequent observations when possible, leading to a structure in which somewhat large gaps occur when no observer can detect an object, then a set of measurements frequent in succession when the object becomes detectable.

The second class of objects requiring comparatively high frequency observations originates from trajectories that remain distant in position space from both the ground-based observer and the space-based observer. The observation structure of this class is an artifact of the reward methodology for Monte Carlo

Tree Search, in which covariances for each object in the catalog are weighted on the same scale. Prior discussion outlined the influence of distance between an observer and a target on positional information received in an observation. Intuitively, then, more frequent observations are required to maintain positional uncertainties for more distant objects at the same scale as positional uncertainties for nearby objects. Further analysis could be performed in the future to better elucidate this point, looking at distinct families with greater fidelity to make a direct comparison.

3.2.3 Discussion and Conclusions

In this contribution, MCTS is applied to the catalog maintenance problem. alongside realistic sensor models, measurement models, trajectories, and object parameters. Results demonstrate the effectiveness of Monte Carlo Tree Search, and offer unique perspectives on the viability of space-based optical observers for cislunar catalog maintenance. These results allow for a gain in understanding of factors of importance for maintaining estimates on cislunar targets.

Further developments on this methodology could consider a variety of differing rollout schemes with information-theoretic foundations. In addition, the methodology may be benefited by recent developments utilizing policy learning with deep neural networks for action sampling. It would be feasible for this methodology to be augmented by learning via self-play.

The developed simulations also generated a vast amount of data; further analysis is possible regarding observer efficacy with a narrower focus on target families of interest. Future work could attempt to probe specific benefits of observer suites for orbits of interest, especially the use of multiple space-based and lunar ground-based observers.

Finally, it is important to keep in mind that this research makes specific focus on the problem of catalog maintenance, rather than the search for new objects or follow-up tasking on prior tracklets. Future research could pose the larger problem as a multi-objective optimization, with a desire to unify these varying modes of operation.

3.3 Sensor Tasking for Feasible Set Search and Follow-up Observation

This contribution instead considers application to the sensor tasking problem in which an admissible region (AR) is previously generated from a too-short-arc and follow-up observation is desired. In this scenario, an optical sensor may search over the AR or another extended projection in measurement space at some point after the full state becomes observable. Prior literature studying this problem includes research by Murphy and Hobson [80, 53], and it is first useful to revisit the analysis Murphy performs, which may be utilized to infer useful actions for the MCTS expansion phase. Methods for estimation over the search set as the problem are then considered, and finally, a result scenario is presented for a geostationary follow-up observation task.

3.3.1 Search Set Behavior

A more extensive mathematical basis for the evolving behavior of search sets is outlined by Murphy et al. [81]. This section provides a general overview of that work as needed for analysis of the search sets and development of the metrics utilized in this contribution. The intention of this section is to first consider in a general manner how a projected subset of state space might evolve over time. This analysis is then applied to an optical observer, and knowledge of search set behavior is used to inform sampling methods for MCTS tasking.

Assume that the dynamical system

$$\dot{\vec{x}} = \vec{f}(\vec{x}, t; \mathbf{p}) \tag{3.18}$$

is associated with the flow function ϕ.

$$\mathbf{x}(t_1) = \phi(t_1; \mathbf{x}(t_0), t_0; \mathbf{p}) \tag{3.19}$$

We assume the flow function may be applied to a subset of states, $\mathcal{S}$, defined by a p-dimensional vector of constraint equations, $\mathbf{c}$, in the global set of state dimension m, $\mathcal{X} = \mathcal{R}^m$.

$$\mathcal{S}(t) = \phi(t; \mathcal{S}(t_0), t_0) \tag{3.20}$$

$$\mathcal{S}(t_0) = \{\mathbf{x} \in \mathcal{X} : c_i(\mathbf{x}) \leq 0,\ i = 1 : p\} \tag{3.21}$$

Sets are analyzed in the measurement space $\mathcal{H}_o$ of dimension n associated with observer o. At a given time, the search set may be projected onto the field of regard of the observer using the measurement function

$$\mathbf{h} : \mathcal{R}^m \to \mathcal{R}^n \tag{3.22}$$

Note that the full state $\mathbf{x}$ may be partitioned into a determined subset d and unobservable subset u.

$$\mathbf{x} = \begin{bmatrix} \mathbf{x_d} & \mathbf{x}_u \end{bmatrix} \tag{3.23}$$

The partition $\mathbf{x}_d$ is representative of the subset of the full state that is observable through $\mathbf{h}$. The observable portion of the set is described as

$$\mathcal{S}_d = \{\mathbf{x}_d : \mathbf{x}_d = \mathbf{h}(\mathbf{x}),\ \mathbf{x} \in \mathcal{S}\} \tag{3.24}$$

Of particular interest for this contribution is the area of the projected set and how this changes over time. Leveraging the representation of the projected set as a vector field, one may apply Gauss's theorem. For an arbitrary vector field $\mathbf{F}$ defined over a compact subset of $\mathcal{R}^n$, $\mathcal{T}$, with a piecewise smooth boundary $\partial\mathcal{T}$, the theorem relates the flux of that vector field through the closed surface to the divergence of the vector field within the region. Note that a n-dimensional integral over a subset is designated a volume integral dV, while a $n-1$-dimensional integral over the boundary of that subset is designated a surface integral dS.

$$\int_{\mathcal{T}} \nabla \cdot \mathbf{F} dV = \oint_{\partial\mathcal{T}} (\mathbf{F} \cdot \hat{n}) dS \tag{3.25}$$

Using Gauss' theorem, one may consider how the projected area of the search set will change over time. The area of the region of interest may be expressed as the total integral,

$$A_{\mathbf{h}}(\mathcal{S}) = \int_{\mathcal{S}_d} dV = \frac{1}{n} \int_{\mathcal{S}_d} \nabla \cdot \mathbf{x}_d dV = \frac{1}{n} \oint_{\partial\mathcal{S}_d} (\mathbf{x}_d \cdot \hat{n}) dS \tag{3.26}$$

where n is the measurement space dimensionality and $\partial\mathcal{S}_d$ is the boundary of the projected set. Following this result, the Leibniz integral rule may be applied to take time derivatives to arbitrary order. In general, the following result is found.

$$\frac{d^n}{d^n t} A_{\mathbf{h}}(\mathcal{S}) = \oint_{\partial\mathcal{S}_d} (\sum_{i=1}^{n} \mathbf{x}_d^{(i)} \cdot \frac{\partial \mathbf{x}_d^{(n-i)}}{\partial \mathbf{x}_d}) \cdot \hat{n}\ dS \tag{3.27}$$

This result may be utilized to describe the area of a subset of the projected set $\mathcal{S}$ at an arbitrary epoch. Murphy discusses additional considerations that must be made for this projected set. Generally the projection itself is not necessarily an injective function, and $\mathcal{S}$ may be "folded" in a variety of manners, such that multiple points in the unobservable set may map to the same point in $\mathcal{S}_d$. The supremum of flux out of a boundary associated with this set of feasible states is necessarily applied to describe changes in the search set over time. Fortunately, at least in the case of admissible regions, the projection may be expected to be injective in the short term, when follow-up observation is desired.

3.3.1.1 Optical Applications

The developed background can immediately applied for an optical observer. It is assumed that an admissible region-based search set $\mathcal{A}$ is previously defined at time t_0, and right ascension and declination measurements are desired at time t. This admissible region is a two-dimensional manifold in the range (ρ) and range-rate ($\dot{\rho}$) half plane projected into six-dimensional state space. Note that the admissible region is formed from an attributable vector with initially determined angles and angular rates. It is most useful to define the problem as follows. Right ascension is defined as α and declination is defined as δ.

$$\mathbf{x}_d = \begin{bmatrix} \alpha & \delta \end{bmatrix}^T \tag{3.28}$$

$$\mathcal{A}_d(t) = \{ \begin{bmatrix} \alpha & \delta \end{bmatrix}^T : \begin{bmatrix} \alpha & \delta \end{bmatrix}^T = \mathbf{h}(\phi(t; \mathbf{x}, t_0); \mathbf{o}(t)), \mathbf{x} \in \mathcal{A} \} \tag{3.29}$$

The divergence of the velocity vector field for this sensor $\dot{\mathbf{x}}_d = \begin{bmatrix} \dot{\alpha} & \dot{\delta} \end{bmatrix}^T$ may be expressed as

$$\nabla \cdot \dot{\mathbf{x}}_d = \frac{d\dot{\alpha}}{d\alpha} + \frac{d\dot{\delta}}{d\delta} \tag{3.30}$$

The area of the set in the sensor field of regard is computed as

$$A_{\mathbf{h}}(\mathcal{A}) = \frac{1}{2} \oint_{\partial \mathcal{A}_d} (\mathbf{x}_d \cdot \hat{n}) d\mathcal{A} = \frac{1}{2} \oint_{\partial \mathcal{A}_d} -\delta \cos(\delta) d\alpha + \alpha d\delta \tag{3.31}$$

Assuming the set is projected as a grid, set of particles or triangulation, this line integral may be computed numerically over the surface bounds. The area rate of change reflects a similar pattern and is

computed as

$$\frac{d}{dt}A_{\mathbf{h}}(\mathcal{A}) = \oint_{\partial\mathcal{A}_d}(\dot{\mathbf{x}}_d \cdot \hat{n})d\mathcal{A} = \oint_{\partial\mathcal{A}_d} -\dot{\delta}\cos(\delta)d\alpha + \dot{\alpha}d\delta \tag{3.32}$$

These results may be extended to compute higher order time derivatives of the search set area. It is noted that these higher order derivatives are explicitly dependent on the dynamical system associated with the problem, and become more challenging to compute. The second order derivative is generalized in Equation 3.33 as

$$\frac{d^2}{dt^2}A_{\mathbf{h}}(\mathcal{A}) = \oint_{\partial\mathcal{A}_d}\frac{\partial}{\partial t}(\dot{\mathbf{x}}_d \cdot \hat{n})d\mathcal{A} = \oint_{\partial\mathcal{A}_d}\left(\ddot{\mathbf{x}}_d \cdot \hat{n} + \dot{\mathbf{x}}_d \cdot \frac{\partial \dot{\mathbf{x}}_d}{\partial \mathbf{x}_d} \cdot \hat{n}\right)d\mathcal{A}. \tag{3.33}$$

Here, both $\ddot{\mathbf{x}}_d$ and $\dot{\mathbf{x}}_d \cdot \frac{\partial \dot{\mathbf{x}}_d}{\partial \mathbf{x}_d}$ are a function of the full state at any point in the projected set. $\ddot{\mathbf{x}}_d$ may be found analytically, but requires acceleration information from the full state, and further analysis is needed for the second term. The variation of the determined subset at time t with respect to initial admissible region coordinates may be expressed as

$$\frac{\partial \mathbf{x}_d(t)}{\partial \mathbf{x}_u(t_0)} = \frac{\partial \mathbf{x}_d(t)}{\partial \mathbf{x}(t)}\frac{\partial \mathbf{x}(t)}{\partial \mathbf{x}(t_0)}\frac{\partial \mathbf{x}(t_0)}{\partial \mathbf{x}_u(t_0)}\Big|_{\mathbf{x_u}(t_0)} \tag{3.34}$$

$$= H\Phi(t,t_0)\frac{\partial \mathbf{x}(t_0)}{\partial \mathbf{x}_u(t_0)}\Big|_{\mathbf{x_u}(t_0)} \tag{3.35}$$

where H is the measurement Jacobian and $\Phi(t,t_0)$ is the state transition matrix. Similarly,

$$\frac{\partial \dot{\mathbf{x}}_d(t)}{\partial \mathbf{x}_u(t_0)} = \frac{\partial \dot{\mathbf{x}}_d(t)}{\partial \mathbf{x}(t)}\frac{\partial \mathbf{x}(t)}{\partial \mathbf{x}(t_0)}\frac{\partial \mathbf{x}(t_0)}{\partial \mathbf{x}_u(t_0)}\Big|_{\mathbf{x_u}(t_0)} \tag{3.36}$$

$$= \dot{H}\Phi(t,t_0)\frac{\partial \mathbf{x}(t_0)}{\partial \mathbf{x}_u(t_0)}\Big|_{\mathbf{x_u}(t_0)} \tag{3.37}$$

and $\dot{H}$ is the Jacobian of the derivative of the measurement function. The desired matrix may then be found as

$$\frac{\partial \dot{\mathbf{x}}_d}{\partial \mathbf{x}_d} = \frac{\partial \dot{\mathbf{x}}_d(t)}{\partial \mathbf{x}_u(t_0)}\left(\frac{\partial \mathbf{x}_d(t)}{\partial \mathbf{x}_u(t_0)}\right)^{-1}\Bigg|_{\mathbf{x_u}(t_0),t} \tag{3.38}$$

These derivatives may then be applied to a Taylor series expansion to provide an analytic solution as an infinite series or an approximate series for the area of the search set projection over time. [81] provides

analysis for the error in area computation as a function of the number of terms utilized in the expansion, as well as a generalization for higher order derivatives.

The strategies in this paper for tracking the growth of the search set over time benefit from this theory and apply a particle representation of the admissible set. Equation 3.32 may be explicitly computed for subsets of the region using knowledge of particles over time, and long term growth of subsets of the region may be observed by tracking particles over time. In using this methodology, this work diverges from the strategy in [81] requiring computation of high order derivatives, but recognizes the clear application of these analytic concepts to developing search heuristics.

3.3.2 Rollout Heuristics for Follow-up Tasking

Several methods for sampling actions over a feasible set in measurement space are next outlined. First, the relative density of the projected set in measurement space is considered. Using a particle representation of the admissible region, it is trivial to approximate a portion of the probability distribution being captured by a given action by studying what particles would lie in the associated field of view. This methodology, described when referenced further as the probability of detection heuristic, is visualized in Figure 3.13. If a Gaussian mixture representation is utilized, this methodology is easily modified. A Gaussian integral may be computed over the field of view using methodologies like Genz integration [43] or expectation propagation [27].

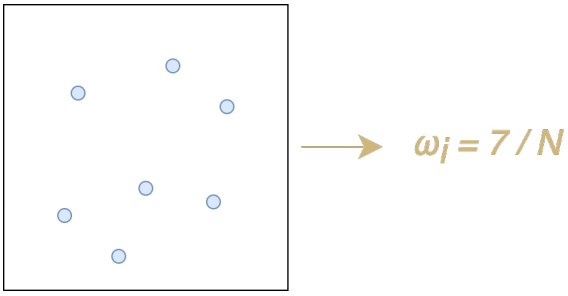


Figure 3.13: Probability of detection heuristic.

Alternatively, one can weight actions by considering how a region may change in the future. Two approaches are developed in this regard. Given a desired action, the projected area of all unobserved particles

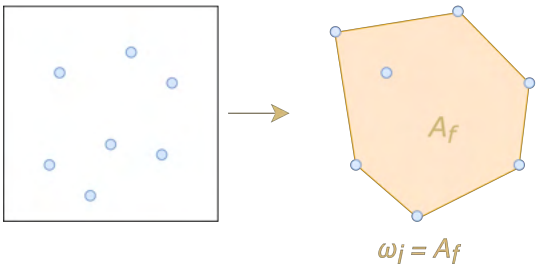


Figure 3.14: Final area lookahead heuristic.

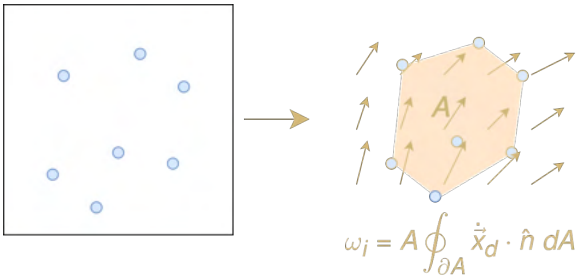


Figure 3.15: Immediate change in area heuristic.

within the field of view at the end of the search period may be considered as a weight from which to sample. This process is shown in Figure 3.14, and in further reference, is described as the lookahead heuristic. This quantity may be evaluated by propagating particles or mixands forward in time, or the methods outlined by Murphy may again be applied [81]. In addition, the more immediate change in area of these particles may be considered. The immediate rate of change in area of the region studied may be computed using the divergence of the measurement velocity vector field as outlined in the previous section. This rate of change is then reweighted by the area of the studied region to form a third set of sampling weights, hereby referred to as the immediate change in area heuristic. This weighting scheme is shown in Figure 3.15. Note that this final methodology leads to additional challenges, in which the area rate of change is not strictly positive. Because of this, it is useful to here consider a deterministic rollout method, in which at a given timestep, the nth new action chosen is the nth "best" as evaluated by the rollout policy. Further consideration for these policies is given after development of a mathematical basis in the next section.

These methodologies may be modified for a Gaussian mixture representation in several ways. First,

the mean states of each mixand may be considered a particle for the purposes of sampling. Second, the final area lookahead heuristic may be evaluated to an appropriate level of variance, again utilizing Gaussian integrals.

3.3.3 Estimation over Feasible Sets

Given a generated set of mixands, one may now consider how mixands are updated when tasking decisions are made and measurements are taken. The simpler scenario is that of a newly made detection. Here, a typical measurement update for a Gaussian sum filter may be performed. Unscented measurement updates are utilized [59], with likelihood-based weight updates in addition such that

$$\omega_i = \frac{\omega_i' \mathcal{L}_i(\mathbf{y}, R)}{\sum_{j=1}^{L} \omega_j' \mathcal{L}_j(\mathbf{y}, R)} \tag{3.39}$$

and

$$\mathcal{L}_i(\mathbf{y}, R) = \mathcal{N}\left(\mathbf{y} - \mathbf{h}(\mathbf{x}_i), H P_i H^T + R\right). \tag{3.40}$$

Updating the PDF when no detection is made introduces further complexities. First, one must consider the probability that the target SO lies in a given field of view (FOV) of an optical sensor. If mixand uncertainties are sufficiently small in range space relative to the range between the mixand and the optical sensor, one may assume that the probability of detection p_D is uniform over the mixand. Then, the cumulative likelihood of observing the target is

$$P(\mathbf{y} \neq 0) = \int_{\text{FOV}} p_D(\mathbf{z}) P(\mathbf{z}) d\mathbf{z} \tag{3.41}$$

Applying the mixand representation of the PDF, a distinct probability of detection may be assumed for each mixand, and

$$P(\mathbf{y} \neq 0) = \int_{\text{FOV}} \sum_{i=1}^{L} p_{D,i} \omega_i P(\mathbf{z}|k=i) d\mathbf{z} \tag{3.42}$$

One must then evaluate the observation likelihood $P(\mathbf{z}|k=i)$ conditioned on mixand i. Note that the transformation from state space to measurement space is locally linear with the same assumptions on range. Therefore, the mixand density projected into measurement space is still Gaussian, with

$$P(\mathbf{z}|k=i) \approx \mathcal{N}\left(\mathbf{h}(\mu_i),\ HP_iH^T\right). \tag{3.43}$$

It is also clear that the integral is separable, and thus,

$$P(\mathbf{y} \neq 0) = \sum_{i=1}^{L} \left(p_{D,i}\omega_i \int_{\text{FOV}} P(\mathbf{z}|k=i)d\mathbf{z} \right) \tag{3.44}$$

To compute the likelihood of observation, one must then evaluate the Gaussian integral in measurement space over the rectangular FOV. Methods such as Genz integration [43] or expectation propagation [27] may be applied for this purpose. Given this result, one must now consider how a null detection may affect existing mixands. It is clear that the probability a null detection occurs is the complement of Equation 3.44. Then, one must consider how to apply this result to knowledge on each mixand. Working from first principles, we may apply Bayes rule.

$$P(\mathbf{x}|k=i,\ \mathbf{y}=\emptyset) = \frac{P(\mathbf{y}=\emptyset|\mathbf{x},k=i)P(\mathbf{x}|k=i)}{P(\mathbf{y}=\emptyset)} \tag{3.45}$$

Immediately, challenges arise when considering the term $P(\mathbf{y}=\emptyset|\mathbf{x},k=i)$, the probability a null detection is made, conditioned on the SO state captured by mixand i. Consider the PDF for mixand i in further detail, with the temporary assumption that the projection of probability density into measurement space is larger in spread than the sensor field of view. At any point within the support of the projected PDF outside of the field of view, the probability of a null detection must be unity, since that point simply cannot be captured during the observation. This leads to a scenario in which a subset of the mixand is scaled as a function of the probability of detection, while the remainder is unaffected. The structure of this update is clearly non-Gaussian, and is further illustrated in Figure 3.16.

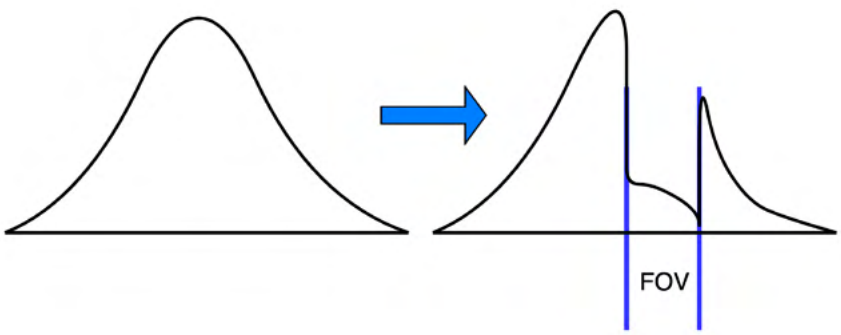


Figure 3.16: Non-Gaussianity in a negative information update.

This behavior may be broadly categorized into distinct groups. First, the density captured in the field of view can be relatively small; this commonly occurs when the mixand is either a large normalized distance away from the sensor FOV in measurement space, or quite large in spread in measurement space as compared to the field of view. In this case, a null detection's impact on the PDF is negligible, since the probability $P(\mathbf{y} = \emptyset|\mathbf{x}, k = i)$ is effectively unity. Alternatively, the probability density may be almost entirely captured by the sensor FOV, in which case the entire mixand is rescaled and the null detection probability for the given mixand nears zero. The third case, in which the sensor field of view overlays a significant portion of the mixand, but not the entirety, requires further consideration. To resolve this case, a splitting method is developed to reduce the projected spread of mixands in measurement space and ensure negative information updates remain Gaussian.

3.3.3.1 Oriented Gaussian splitting

Consider a mixand with mean μ and covariance P in state space $\mathcal{S}$. This mixand may be defined relative to an observer $\mathcal{O}$ with measurement function $\mathbf{h}$. Letting the measurement function be differentiable, the local behavior of $\mathbf{h}$ may be examined utilizing the gradient. For each scalar measurement, this leads to a tangent that may be normalized and considered as the direction in state space leading to a maximal change in the associated scalar measurement. The resultant set of tangent vectors forms the basis of a tangent space of dimension n, where n is the rank of the measurement Jacobian. Each tangent vector may be explicitly computed as

$$\hat{l}_i = \frac{\frac{\partial h_i}{\partial \mathbf{x}}|_\mu^T}{|\frac{\partial h_i}{\partial \mathbf{x}}|_\mu^T|}. \tag{3.46}$$

It is desired to split the mixand into a set of mixands with unknown means and equivalent covariance P, while enforcing that the combined PDF of the resultant mixands captures the same first and second moment of the original mixand. Additionally, it is desired that the observer is also taken into consideration, such that the new mixands are perturbed about the defined measurement bases. Without loss of generality, let the new set contain $2m+1$ mixands, where m is the dimension of the measurement space. Let one mixand be placed at the original mean, with another pair of mixands evenly distanced along the the tangent vector associated with each scalar measurement with distance $a_k P^{\frac{1}{2}}$, where $P^{\frac{1}{2}}$ is the matrix square root of the original mixand covariance. Note that this methodology is typical when considering transformations on Gaussians, exemplified by the work of Havlak and Campbell [49]. The matrix square root is incorporated to ensure that similar mixand structure is in place, and to enforce positive-definiteness. Additionally, let each new mixand have equivalent weight $\omega = \frac{1}{2m+1}$. The mean of the resultant distribution is then

$$\mu_{\text{TOT}} = \sum_{k=0}^{2m} \omega_k \mu_k \tag{3.47}$$

$$= \frac{1}{2m+1}\left(\mu + \sum_{k=1}^{m}\left(\mu + a_k P^{\frac{1}{2}}\hat{l}_k\right) + \sum_{k=1}^{m}\left(\mu - a_k P^{\frac{1}{2}}\hat{l}_k\right)\right) = \mu. \tag{3.48}$$

The covariance of the new distribution must also be determined. For a Gaussian mixture, the total covariance is

$$P_{\text{TOT}} = \sum_{i=1}^{N} \omega_i P_i + \sum_{i=1}^{N} \omega_i \left(\mu_i - \mu_{\text{TOT}}\right)\left(\mu_i - \mu_{\text{TOT}}\right)^T. \tag{3.49}$$

In this case, then, we find

$$P_{\text{TOT}} = \sum_{k=0}^{2m} \omega_k P^* + \sum_{k=0}^{2m} \omega_k \left(\mu_k - \mu\right)\left(\mu_k - \mu\right)^T. \tag{3.50}$$

$$= P^* + 2\sum_{k=1}^{m} \frac{1}{2m+1} a_k^2 P^{\frac{1}{2}} \hat{l}_k \hat{l}_k^T P^{\frac{T}{2}} = P \tag{3.51}$$

With this result, one can determine the updated covariance

$$P^* = P - 2\sum_{k=1}^{m} \frac{1}{2m+1} a_k^2 P^{\frac{1}{2}} \hat{l}_k \hat{l}_k^T P^{\frac{T}{2}} \tag{3.52}$$

With this result in mind, it is still important to consider whether P^* is positive definite. The key determination is whether the eigenvalues of P^* are all positive. It is possible to left and right multiply P^* by $P^{-\frac{1}{2}}$ and maintain the definiteness of the matrix such that

$$P_{\text{NORM}} = P^{-\frac{1}{2}} P^* P^{-\frac{T}{2}} \tag{3.53}$$

$$= I - 2\sum_{k=1}^{m} \frac{1}{2m+1} a_k^2 \hat{l}_k \hat{l}_k^T \tag{3.54}$$

Now, any eigenvector for the summation must also be an eigenvector of P_{NORM}. For each eigenvector ν_i, the associated eigenvalue of the summation is λ_i, and the eigenvalue of P_{NORM} must be enforced to be strictly greater than zero such that

$$\lambda_P = 1 - \lambda_i > 0 \tag{3.55}$$

This allows for gains a_k to be chosen as a function of the structure of the cumulative outer product of the tangent space. First, consider the outer product $\hat{l}_k \hat{l}_k^T$. Since the tangent vectors are normalized, this matrix is symmetric and positive semi-definite with a single eigenvalue at unity. This offers an upper bound on the eigenvalues of the summation. If each tangent vector is collinear, the summation will then have a single nonzero eigenvalue

$$\lambda^* = \frac{2ma_k^2}{2m+1} \tag{3.56}$$

that must be less than unity. Gains must in general then be no greater than

$$a_k < \sqrt{\frac{2m+1}{2m}}. \tag{3.57}$$

Note that these gains may be increased if there is further knowledge of the measurement space. If two unit vectors are orthogonal, one may infer that the sum of outer products of these vectors has two eigenvalues at unity in addition to zero eigenvalues. This argument may be expanded, considering the full set of tangent vectors in the space. The maximum eigenvalue of the summed matrix must be no greater than, assuming gain is held constant,

$$\lambda^* = \frac{2a_k^2}{2m+1}(m+1-\text{rank}(H)) \tag{3.58}$$

where rank(H) is the rank of the measurement Jacobian, which is equivalent to the rank of the set of tangent vectors.

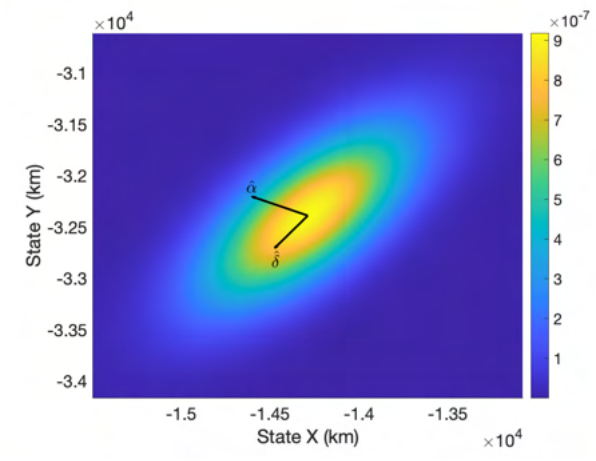
This result may explicitly be demonstrated for an optical case in which right ascension and declination measurements are taken. This is the critical case for negative information updates, because projected mixands must be split in angular space to ensure the update remains Gaussian. For this case, the dimension of the measurement is m = 2. It is also known that the gradients of right ascension and declination are orthonormal in state space. Therefore, we have

$$\lambda^* = \frac{2a_k^2}{5} < 1 \tag{3.59}$$

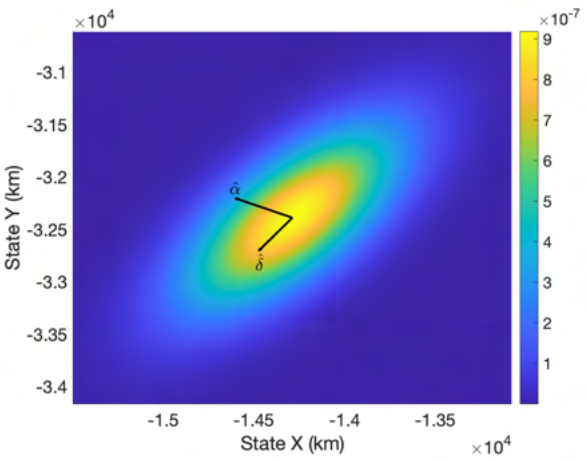
and

$$a_k < \sqrt{\frac{5}{2}}. \tag{3.60}$$

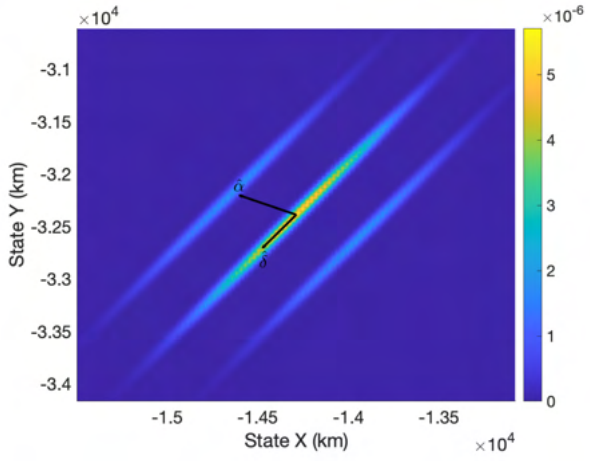
It is then ensured that newly generated mixands continue to have positive definite covariances as splitting occurs. Note that while this gain is the theoretical maximal bound to ensure positive definite covariances, it is not advisable to use. As gain nears this quantity, the projected PDF of the split mixands will become drastically small in the eigendirection associated with the eigenvalues nearing zero. This behavior is visualized in Figure 3.17, where the PDF is presented for a mixand before a split, with a split using reasonable gains, and a split with gains nearing the theoretical maximum. With this projection, it is most clear that mixand uncertainties become quite small in the $\hat{\alpha}$ direction, and they also become quite



(a) Original mixand. Measurement tangent vectors $\hat{\alpha}$ and $\hat{\delta}$ are also displayed.



(b) Mixands split with gain $a = 1.0$. Density is approximately equal to a).

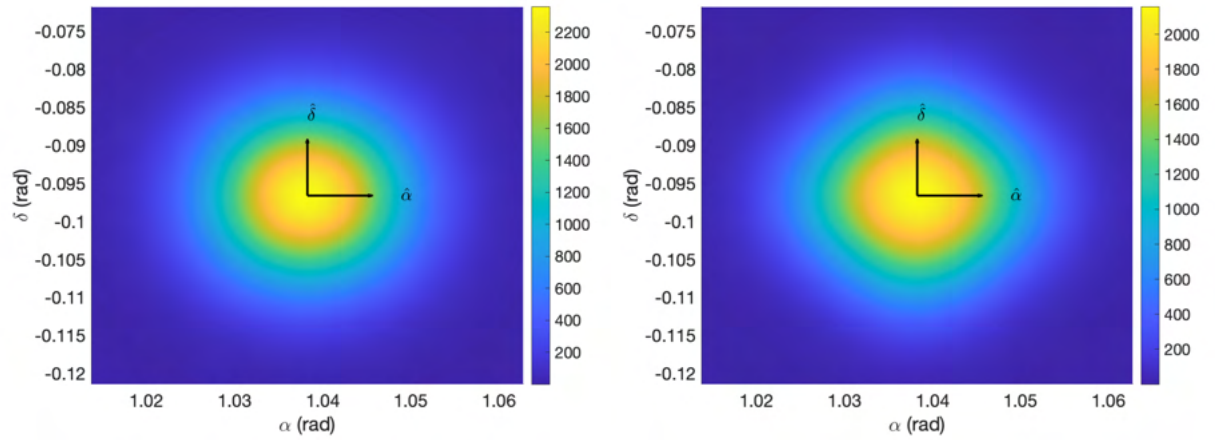


(c) Mixands split with gain $a = \sqrt{\frac{5}{2}}$. The first and second moments match those of a), but relative entropy is large compared to b).

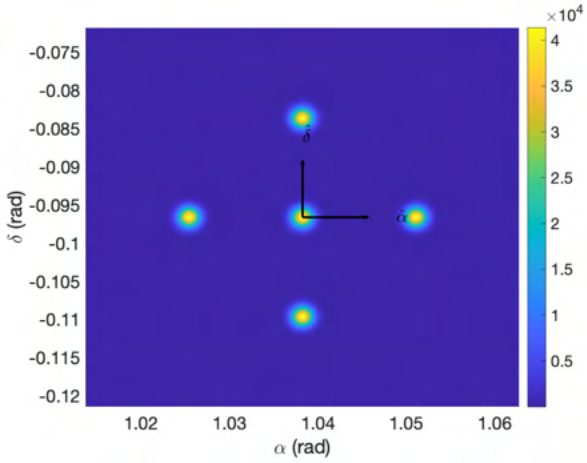
Figure 3.17: Mixand densities in a subset of position space before and after splitting.

small in the $\hat{\delta}$ direction (which is largely out of the plane). More specifically, as the gain approaches the theoretic maximum, the projections of uncertainties into measurement space collapse into discrete points. This behavior is demonstrated in Figure 3.18, where mixands are projected into measurement space.

Understanding this behavior helps visualize a trade in gain selection. Increased gain ensures a reduction in spread for the mixands utilized, but also increases relative entropy between the original mixand and resultant mixand. Care must be taken to allow new mixands to become as small as needed while maintaining



(a) Original mixand projected into measurement space. (b) Mixands split with gain $a = 1.0$ in measurement space.



(c) Mixands split with gain $a = \sqrt{\frac{5}{2}}$. Mixands are reduced into discrete points.

Figure 3.18: Mixand densities in measurement space before and after splitting.

a distribution sufficiently close to the original.

3.3.3.2 Updating Gaussians

Now that a formulation for splitting Gaussians such that they will be sufficiently small in measurement space, a criterion for determining whether a mixand shall be split must be established. It is logical to incorporate some measure of offset from the center of the sensor FOV in measurement space and the comparative spread of uncertainty to the sensor FOV. A critical Mahalanobis distance may be defined, such that a mixand shall only be split if

$$D_M(\mu_i, P_i; \mathcal{O}) < d^*. \tag{3.61}$$

Additionally, the angular spread may be defined as the square root of the maximal diagonal value of the projected covariance trace

$$s = \sqrt{\max(HPH^T)} > s^*. \tag{3.62}$$

This may be compared with a critical value that is a function of the diagonal field of view of the sensor.

Once mixands are rescaled such that they are either sufficiently distant in measurement space from the sensor FOV or of comparable size to the sensor FOV, the negative information update may be considered in further detail. Revisiting Equation 3.44, the problematic term may now be considered approximately discrete in that the mixand is either fully covered by the sensor FOV or is sufficiently far from the sensor. As such, it is now logical to consider Equation 3.44 as a weight update on each mixand in much the same manner as a particle filter, using the intermediate density

$$g(\mathbf{y}|k=i) = \frac{P(\mathbf{y}=\emptyset|\mathbf{x}, k=i)}{P(\mathbf{y}=\emptyset)} \tag{3.63}$$

$$= \frac{P(\mathbf{y}=\emptyset|\mathbf{x}, k=i)}{\sum_{j=1}^{L} P(\mathbf{y}=\emptyset|\mathbf{x}, k=j)} \tag{3.64}$$

The denominator can simply be considered a normalization, while the numerator may be approximately evaluated as

$$P(\mathbf{y}=\emptyset|\mathbf{x}, k=i) \approx \begin{cases} 1 & i \notin \text{FOV} \\ 1-p_D & i \in \text{FOV} \end{cases}. \tag{3.65}$$

Note that it is still useful to explicitly compute the Gaussian integrals over the FOV because of the stopping criterion on splitting; these results may be scaled by the tail probabilities computed.

3.3.4 Merging and Filter Outline

With the filter update fully expressed, one now must ensure there is no hypothesis explosion in mixands so that the filter remains computationally efficient. With a goal of minimizing Kullbeck-Liebler divergence during the merging process, the well-known Runnall's method is utilized [84]. A discrimination bound,

$$B(i,j) = \frac{1}{2}\left[(\omega_i + \omega_j)\log|P_{ij}| - \omega_i \log|P_i| - \omega_j \log|P_j|\right] \tag{3.66}$$

may be iteratively computed with the merged covariance for mixands i and j, P_{ij}. Merging is iteratively performed until a threshold maxima of mixands is reached. Pruning may also be applied if weights are sufficiently small, but care must be taken to ensure that this does not disregard mixands split during the negative information update. With this methodology in place, the full filter is outlined in Figure 3.19.

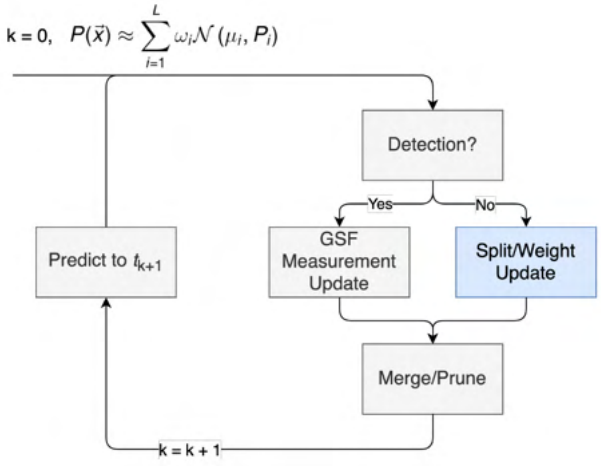


Figure 3.19: Gaussian Sum Filter diagram. Major contribution highlighted in blue.

3.3.5 Follow-up observation in practice

The methodology is now presented for a case in which a prior detection is made and follow-up observation is desired. This problem is modeled as a MDP, with

- $\mathcal{S}$: The feasible region of measurement space within which the space object resides.
- $\mathcal{A}$: Binned subsets of measurement space that an agent may choose to observe.

- $\mathcal{T} : \mathcal{S} \times \mathcal{A}$: Propagation of the unobserved region in the full state space, which is again projected into measurement space. Generally, Keplerian dynamics are utilized, as they are applied for initial formulation of the admissible region.
- R: Change in projected area of the unobserved region or cumulative probability of detection.
- $\gamma \in [0, 1]$: the discount factor over time, impacting the prioritization of short term rewards.

The object tracked has the true initial state (in kilometers and kilometers per second)

$$\mathbf{x} = \begin{bmatrix} -27100 & -32300 & -100 & 2.36 & -1.98 & 0 \end{bmatrix} \tag{3.67}$$

in the Earth-centered inertial (ECI) frame. An observation is made by an observer at the initial ECI position (in kilometers)

$$\mathbf{o} = \begin{bmatrix} 517.859 & -5281.538 & 3526.190 \end{bmatrix}. \tag{3.68}$$

An admissible region is then formed from knowledge of observer state and the attributable vector (in radians and radians per second) as

$$\mathbf{y} = \begin{bmatrix} \alpha & \delta & \dot{\alpha} & \dot{\delta} \end{bmatrix} = \begin{bmatrix} -2.36716 & -0.093581 & 7.30762e-05 & -1.53752e-09 \end{bmatrix}. \tag{3.69}$$

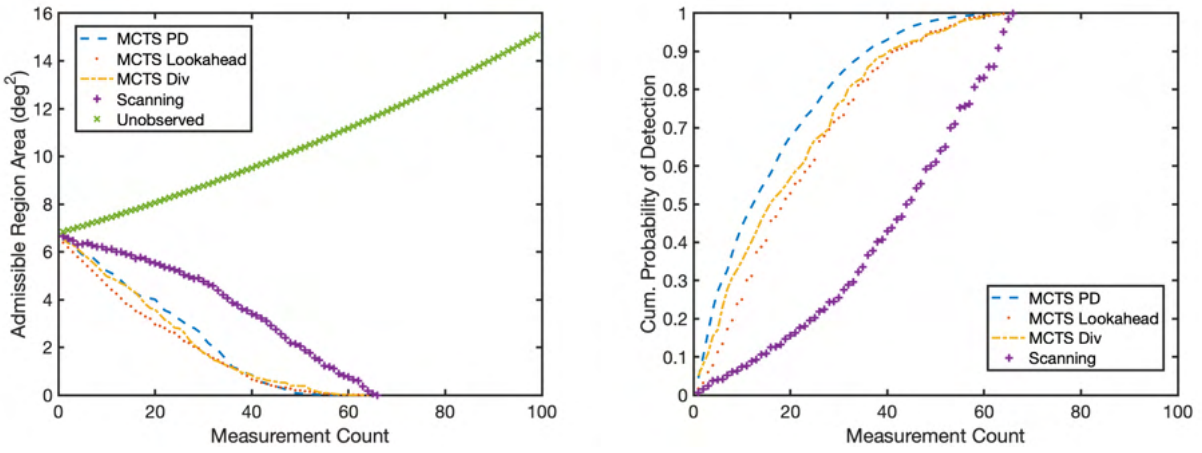
To constrain the admissible region, several assumptions are made on feasible orbits. The assumption is made that the target SO is on an elliptic trajectory following two body dynamics, with energy less than 0. A maximal eccentricity of 0.3 is applied. Finally a minimum radius of periapsis of 6500 km is assumed, ensuring the target will not collide with Earth. The admissible region is approximated using a mixture representation.

In order to provide a comparison to the MCTS methodologies, a scanning strategy is developed, utilizing scenario-specific knowledge that the fastest-growing subset of the search space is that where right ascension is largest. A striping method is applied, moving along increasing declination and decreasing

right ascension, hoping to capture these quickly growing subsets early in the search scenario. As the reward functions utilized in MCTS, admissible region area and probability of detection will be the primary indicators of the relative merits of differing search strategies.

This baseline scanning pattern is able to complete the search campaign in the allocated period of 100 observations at a 15 second observation cadence (25 minutes of instrument time). 66 observations are required, and this result may be assumed as a minimum goal for the MCTS methodology. With this knowledge, tree search results are presented in comparison. MCTS is run for 100 iterations down the search tree at each time step, with a search depth of 40 observations. As the search period nears the end of the 100 allocated observations, this depth is reduced accordingly such that only 100 observations are considered across the scope of the problem.

In Figure 3.20, it is clear that each MCTS method offers advantages over the naive scanner. One may observe that the area rate of change heuristic offers a middle ground in that it trails the lookahead heuristic in reducing area and it trails the probability of detection heuristic in achieving cumulative probability of detection. Each method offers clear advantages over the scanning methodology. The general structure of the growth of the unobserved region over time is also shown.



(a) Admissible region area as a function of observations taken.

(b) Cumulative probability of detection as a function of observations taken.

Figure 3.20: Comparisons between each MCTS scenario and the naive scanner.

With these results outlined, MCTS is then applied in combination with the developed estimation

paradigm. As in the initial search case, observations are taken at a 15 second cadence until the search region is exhausted; the first follow-up observation is performed two hours after the initial detection. The admissible region, with initial area in measurement space of approximately 7 $\deg^2$, is exhausted with a sequence of 66 observations, using a sensor with a square field of view of 0.25 $\deg^2$. At the time of the 66th observation, the projected admissible region has an area of approximately 12 deg^2. Over the course of this tasking solution, a single observation is made at timestep $t_{15} = 7410\ s$ when the observer is pointing at the angular coordinates $\alpha = -1.82939$ rad, $\delta = -0.093562$ rad. In each other case, no detection is made, and the negative information is processed. Figure 3.21 visualizes the effects of processing negative information 7 observations into the tasking scenario. Note that probability density in the observed subset of measurement space has been greatly reduced.

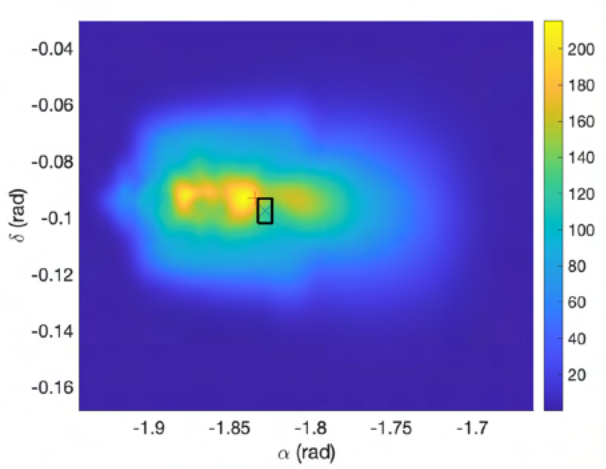


Figure 3.21: The admissible region projected into measurement space after 7 null detections.

In the Figure, the most recent observation is represented by a shaded rectangle, and the true state, at $\alpha = -1.8346$ rad and $\delta = -0.09318$ rad, is marked by a plus sign. Observed regions are most visible from $\alpha = -1.87$ to -1.85 radians and $\delta = -0.1$ to -0.09 radians. Because of these reductions, unobserved subsets of the projected admissible region are now comparatively more likely. Indeed, the density at the true state is approximately 10 percent higher than prior to the processing of any negative information. The negative information in this case effectively describes missed detections on a subset of mixands in the ensemble, increasing the likelihood that each other mixand is "truth" and may be associated with the true state.

It is also important to consider the behavior of the filter when a measurement is received. Figure 3.22 demonstrates the reduction of the mixture when this occurs at time t_{15}. Here, the projected area of the mixture is now reduced to that of measurement uncertainty, assumed to be on the order of 5 arcseconds in this simulation. After this observation is made, there is negligible effect on the state estimate through further processing of negative information, but this is still critical to do in real scenarios, when the likelihood of false alarm measurements is non-negligible. Finally, Figure 3.23 visualizes estimation error over the course of the simulation. Note that the estimation error remains within the covariance bounds throughout the simulation, and is greatly reduced when the follow-up observation is received.

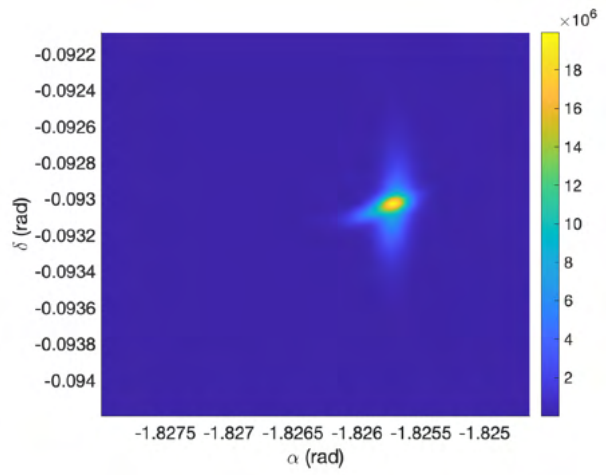
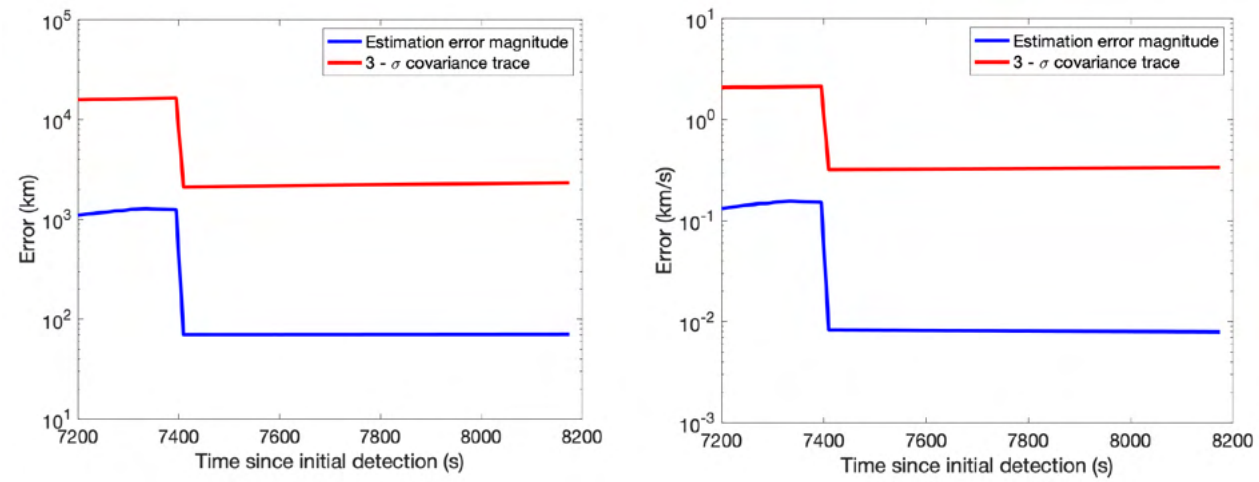


Figure 3.22: The admissible region projected into measurement space after a follow up detection is made. Uncertainty in angular space is equivalent to measurement uncertainty.

The covariance bounds are a bit conservative in this context because of the time between observations. Interestingly, the full state is only weakly observable, and there is still significant uncertainty in the sensor line of sight direction after the measurement, on the order of 2000 km $3-\sigma$. However, uncertainty is sufficiently reduced such that only a small number of mixands are needed to effectively represent probability density after the measurement. There is little impact on the processing of negative information on estimation error, but this is as expected. Before a detection is made, the PDF is still quite large in state space; while some mixands are eliminated through processing negative information, the ensemble mean is not very useful for such a large PDF, and in this case, it just happens to be that the original mean is somewhat close to the

true state. It would be interesting to determine whether the ensemble mean becomes more useful in cases where most of the feasible region is exhausted before a follow-up detection is made.



(a) Positional estimation error over the course of the simulated observation campaign.

(b) Velocity estimation error over the course of the simulated observation campaign.

Figure 3.23: Estimation errors before and after the detection is made.

3.3.6 Discussion and Conclusions

Generally, the rollout policies developed performed very well for a realistic case. An interesting avenue of further research would be in the process of discretizing the search space. A relatively efficient gridding method is utilized, but it is likely that further improvements could be seen in treating the sub problem of determining feasible actions as a covering or packing problem. That problem is NP-hard, but it is likely that utilization of more rigorous partitioning of the action space using approximate covering algorithms could lead to more efficient options, especially when the search region becomes somewhat sparse.

Also of interest for further study is the extension of this research to probabilistic admissible regions (PARs) and reachable sets. Incorporation of a non-unity observation probability when actions are taken is a useful consideration for challenging cases with dim and distant, near-cislunar objects, as well as for flexibility if comparatively poor sensors are in operation. Searching reachable sets is a critical problem as well, and this methodology is quite applicable for maintaining state estimates on maneuvering objects. It is important to note that the tasking process is essentially analogous for these extensions, but that particle weights may

change according to the probability density of the PAR or a priori knowledge on a maneuver or unmodeled force.

Additional avenues may also considered for the second contribution of this work. While the logical focus for processing negative information in this paper is the context of optical observations, this methodology could also be extended for utilization with radar measurements and a variety of novel observational techniques such as event sensors. The key use case is any situation in which the spread of the projected state estimate exceeds the sensor field of regard in measurement space.

Theoretic bounds on the size of Gaussian mixand splits are outlined, but it is noted that further perturbation along measurement axes comes with an increase in Kullbeck-Liebler divergence from the original mixture. Future work may aim to formally pose this splitting methodology as an optimization in which Kullbeck-Liebler divergence is minimized with an additional cost objective comparing the spread of resultant mixands to a target value (for example, a sensor field of view).

This splitting methodology is utilized to ensure that the weight update remains Gaussian, and it is important to make further comparisons to particle representations of the state estimate, as this state representation is considered for instantiation of the tasking contributions. Particles may also be utilized with likelihood updates within a sensor field of view when no detection is made, but the Gaussian mixture representation may be considered advantageous in this context for several reasons. First, it allows for a smooth representation of probability density over the support of the state estimate. This is quite useful in regions of measurement space where particles may be quite sparse, avoiding concerns of depletion. Additionally, there are computational advantages to utilizing Gaussian mixture representations, in that a large set of particles need not be sampled; this is especially critical when particles are sampled from a high-dimensional state space. In the context of admissible regions, this is not the case, but a curse of dimensionality becomes rather important when considering search over reachable sets in Cartesian position-velocity space.

In general, results demonstrate advantages in utilization of negative information before follow-up detections are made. Extensions of these results will be made considering a variety of target orbits, utilization of more interesting dynamics, and follow-up tracking of maneuvering objects.

Finally, it is important to note that an end goal of this contribution is incorporation in an online

manner. The problem of tracklet association for multiple generated admissible regions is solved, but this research is beneficial when there is immediate need for follow-up observation. It would be quite interesting to consider whether this tasking objective may be leveraged in tandem with other objectives such as pure search and catalog maintenance.

Chapter 4

Catalog Maintenance of Maneuvering Space Objects

With methods in place that address major sensor tasking objectives, MCTS may now be considered for more complex scenarios. This chapter presents applications of MCTS to the realistic problem of tasking agile space objects. Goals in this chapter are twofold. We first wish to ensure that state estimation is successful through maneuvers across a catalog of maneuvering objects; this is an especially challenging problem in environments that are already nonlinear such as the XGEO domain. Further, we wish to develop a tasking methodology that considers the potential of future maneuvers, such that target SOs might be observed as they perform stationkeeping or transfer maneuvers. To the first point, this chapter develops a novel estimator for maneuver detection and estimation. Analysis of the local dynamical flow is then leveraged to better inform maneuver potential. These points of analysis may then be utilized to augment the previously established rollout policies in catalog maintenance. A result scenario is then applied demonstrating the use of lunar optical sensors for detection and tracking of SOs performing stationkeeping maneuvers in Earth-Moon Halo orbits. These observers are then stress-tested, studying feasibility of the developed estimation for tracking transfers between L1 and L2 Halo orbits in chaotic domains.

4.1 Maneuver Estimation in Space Object Tracking

The maneuver estimation problem is well-studied in both the SDA community and across target-tracking literature as a whole. Outside of SDA, applications range from missile defense and air traffic control to the study of crowd movement and vehicle collision avoidance. Classical methods for the problem include the Interacting Multiple Model filter [76], Dynamic Model Compensation [103], or Input Estimation [20].

These methods either aim to estimate an unknown acceleration as a function of measurement residuals, or utilize a set of estimators, each with a probability of accurate representation of the underlying system and a Markovian model for switching between estimators.

More recently, there has been much interest in the utilization of optimal control methods for maneuver detection and estimation. Holzinger presented the application of control distances to correlate object detections and detect maneuvers [56]. Lubey incorporates optimal control policies into the Optimal Control Based Estimator (OCBE) [74], to automatically correct for and estimate dynamics mismodelling. More recently, such methods were extended into nonlinear domains as the unscented OCBE (U-OCBE) by Greaves [46]. The next section aims to address a gap in this literature, presenting optimal smoothing techniques for the U-OCBE.

4.1.1 The Optimal Control Based Estimator

The optimal control based estimator, developed by Lubey et al. [74], combines optimal control processes and optimal estimation. Cost functions are formulated combining boundary estimation error with a control Lagrangian, such that the L2 norm of control effort is minimized. With assumptions on a nominal adjoint, the ballistic OCBE offers a Kalman-like structure. Adjoints evolve alongside the estimated state in an augmented state transition matrix (STM) Φ of dimension $2n$, where n is the dimension of the state. The STM is evaluated using the nominal state $\bar{x}(t)$ and adjoint $\bar{p}(t)$ trajectories with

$$\dot{\Phi}(t, t_0) = A(t)\Phi(t, t_0) \tag{4.1}$$

$$A(t) = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial \mathbf{x}} & -B(t)\tilde{Q}(t)B(t)^T \\ -\frac{\partial^2}{\partial \mathbf{x}^2}\left(\mathbf{f}^T\mathbf{p}\right) & -\frac{\partial \mathbf{f}}{\partial \mathbf{x}} \end{bmatrix}_{(t,\tilde{x}(t),\tilde{p}(t))} \tag{4.2}$$

The resultant STM is partitioned into 4 $n \times n$ submatrices relating to the state x, adjoint p, and cross terms. Adjoint estimates are evaluated at the prior epoch using the single step smoothed gain L_{k-1} alongside the measurement residual and mapped into control space using the adjoint subset of the STM. Note that since the ballistic OCBE nominally assumes no control, this is the first order estimate for the full control effort at a given epoch.

$$\delta\hat{\mathbf{p}}_{k-1|k} = -\hat{P}^{-1}_{k-1|k-1} L_{k-1} \left[\delta\mathbf{y}_k - H_k \left(\delta\hat{\mathbf{x}}_{k|k-1} + \nu_x(t_k)\right)\right] \tag{4.3}$$

$$\hat{\mathbf{u}}(t) = -\tilde{Q}(t)\frac{\partial \mathbf{f}^T}{\partial \mathbf{u}}\Phi_{pp}(t, t_{k-1})\delta\hat{\mathbf{p}}_{k-1|k} + \omega(t) \tag{4.4}$$

This structure is the critical factor for incorporation of the OCBE into the tasking methodology because control estimates are inferred over an arbitrary time period utilizing an adjoint estimate. The adjoints are normalized by the prior information, and as is observed in Equation 4.3, are explicitly dependent on measurement residuals. Logically, the ballistic OCBE utilizes control estimates to explain measurement residuals, but control authority is effectively maximized with the process noise term $\tilde{Q}(t)$. Care must be taken to correctly select process noise, especially in the context of maneuvering space objects, and methods that autonomously update the admissible control are critical to the problem.

For this purpose, further inspiration is taken from the adaptive process noise methods utilized by Lubey and Scheeres [73]. Lubey presents several metrics for evaluating whether dynamics mismodelling is present in the studied system, and for this problem, the OCBE measurement distance metric is selected as a means for evaluating whether process noise should be adapted. If this is the case, the metric is used as an output to a nonlinear root finder problem, and Newton-Secant methods are applied to determine a measurement distance-minimizing process noise. The OCBE measurement distance metric is evaluated as

$$D_M(\hat{\mathbf{x}}_{k|k}) = \frac{1}{2}\left(\mathbf{y}_k - \mathbf{h}(t_k, \hat{\mathbf{x}}_{k|k}\right)^T R_k^{-1}\left(\mathbf{y}_k - \mathbf{h}(t_k, \hat{\mathbf{x}}_{k|k}\right) \tag{4.5}$$

It is critical to determine both whether the process noise shall be adjusted and the horizon of observations to reprocess with an updated process noise. Modifications are made to the methods discussed by Lubey with this intention, recognizing that several scenarios may occur. In the case of a weak mismodelled force or small maneuver, measurement residuals are expected to grow slowly over a long time horizon. However, if a larger maneuver occurs, a sudden spike in the applied metric will be observed. As such, factors such as the magnitude of the distance metric and time between observations should be considered when determining a reprocessing horizon. This work utilizes a lagging horizon of 5 observations for triggering adaptive process noise evaluation. If the normalized OCBE measurement distance averages exceed a specified threshold

over the observation horizon, the adaptive process is applied. In addition, spikes in the most recent OCBE measurement distance are studied to determine whether the dynamics mismodelling occurred over a short time horizon. If this is the case, a large control authority is assumed, and only the most recent time and measurement updates are reevaluated with an updated process noise. If this is not the case, the lagging horizon is reprocessed.

Even with careful methods for applying process noise, the linearized ballistic OCBE is insufficient for state estimation in nonlinear dynamical systems such as the cislunar regime. Greaves presents an evaluation demonstrating that filters such as the extended Kalman filter struggle to converge for cislunar orbit determination and presents an unscented form of the ballistic OCBE as a successful solution [46]. The unscented modification acts quite similarly to the unscented Kalman filter [59], with the added need for tracking of the state transition matrix, a component that remains necessary for adjoint estimation. The process evolves as follows. The state estimate is first augmented by a nominal adjoint, and the unscented transform s is applied. Resultant sigma points are transformed through the dynamical system, and desampling s^{-1} is performed to apply the unscented time update. Note that the covariance should be desampled with the process noise matrix Q, which is extracted from the augmented state transition matrix. In the time update, it is also useful to track the cross-covariance $C_{xx,k}$ and a newly introduced unscented smoother gain $\tilde{L}_{k-1}$, where

$$C_{xx,k} = \sum_i W_{i-1}^c \left(\hat{\mathbf{x}}_{k-1,i} - \mu_{k-1}\right) \left(\hat{\mathbf{x}}_{k,i}^- - \mu_k^-\right)^T \tag{4.6}$$

$$\tilde{L}_{k-1} = C_{xx,k} \hat{P}_{k|k-1}^{-1}. \tag{4.7}$$

The full prediction step is outlined as follows.

$$\hat{X}_{k-1} = s(\hat{\mathbf{x}}_{k-1|k-1}, \hat{P}_{k-1|k-1} | \alpha, \beta, \kappa) \tag{4.8}$$

$$\hat{X}_k, \Phi(t_k, t_{k-1}) = \mathbf{f}(\hat{X}_{k-1}) \tag{4.9}$$

$$Q = -\Phi_{xp}(t_k, t_{k-1}) \Phi_{xx}(t_k, t_{k-1})^T. \tag{4.10}$$

$$\hat{\mathbf{x}}_{k|k-1}, \hat{P}_{k|k-1} = s^{-1}(\hat{X}_k, Q | \alpha, \beta, \kappa) \tag{4.11}$$

The unscented ballistic OCBE measurement update proceeds in an equivalent manner to the unscented Kalman filter. Revisiting Chapter 2, the sigma points may further be transformed into measurement space via some measurement function $\mathbf{y}$ resulting in a set of sigma points $\mathcal{Y}_{k|k-1}$ with dimension $k \times 2n$, where k is the dimension of the measurement. The mean and innovation are evaluated as

$$\hat{\mathbf{y}} = \sum_{i=1}^{2n} W_i^{\mu} \mathcal{Y}_{k|k-1,i} \tag{4.12}$$

$$S_{k|k-1} = \sum_{i=0}^{2n} W_i^{P} \left(\mathcal{Y}_{k|k-1,i} - \hat{\mathbf{y}}\right) \left(\mathcal{Y}_{k|k-1,i} - \hat{\mathbf{y}}\right)^T + R_k. \tag{4.13}$$

The innovation represents the nonlinear projection of state uncertainty into measurement space with measurement uncertainty incorporated. The cross-covariance between state and measurement space C_{xy} is also needed, and may be evaluated from the sigma points as

$$C_{xy} = \sum_{i=0}^{2n} W_i^{c} \left(\chi_{k|k-1,i} - \hat{\mathbf{x}}_{k|k-1}\right) \left(\mathcal{Y}_{k|k-1,i} - \hat{\mathbf{y}}\right)^T \tag{4.14}$$

The resultant form may then be used to directly apply the unscented measurement update.

$$K_k = C_{xy} S_{k|k-1}^{-1} \tag{4.15}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k(\mathbf{y}_k - \hat{\mathbf{y}}) \tag{4.16}$$

$$P_{k|k} = P_{k|k-1} - K_k S_{k|k-1} K_k^T = P_{k|k-1} - C_{xy} S_{k|k-1}^{-1} C_{xy}^T \tag{4.17}$$

In addition to the unscented ballistic OCBE, a novel U-OCBE Rausch-Tung-Streibel (RTS) smoother is presented, inspired by the unscented RTS smoother [86]. This method is applicable to the single step OCBE smoother and full OCBE smoother [74]. In order to transition these smoothers to an unscented form, several components are desired. First, we note a heavy dependence on the STM Φ. For nonlinear systems, a higher order representation of the flow is desired to accurately transform state information between epochs. In addition, a sigma point representation of the state estimate offers great utility throughout the smoother. To address the first need, the illustrated cross-covariance is utilized extensively. Comparing the gain $\tilde{L}_{k-1}$ to

the form used by Lubey, we recognize that the crosscovariance acts as a means to transform state innovation at epoch k into the state space at epoch $k-1$. With careful selection of unscented transform hyperparameters, this modification is expected to capture nonlinearities of the transformation to higher order than utilization of the state transition matrix. A sigma point approximation of the state component of the STM may further be evaluated as

$$\Phi(t_k, t_{k-1}) = \left(\hat{P}^{-1}_{k-1|k-1} C_{xx,k}\right)^T \tag{4.18}$$

The unscented OCBE smoother may then be expressed as

$$\hat{\mathbf{x}}_{k-1|l} = \hat{\mathbf{x}}_{k-1|k-1} + \tilde{L}_{k-1}\left[\hat{\mathbf{x}}_{k|l} - \left(\hat{\mathbf{x}}_{k|k-1} + \nu_x(t_k)\right)\right] \tag{4.19}$$

$$\hat{P}_{k-1|l} = \hat{P}_{k-1|k-1} + \tilde{L}_{k-1}\left[\hat{P}_{k|l} - \hat{P}_{k|k-1}\right]\tilde{L}^T_{k-1} \tag{4.20}$$

$$\hat{\mathbf{p}}_{k-1|l} = -\tilde{L}_{k-1}\left[\hat{\mathbf{x}}_{k|l} - \left(\hat{\mathbf{x}}_{k|k-1} + \nu_x(t_k)\right)\right] \tag{4.21}$$

$$\hat{\mathbf{u}}(t|l) = -\tilde{Q}(t)B(t)^T\Phi_{pp}(t, t_{k-1})\hat{\mathbf{p}}_{k-1|l} + \omega(t) \tag{4.22}$$

The resultant smoother may then be utilized to inform decision making. The smoother is applied over a lagging horizon, and a threshold adjoint norm is applied as a means for identifying maneuver likelihood. The time at which a potential maneuver was most recently flagged is tracked throughout tree search, and any object flagged within a decay time τ is given a scaled observational weight.

4.2 Stretching Dynamics for Maneuver Feasibility

It is critical to prioritize observation of objects during intervals over which maneuvers are most impactful in addition to those that are known to have recently maneuvered. With the assumption that targets are perturbed and maneuver about a nominal periodic orbit or the current best state estimate, one may explore the relative utility of maneuvers. The local effects of small maneuvers are considered using the linearized flow about the periodic orbit, and analysis of the state transition matrix will be leveraged, with

$$\Phi(t, t_0) = \int_{t_0}^{t} A(\tau)\Phi(\tau, t_0)d\tau, \quad \Phi(t_0, t_0) = \mathcal{I} \tag{4.23}$$

The state transition matrix is evaluated at a full period in order to form the monodromy matrix, and from the monodromy matrix, one may determine the stable and unstable manifolds about the orbit, as well as associated eigenvalues. It is largely useful to consider the unstable components of the perturbation from nominal. There is an array of literature considering Floquet mode control for stationkeeping, in which it is desired to cancel the unstable mode [93]. Intuitively, then, one may assume that a stationkeeping maneuver is more likely the larger the unstable mode grows. This feature is also worth considering for maneuvers not necessarily associated with stationkeeping. If a maneuver is performed to depart from the nominal orbit, it is expected that the unstable mode shall very quickly grow. As such, this feature is utilized to prioritize valuable departing maneuvers.

In addition to the monodromy matrix, recent stationkeeping literature has proposed the use of the Cauchy-Green stress tensor [48, 79], evaluated as

$$C(t, t_0) = \Phi(t, t_0)^T \Phi(t, t_0). \tag{4.24}$$

While the state transition matrix maps an initial perturbation to a final perturbation, the Cauchy-Green stress tensor acts as a means to evaluate the relationship between initial and final perturbation distance, where

$$|\delta\mathbf{x}(t)|^2 = \delta\mathbf{x}(t_0)^T C(t, t_0)\delta\mathbf{x}(t_0). \tag{4.25}$$

This relationship in itself could be used as an evaluation tool in rollout heuristics, but much as in the case of the monodromy matrix, eigendecompositions of the CGT are performed to evaluate stretching and compressing perturbation directions. It is useful to relate the eigendecomposition of the CGT to the state transition matrix, and in doing so, it is noted that the eigenvalues of the CGT are explicitly related to the singular values of the state transition matrix.

$$C(t,t_0) = U\Gamma U^T = \Phi(t,t_0)^T\Phi(t,t_0) = \left(V\Sigma U^T\right)^T\left(V\Sigma U^T\right) = U\Sigma^T\Sigma U^T \tag{4.26}$$

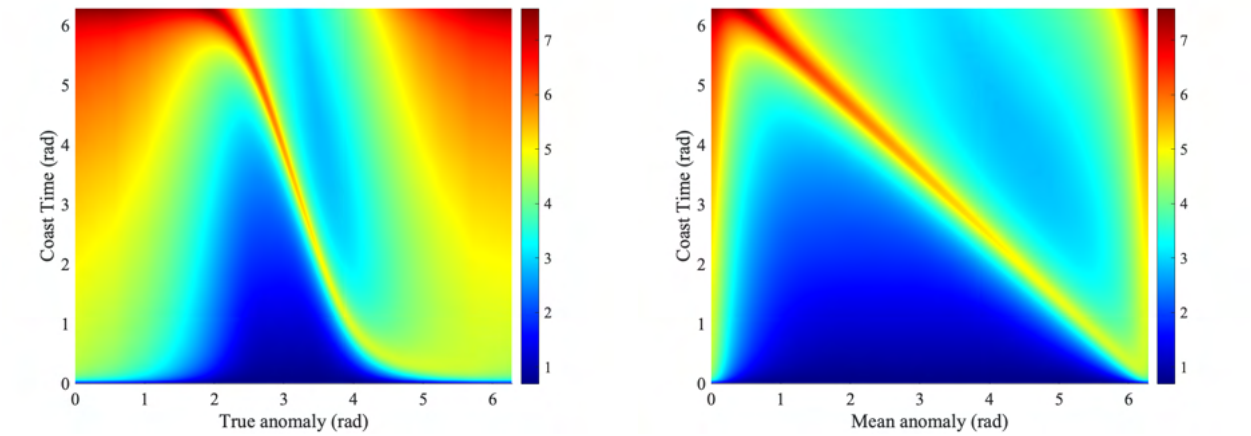
As Σ is a positive, real, diagonal matrix, the eigenvalues of the CGT are guaranteed to be real and positive. These eigenvalues are the square of the singular values of the state transition matrix and correspond to the magnitude to which stretching occurs in the associated stretching direction.

Without any knowledge of maneuvers occurring, it is sensible to make use of the eigenvalues of the CGT. Future work shall study the transformation of the estimated perturbation into the basis associated with the stretching directions, but initially, it is critical to consider the minimal and maximal eigenvalues of the CGT. Intuitively, these eigenvalues provide bounds on departure and restoring maneuvers, respectively. Taking inspiration from the stability index, one may aggregate these bounding eigenvalues using the CGT indicator, where

$$\rho(t,t_0) = \log_{10}\left(\sqrt{\lambda_{c,max}} + \frac{1}{\sqrt{\lambda_{c,min}}}\right). \tag{4.27}$$

The indicator is applied in log space to better address the relative magnitude of perturbations, especially since eigenvalues can be quite large near perilune. A square root is applied to consider the indicator in log-distance space. The CGT indicator is visualized for a sample L1 Southern Halo orbit across a variety of initial phases and coast times in Figure 4.1.

These results are compared with those presented by Muralidharan and Howell [79], noting that stretching is considered in the full state space, rather than in submatrices, and that coast time is always in mean anomaly. Incorporation of the minimum eigenvalue has a small effect in terms of damping minima and maxima, and values of the CGT indicator are largely driven by proximity to perilune at the termination of a coast arc. Visualization of the CGT indicator as a function of osculating true anomaly provides insight into maneuver impacts in a spatial sense, while visualization of the CGT indicator against mean anomaly provides broader understanding of maneuver response across time. Use of the CGT in tasking should be expected to prioritize observation of SOs during lunar flybys. This is counter to typical stationkeeping practices for Halo



(a) A L1 SHalo CGT indicator across initial true anomaly and coast time.

(b) A L1 SHalo CGT indicator swept against mean anomaly and a range of coast durations.

Figure 4.1: Characteristic structures in the CGT indicator.

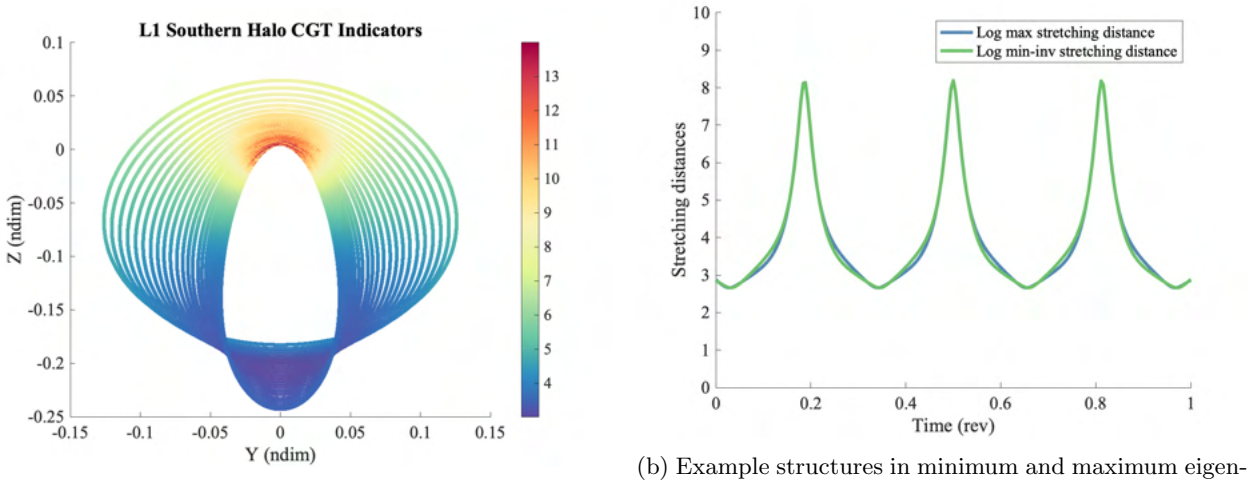
orbits, for which maneuvers are typically performed about apolune. This practice is largely implemented because of increased risk as injection errors are introduced; even though minimum stretching distance decreases, any impulse misaligned with this stretching direction can greatly increase the perturbation off nominal.

Additionally, the eigenvalues of the periodic CGT are closely related, and in Figure 4.2, we qualitatively observe that for any given true anomaly from perilune θ,

$$\lambda_{c,max}(\theta) = \frac{1}{\lambda_{c,min}(-\theta)} \tag{4.28}$$

Further research shall be applied to demonstrate this behavior analytically, but the key outcome of this is that because of orbital symmetry, the CGT indicator for an orbit is fixed at a given lunar distance, even though neither eigenvalue is fixed at these distances and symmetric about perilune. Because of this, a future avenue of research may consider whether this indicator may be approximated as a function of lunar distance with nominal orbit assumptions relaxed. This trend appears to generally hold when visualized across an orbit family, with additional correlation between stability index and mean stretching distance. To support this point, in Figure 4.2, the CGT indicator is presented across the L1 SHalo family and segmented by minimum and maximum eigenvalue for a sample orbit.

The CGT indicator may be applied to prioritize scenarios in which maneuvers are feasible and impactful in terms of object custody. Combined with covariance-based tasking and information on known



(a) The CGT indicator for the L1 Southern Halo family.

(b) Example structures in minimum and maximum eigenvalues of the CGT over an Earth-Moon synodic period.

Figure 4.2: Broad structures of the CGT indicator.

maneuvers, one may then scalarize a combined objective about which to prioritize tasking decisions.

4.3 A combined rollout heuristic

With methods for prioritizing objects that have previously maneuvered and identifying scenarios for impactful maneuvers in place, the developed methods must be unified in a manner that is ingested into MCTS rollout-based planning. A successful rollout policy shall combine these methods with the covariance minimization goal of catalog maintenance, incorporating state covariance and observer knowledge. The previous chapter applied measures such as the projection of covariance into the field of regard of an observer as a means for scoring the immediate value of an action. If the trace of the projection is significantly larger than that of measurement uncertainty, the observation offers significant information gain. Letting both this trace and the CGT indicator be normalized across the tracked catalog, one may assign feature weights

$$\omega_{y,i} = \frac{\mathrm{tr}(H_i P_i H_i^T)}{\sum_j \mathrm{tr}(H_j P_j H_j^T)} \tag{4.29}$$

$$\omega_{\rho,i} = \frac{\rho_i(t+T_i,t)}{\sum_j \rho_j(t+T_j,t)} \tag{4.30}$$

A maneuver flag ξ may then be incorporated to reweight actions associated with objects that have

recently maneuvered. The full sampling weight is then expressed as

$$\omega_i = \xi_i \left(\nu \omega_{y,i} + \eta \omega_{\rho,i} \right) \tag{4.31}$$

$$\xi_i = \begin{cases} 1 & t - t_{M,i} > \tau \\ \omega_M & t - t_{M,i} \leq \tau \end{cases} \tag{4.32}$$

Further tuning may be applied to prioritize covariance minimization, CGT custody, and reacquistion of maneuvering targets with hyperparameters ν, η, and ω_M.

4.4 Stationkeeping Catalog Maintenance with Lunar Optical Sensors

The developed methods are now applied to a sensor tasking scenario in which lunar observers are utilized to maintain estimates on a population of 100 space objects following Halo trajectories about the L1 and L2 Lagrange points. The problem is formulated as a POMDP as follows.

- $\mathcal{S}$: an ensemble of N space object states modeled as multivariate Gaussian random variables, along with completely known observer states.
- $\mathcal{A}$: Each observer may choose to observe a single space object, leading to an action space with dimension $M \times N$, where M is the number of observers and N is the number of space objects.
- $\mathcal{T} : \mathcal{S} \times \mathcal{A}$: a transition function between states over time conditioned on global sets of actions. In this case, Kalman measurement updates are performed on each tasked action if an object is detected, and space object states dynamically evolve under CR3BP dynamics. Note the major difference in this case as compared to the presented case in Chapter 3, in which an object may maneuver at this time; as such, unmodeled dynamics may also be applied within this transition function.
- R: reduction in covariance traces.
- $\mathcal{O}$: the space over which the environment may be observed, in this case, right ascension, declination, and associated rates.
- $\mathcal{H}$ the transformation of spacecraft states onto the celestial sphere in the field of regard of the applied observer.

Specification	Lunar Sensors
Aperture (m)	0.2
f-number	3
Pixel pitch (μm)	5
QE	0.8
Read noise (pix/s)	2
Optical Transmission	0.756
Atmospheric Transmission	1.0

Table 4.1: Lunar sensor specifications.

- $\gamma \in [0, 1]$: the discount factor over time, impacting the prioritization of short term rewards.

When a tasking action is considered, the probability of detecting the target object is computed and not assumed unity. Spherical cannonball models are assumed; simulated space objects are given uniformly sampled surface areas between 1 and 16 m^2 and uniformly sampled albedos between 0.2 and 0.4.

Two primary observers are applied in this study and are placed at the lunar north and south poles, respectively. Sensor designs consistent with a small slewable instrument are applied at each location, and a complete overview of the sensor schema is detailed in Table 4.1. Sensors are prescribed a maximal slew rate of 5 degrees per second. In prior studies, these sensor parameters were deemed sufficient for accessible observation throughout the Earth-Moon Halo families, with a minimum necessary photometric signal to noise ratio of 3 assumed for successful detection.

These sensors are then used to maintain estimates on a set of 100 objects following trajectories in Halo orbits about the Earth-Moon L1 and L2 Lagrange points. Orbits are split evenly between northern and southern Halos about each Lagrange point, and object populations are detailed in Table 4.2. For orbit selection, a maximal stability index of 10 is used as a threshold, and near-rectilinear orbits are prioritized. The resultant nominal catalog orbits range in period from approximately 6.5 to 12 days. True trajectories are perturbed from the nominal orbits, and stationkeeping procedures are applied about the assumed nominal. Consistent with the expected norms for future missions such as CAPSTONE or Artemis, downstream x-crossing control is applied for stationkeeping following the methods of Guzzetti et al. [48]. Maneuvers are permitted at nominal orbital phases between 135 and 225 degrees, and trajectories are considered over a lunar synodic period.

Orbit Type	Object Count
L1 Northern Halo	25
L1 Southern Halo	25
L2 Northern Halo	25
L2 Southern Halo	25
Total	**100**

Table 4.2: Periodic orbits performing stationkeeping maneuvers.

4.4.1 Uniform Observational Cadences

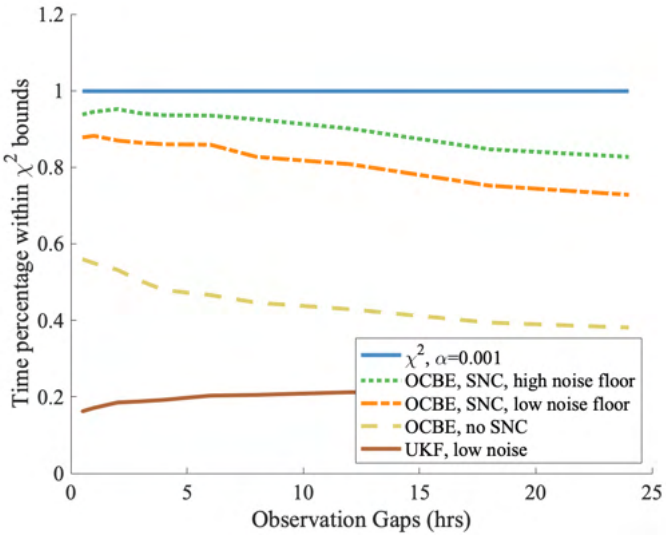


Figure 4.3: χ^2 rates across all catalog objects and epochs for studied filters and times between observation.

Object state estimates are first considered with observations tasked at a uniform rate. To maximize optical information, at each epoch, the positionally closest lunar observer for which a space object is accessible is chosen to observe. This tasking paradigm is applied in combination with the U-OCBE with adaptive state noise compensation, and each catalog object is studied independently. We first consider whether the applied estimator is successful in the cislunar regime. For comparison, results are compared with the U-OCBE without adaptive state noise compensation and with an unscented Kalman filter with very low process noise.

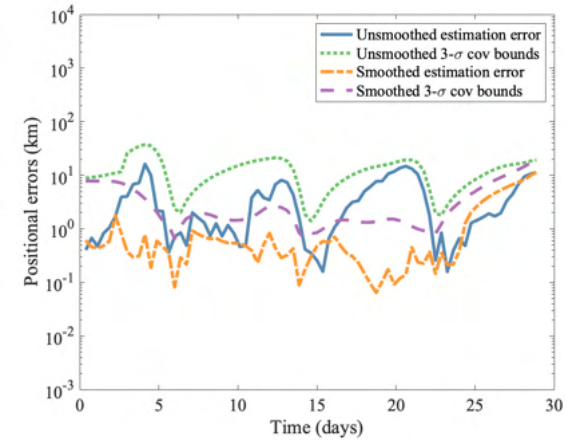
Throughout each simulation, we first consider estimation error with truth at each epoch. Comparing the state estimation error normalized by covariance with a chi squared distribution with degree of freedom 6, one may evaluate when covariance acceptably bounds estimation error. The time percentage over which

estimation error is acceptable is evaluated over every object as a means to provide a notional understanding of filter performance gains. These results are presented in Figure 4.3. Several features are noted from this Figure. Firstly, one may observe that filters very quickly diverge when little process noise is applied, and filter divergence is expected even with significant process noise if no care is taken around epochs when maneuvers occur. As such, the augmentation of the unscented OCBE with a adaptive state noise compensation leads to significant performance gains, and a consistent estimator is observed.

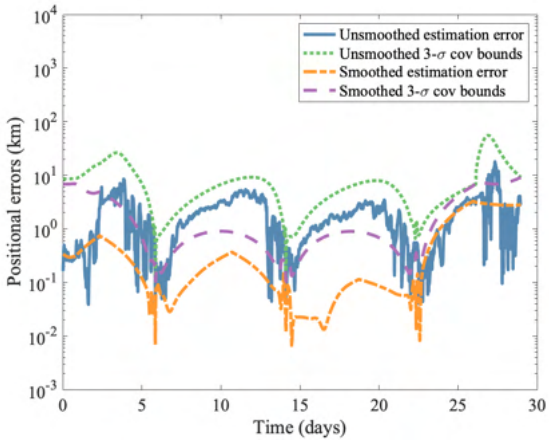
While the time spent with estimation error within covariance bounds is less than the expectation, this figure does incorporate the lag time around a maneuver where adaptive process noise is applied and the filter must re-converge on the true state. It is impossible to achieve a perfectly consistent filter without prior maneuver knowledge, so an exact chi-squared distribution is unrealistic. In the implementation presented, divergence from the covariance bounds must occur for adaptive methods to be triggered. Further impactful data features include maneuver magnitudes and time between a maneuver and the next observation. In combination with the added filtering challenges, this is visible in the decreased time spent within chi-squared bounds as observation gaps increase.

It is next useful to consider the structure of estimates visually over the course of a synodic period at a variety of cadences. In Figure 4.4, estimates over the same trajectory are presented at an observational cadence of 30 minutes and 9 hours. Major features are preserved between the two cases. For this analysis, it is first critical to note a large maneuver that is observed in velocity space in Figures 4.4c and 4.4d at approximately 2.5 days into the simulation. With a larger interval between observations, this maneuver effectively injects a large amount of positional uncertainty; as such, positional covariances in the 9 hour case are relatively large as compared to the high-frequency case, reaching approximately 50 km 3-σ, while the high-frequency case reaches a maximum of approximately 20 km. In the high-frequency case, one may note a large reduction in velocity uncertainties on a shorter timescale as compared to the 9 hour case.

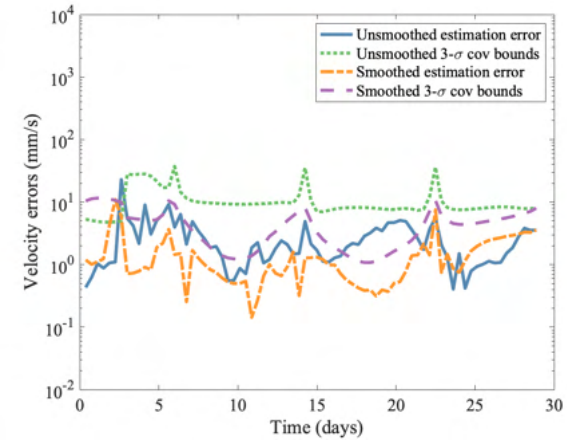
These results are also demonstrative of uncertainty bounds that one may expect for the catalog as a whole. In Figure 4.4, several sharp decreases in positional uncertainties are noted alongside increases in velocity uncertainty at approximately 7, 15, and 23 days into the simulation. This clear structure corresponds to close lunar approaches, at which time observations yield rich positional information and a variety of viewing



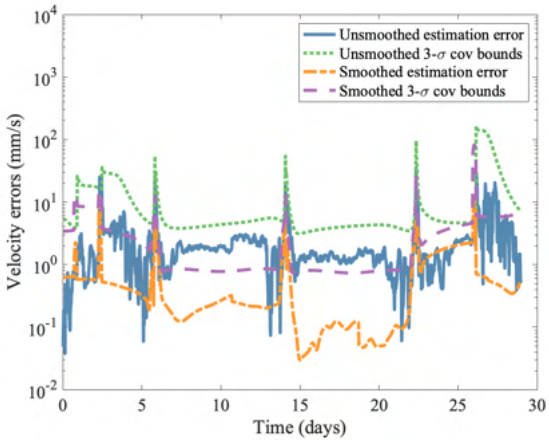
(a) Positional state estimates for a L1 Northern Halo with 9 hour observation gaps.



(b) Positional state estimates for a L1 Northern Halo with 30 minute observation gaps.



(c) Velocity state estimates for a L1 Northern Halo with 9 hour observation gaps.



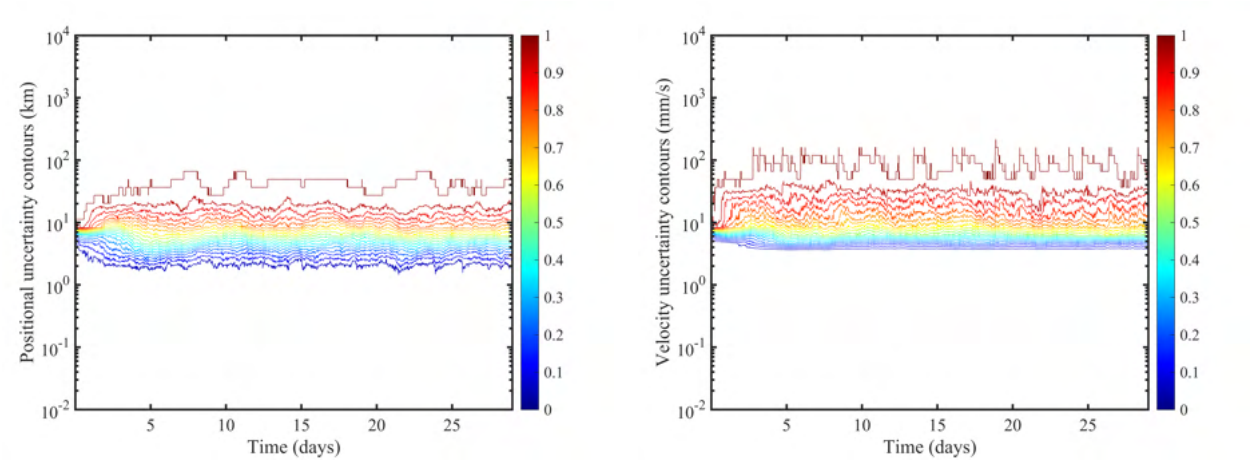
(d) Velocity state estimates for a L1 Northern Halo with 30 minute observation gaps.

Figure 4.4: OCBE estimates for a sample trajectory over a synodic period.

perspectives over a short time scale. On the other hand, positional information is at a minimum as orbits near apolune, a sensible expectation when only lunar observers are utilized. Finally, it is interesting to note the relative scale of covariance bounds between each case. Uncertainties on the same order of magnitude suggest that increasing observation frequencies offers a decreasing trade in new state information, though it is critical to note the benefit of shorter response time in the event of maneuvers. Further considering this point, it is useful to return to Figure 4.3, where filter degradation clearly occurs as measurement cadences exceed 8 hours. Further analysis could be applied to evaluate performance as a function of the angular arc between observations.

It is also important to note structures visible in the smoothed estimation error presented in Figure 4.4. The previously discussed correlation between orbital phase and estimation error is greatly reduced when considering this result, and the smoothed trajectories offer insight into the steady-state capabilities of the studied observers. Gains in smoothed error appear to be visible when comparing between the 30 minute cadence case and the 9 hour case, with smoothed positional errors of 0.42 kilometers on average in the 30 minute case, and 1.09 kilometers in the 9 hour case. The high-frequency case demonstrates that sub-kilometer smoothed covariance bounds below one kilometer are feasible, with smoothed estimation errors evolving on the order of 100 meters.

4.4.2 MCTS-Based Tasking

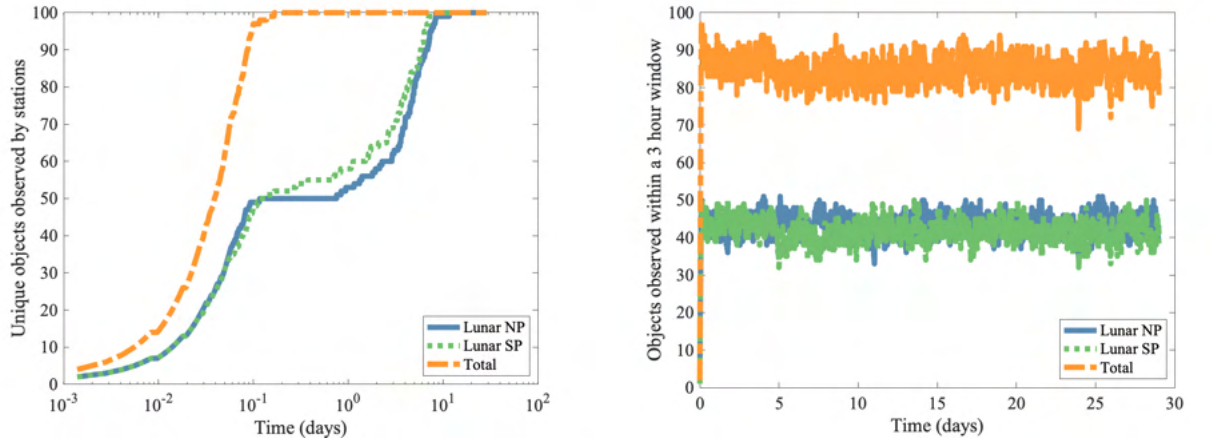


(a) Positional state estimates for a Halo catalog.

(b) Velocity state estimates for a Halo catalog.

Figure 4.5: Covariance bounds across the catalog with MCTS tasking.

Monte Carlo Tree Search is now applied to study the entire object catalog as a whole. The rollout heuristic presented in Section 4.3 is applied for each sensor, and at each epoch, tree search is evaluated over a depth d for an allocated period t_s. The diversity of observed objects over a lag period T_o is applied as a reward, and future rewards are discounted with a factor $\gamma \in (0, 1]$. In the presented results, parameters are selected as $d = 30$, $t_s = 15$ seconds, $T_0 = 30$ minutes, and $\gamma = 0.95$. After the allocated search period is elapsed, the immediate action with maximal simulated value is applied at the next observation step, allowing tree search to proceed. With an allocated search period that is less than the exposure time, this methodology



(a) Unique objects tasked over time by sensor.

(b) Unique objects tasked over a 3 hour period by sensor.

Figure 4.6: Sensor tasking information across the studied catalog.

is applied online.

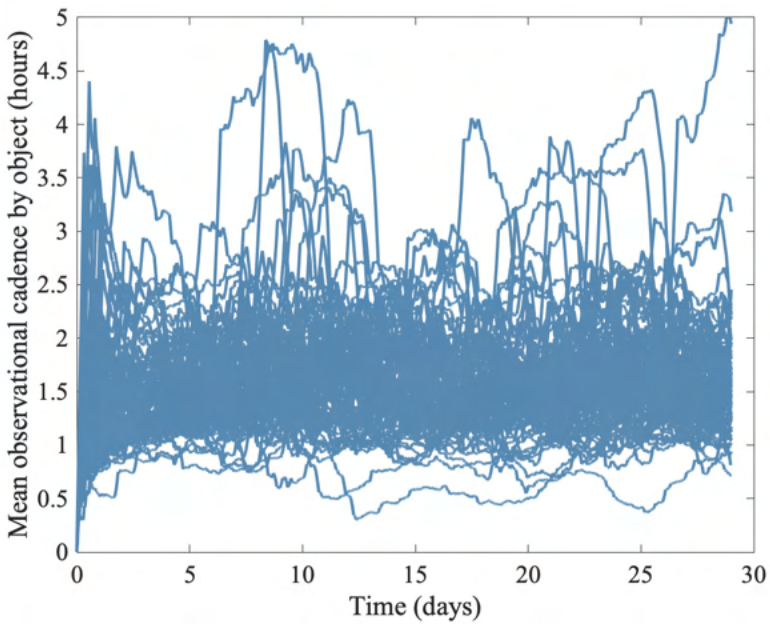
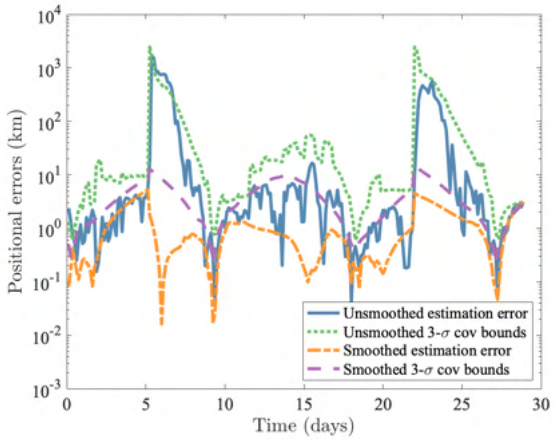


Figure 4.7: Lagging mean observational cadences for each catalog object.

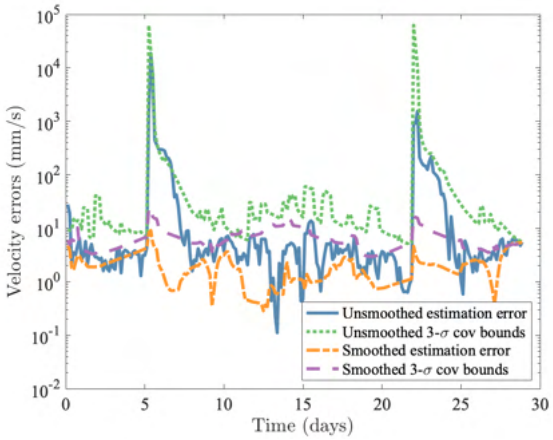
In the presented case, an observational cadence of 120 seconds is applied, corresponding to an averaged observational cadence of 6000 seconds for each object. Figure 4.5 visualizes covariance traces for objects across the catalog with MCTS tasking applied as a contour in position and velocity space. Levels of the

contour describe the percentage of the catalog below a threshold $3-\sigma$ uncertainty. Several features are noted in the Figure. Trajectories are largely maintained with median uncertainties on the order of 3 kilometers 3-σ in position and 5 millimeters/second in velocity. Positional uncertainties never exceed 100 kilometers, implying custody is maintained across the catalog, with the assumption that observations of objects after a maneuver are correctly associated. In Figure 4.5b, many spikes in velocity uncertainty are noted, but this behavior corresponds to application of adaptive process noise in the tasking loop.

It is also of interest to visualize what sensors observe each object, as well as how often objects are observed. Figure 4.6 presents this information, visualizing the diversity of objects each sensor observes over the full time history and a receding three hour window. We first may note that the entire catalog is observed after approximately 4 hours, demonstrating that the combination of observers placed at the lunar north and south poles offers excellent coverage of Halo families about L1 and L2. Additionally, each sensor effectively covers all objects by 8 days of simulation. Even though this is the case, Figure 4.6b demonstrates that each sensor maintains estimates on a distinct population of objects. Combined, the sensors on average tend to observe the vast majority of the catalog population, averaging detection of 90 of 100 objects over the receding horizon.



(a) Positional state estimates for a L2 Northern Halo applying large maneuvers.



(b) Velocity state estimates for a L2 Northern Halo applying large maneuvers.

Figure 4.8: Successful state estimation for an agile spacecraft following a L2 Halo trajectory.

Observational cadences may be further considered with respect to specific objects. Figure 4.7 visualizes

mean cadences across the catalog, and we note a concentration about approximately one hour on average. Note that the mean cadences are taken as the average time since last observation over a receding period of 3 days. Several factors may lead to observation gaps. Depending on solar phase angle, periods of challenged observation of certain space objects are expected. Alternatively, during periods when uncertainties are expected not to grow significantly and maneuvers have less significant effects, objects are deprioritized. Interestingly, a large portion of the objects with lower than average observational cadences followed short period trajectories about L2. This set had high CGT indicators on average, and this behavior demonstrates the capability of MCTS to stochastically prioritize objects for which maneuvers are more impactful.

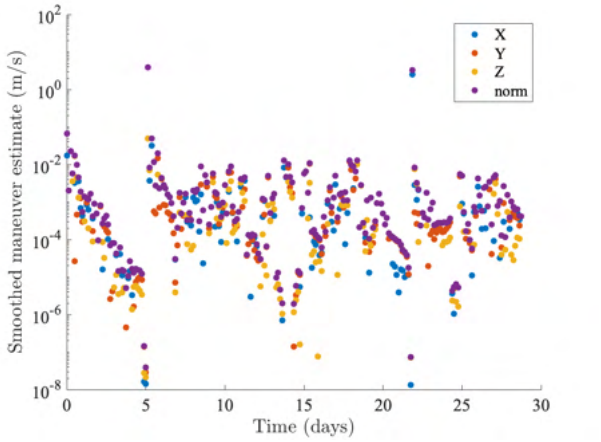
The test case validates MCTS as a means to track large populations of objects in challenging scenarios, as well as lunar observer capabilities for management of growing populations in cislunar space.

4.4.3 Robustness to Large Maneuvers

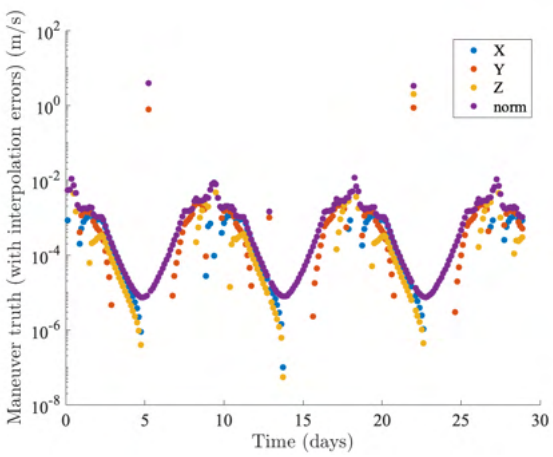
Finally, we consider whether the developed methodology remains convergent in a challenging scenario. In the case presented in Figure 4.8, the space object in question makes large maneuvers at approximately 5 and 22 days into the simulation. These maneuvers are 3.94 and 3.32 meters per second in magnitude, respectively. The object is observed every 4 hours. It first may be noted that the filter is consistent throughout the simulation.

While each maneuver is accompanied by a large spike in uncertainty, covariance bounds decrease relatively quickly after maneuvers occur, especially in position space. Positional covariance traces generally decrease by an order of magnitude on a time scale of hours, despite the fact that these maneuvers coincide with periods where the object is distant from the observers and positional information is comparatively sparse. Smoothed estimation error is even more successful, on the order of a kilometer in position and 5 millimeters per second in velocity. This motivates an avenue of future work, in which the OCBE smoother may be leveraged over a lagging horizon to better inform adaptive process noise selection. If the smoother is used to provide a better understanding of the maneuver that occurred, filter re-instantiation with increased maneuver knowledge may yield further performance gains.

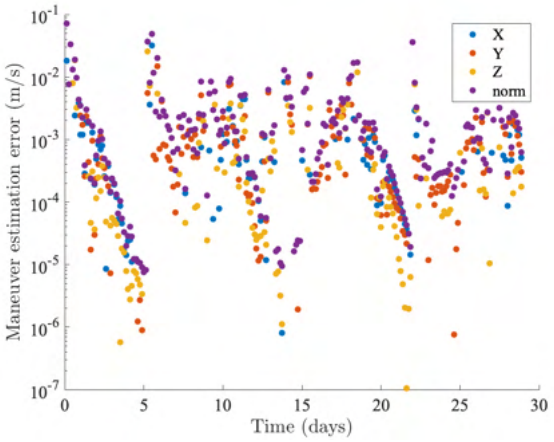
It is also worthwhile to compare estimated maneuvers with the true cataloged maneuvers. Figure



(a) Estimated maneuvers across simulation of an agile object.



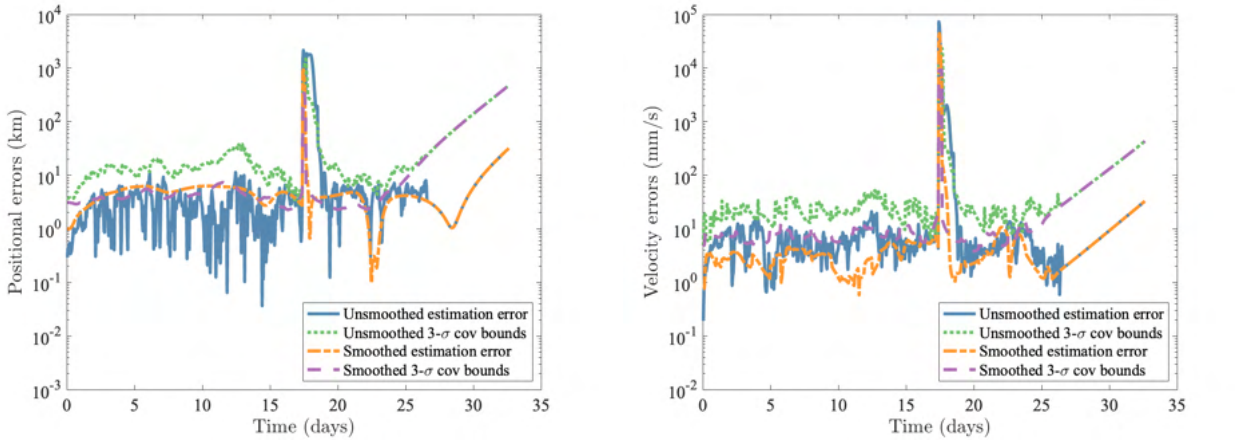
(b) Maneuver truth across object simulation.



(c) Maneuver estimation errors across object simulation.

Figure 4.9: Maneuver estimation and analysis for a stationkeeping space object on an unstable Halo orbit.

4.9 visualizes first order maneuver estimates obtained with the unscented OCBE smoother at each epoch compared with truth. Note a slight oscillatory structure in the true maneuver information, a result of integration and truncation differences within simulation compared to generation of the truth trajectory. Interestingly, the smoother has some success in estimating the magnitudes of the resultant small maneuvers when the filter is relatively certain of the true state, with less success in the aftermath of the large maneuvers. Estimates of large maneuvers in the example are quite accurate, with estimation error on the order of 10 millimeters per second on maneuvers several meters in magnitude. Effectively, this result demonstrates that the sensors utilized are unable to detect impulsive maneuvers on the order of 10 millimeters per second over



(a) Positional state uncertainties across a L1 NHalo to L2 NHalo transfer trajectory.

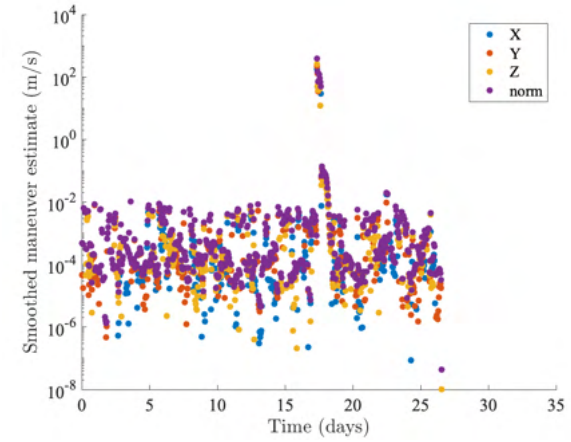
(b) Velocity state uncertainties across a L1 NHalo to L2 NHalo transfer trajectory.

Figure 4.10: Successful state estimation following a challenging transfer trajectory.

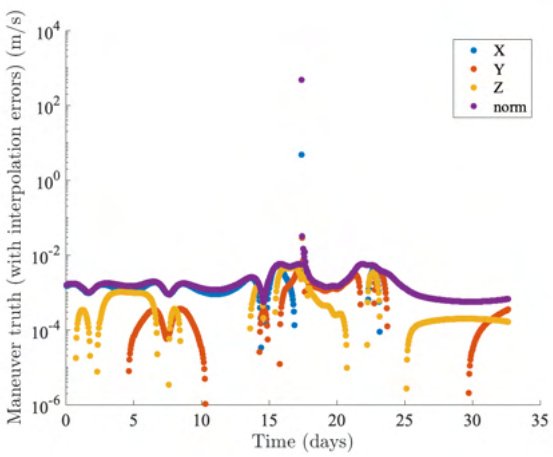
the observation window considered.

We may also consider an example in which maneuvers are much larger. In Figure 4.10, we visualize the utilization of lunar surface observers to maintain state estimates over a transfer from a L1 Halo to a L2 Halo orbit. The SO in question first departs along the unstable manifold of the initial orbit, then performs a single impulsive maneuver on the order of 400 meters per second occurs approximately 17 days into simulation. Following the maneuver, the SO then coasts along the stable manifold of the target orbit until the end of simulation. The estimator struggles to maintain a convergent state estimate immediately after the maneuver but estimation error quickly returns to the covariance bounds. Additional growth in uncertainty is observed at the end of the study and occurs because the object is no longer observable as a result of distance and solar phase angle.

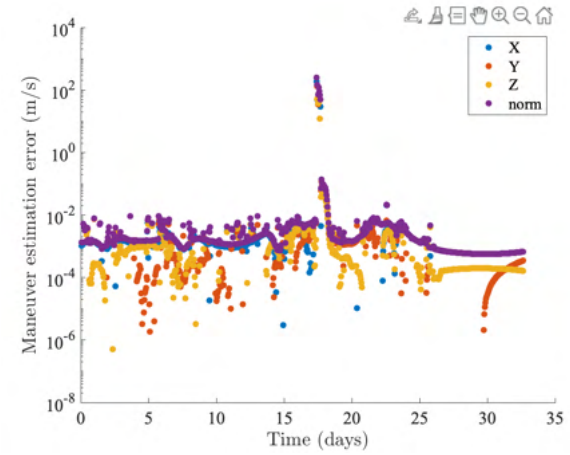
Because of challenges in state estimation around the maneuver epoch, the estimated maneuver in this case (visualized in Figure 4.11) is also quite poor. Interestingly, the unscented smoother rather estimates a smaller, decaying maneuver estimate over several hours after the large maneuver occurs. This is likely a result of the optimal control assumptions in the underlying filter, and a large, impulsive burn in this case is clearly sub-optimal. Nevertheless, this presents a challenge with the methodology used in that a non-cooperative object could maneuver sub-optimally to challenge the tracking methods used.



(a) Estimated maneuvers across simulation of a transfer trajectory.



(b) True maneuvers for a transfer trajectory.



(c) Maneuver estimation errors across a transfer trajectory.

Figure 4.11: Maneuver estimation and analysis using the U-OCBE smoother.

4.5 Discussion and Conclusions

This contribution presents a variety of methodologies for maintaining tracks in the near-lunar environment. The Optimal Control-based Estimator is augmented with novel smoothing and modified adaptive process noise algorithms, allowing for effective state estimation in highly nonlinear problems. Using the Cauchy-Green stress tensor, dynamical knowledge is leveraged to inform maneuver feasibility. The methods are then combined into a rollout heuristic for integration with Monte Carlo Tree Search-based tasking.

We present a variety of results with application to the usage of lunar ground-based optical observers.

A broad study of observation at uniform cadences is performed, demonstrating the utility of the presented filter schema. Insight is given into necessary maximal observation gaps for successful state estimation. Monte Carlo Tree Search is then applied, and the methodology is demonstrated to be robust to highly agile space objects. The sensor configuration demonstrates the capability to maintain estimates over large catalogs of objects, supporting an increasing number of missions in the cislunar regime.

Chapter 5

Decentralized Decision Making using Monte Carlo Tree Search

Another emerging need for SDA sensor tasking is the question of how to best coordinate many sensors for a common goal. The methods presented insofar in this thesis consider the notion of a central decision-maker that passes tasking actions to each agent in a problem, but there are significant challenges in scalability of such a methodology to a many-agent problem. As observers become oversubscribed, it is inevitable that a large number agents must be considered. The rollout policies previously discussed are generally linear in the number of observers, but this still has potential to be a bottleneck when efficient search tree throughout is critical to the utility of MCTS. In addition, the practicality of passing decisions at high frequency in such an architecture is also of concern. To succesfully apply MCTS in a decentralized manner, the SDA sensor tasking problem must be revisited using the Markov game framework, discussed in Section 2.1.

The goals of this chapter are twofold. First, a means for ensuring decentralized communication between agents is developed using random graphs, with specific focus on scalability in many-agent problems. The use of random graphs allows bounds to be established on communication times, connectivity, and robustness of communication between a large number of agents. The MCTS methodology is then modified to consider the effect of communication on rewards, and analysis is performed on the resultant algorithm, which may be applied to many-agent Markov games. The decentralized MCTS methodology is then studied for geostationary and cislunar catalog maintenance scenarios. We first present results for a scenario in which the two telescopes at the VADeR observatory are coordinated in a decentralized manner, sucessfully maintaining geostationary state estimates. We then consider the utility of sensors placed in Lyapunov and Halo orbits as in Chapter 3 for maintaining state estimates on a catalog of Halo orbits. Two studies are

presented for this case; in the first, the two space-based sensors operate in a decentralized manner alongside a large Earth ground-based sensor. In the second scenario, the ground-based sensor is isolated and unable to communicate with any other agents for a subset of the simulation, demonstrating robustness of the methodology to communication failures.

5.1 Communication between Agents over Random Graphs

As the tree search process evolves, agents must intermittently communicate the actions they desire to take with other agents. Such actions may impact the local reward of another agent, with

$$V^i_{\pi^i,\pi^{-i}}(s) = E\left[\sum_{t=0}^{\infty}\gamma^t R^i(\mathbf{s}_t,\mathbf{a},\mathbf{s}_{t+1})\,\middle|\, a^i_t \sim \pi^i(\cdot|s_t), s_0 = s\right] \tag{5.1}$$

Becuase of this effect, new information on the action another agent might take discretely reduces the local value of that action, a scenario we term a "breakpoint." Let each agent select and communicate with another random agent every τ intervals, while maintaining lines of communication with any prior agents. Typically, in random graph literature, the number of vertices is labelled n and outdegree is labelled k. To relate such concepts to multi-agent decision processes and this communication scheme, the number of vertices is defined as the number of agents K, and the outdegree shall be defined as the number of communication rounds τ.

The problem is greatly simplified if it is assumed agents communicate in an undirected manner, and it is worth briefly discussing such a scenario. Wormald demonstrates that any τ-regular graph is τ-connected for $\tau > 3$ with high probability [114]. Connectivity ensures that any agent may communicate with any other agent over the resultant graph.

Furthermore, it is imperative to consider the diameter of the resultant graph, in order to determine a bound for the time it takes agents to communicate with each other. Bollobas and De La Vega present asymptotic bounds [15] for the diameter d of τ-regular random graphs as the minimum integer satisfying

$$(\tau-1)^{d-1} \geq (2+\epsilon)\tau K \log K \tag{5.2}$$

and

$$d \approx 1 + \log_{\tau-1}(2\tau K \log K).$$

These bounds may then be used to ensure communication and constrain communication times between agents. While these bounds have been well established for regular graphs, the directed case has not been studied in detail, and is much more realistic for the context of this problem. After τ rounds of communication, we assume a τ-regular digraph is generated, such that each agent has τ out-neighbors with which it communicates. Frieze and Karonski provide a discussion on the connectivity of τ-out directed graphs, and demonstrate the probablity such a graph is not connected devolves to the event a vertex has indegree 0 with high probability [37]. The probability of this event may be evaluated as

$$\mathcal{P}(\mathcal{E}_0) = \left(\frac{\binom{K-2}{\tau}}{\binom{K-1}{\tau}}\right)^{K-1} = \left(1 - \frac{\tau}{K-1}\right)^{K-1} \approx e^{-\tau} \tag{5.3}$$

with an expected number of vertices with total degree τ over K vertices of

$$E[Z(\tau)] = Ke^{-\tau}$$

The probability the directed graph has edge connectivity of at least one may be evaluated as

$$\mathcal{P}(\lambda(D) \geq 1) = (1 - \mathcal{P}(\mathcal{E}_0))^K = (1 - e^{-\tau})^K \approx 1 - Ke^{-\tau} \tag{5.4}$$

This result is visualized in Figure 5.1. Applying a threshold probability of connectivity Δ, a necessary out-degree may be evaluated as

$$\tau = -\log(1 - \Delta^{\frac{1}{K}}) \tag{5.5}$$

When τ exceeds this threshold, we assume the random graph is strongly connected with high probability. This analysis may be extended to further consider the probability the random graph is strongly r-connected and robust to the denial of r edges. In general, this robustness may be reduced to the probability that every vertex has indegree of at least r. Here, we have

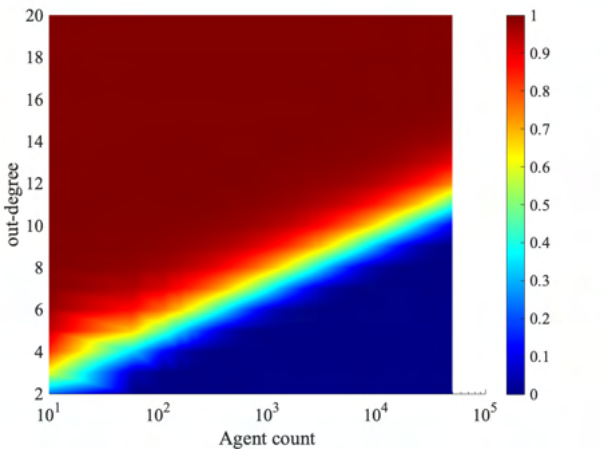


Figure 5.1: Probability of connectivity for many agents and small out-degree.

$$\mathcal{P}(\lambda(D) \geq r) = \left(1 - \sum_{i=0}^{r-1} \mathcal{P}(\mathcal{E}_i)\right)^K \tag{5.6}$$

Within this result, in-degree probabilities may be evaluated as

$$\begin{aligned} \mathcal{P}(\mathcal{E}_i) &= \left(\frac{\binom{K-2}{\tau}}{\binom{K-1}{\tau}}\right)^{K-1-i} \left(1 - \frac{\binom{K-2}{\tau}}{\binom{K-1}{\tau}}\right)^i \binom{K-1}{K-1-i} \\ &= \left(1 - \frac{\tau}{K-1}\right)^{K-1-i} \left(\frac{\tau}{K-1}\right)^i \binom{K-1}{K-1-i}. \end{aligned} \tag{5.7}$$

A bound on necessary out-degree is non-trivial, but can be numerically evaluated. In Figure 5.2, we evaluate this necessary out-degree to ensure a random digraph is r-connected across a wide range of vertices.

Also interesting to this problem is consideration of the diameter of the resultant graph. If each agent forwards all information it has to downstream agents, the diameter of the random graph gives insight into the maximal communication time between arbitrary agents. This result has been well-studied by Addario-Berry et al, and a specific bound on the diameter of the largest connected component is illustrated, with [1]

$$d(D(K,\tau)) = (1 + \eta_\tau + o(1)) \log_\tau K \tag{5.8}$$

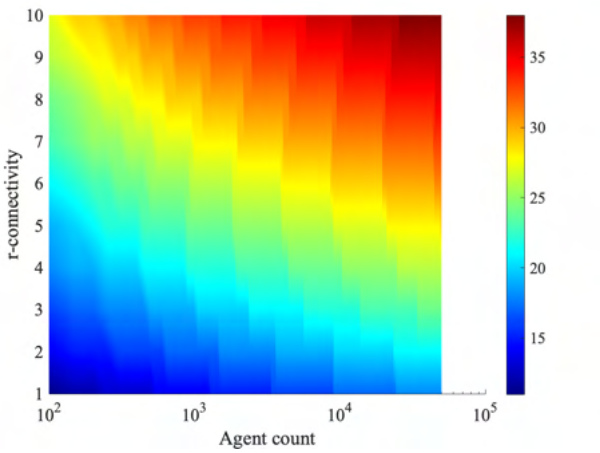


Figure 5.2: Out-degree requirements for 0.999 probability of r-connectivity.

where

$$\eta_\tau = \frac{\log \tau}{\lambda_\tau \tau - \log \tau} \tag{5.9}$$

$$\lambda_\tau = \max 1 - \lambda - e^{-\tau\lambda} \tag{5.10}$$

Note that η_τ is a function of the out-degree, not the number of agents. Interestingly, the diameter of the graph concentrates much more quickly than the connectivity, and the key challenge in the directed sense is ensuring all vertices within the graph are connected. A visualization of the diameter of the largest connected component is given in Figure 5.3.

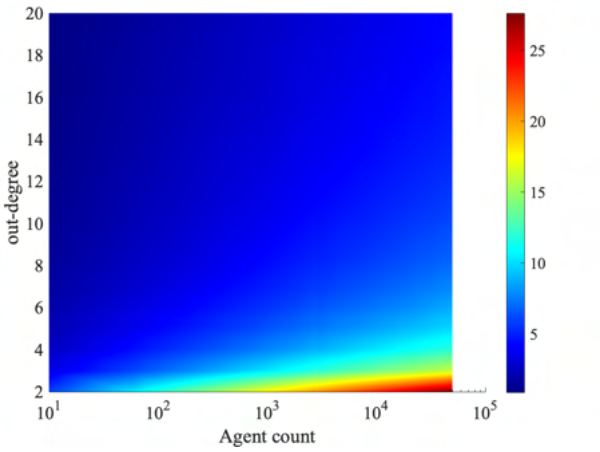


Figure 5.3: Expected diameter of a random digraph.

While theoretic probabilities and expectations are given throughout this section, it is also critical to note that these results have been compared to Monte Carlo trials for confirmation. It is also useful to compare these results to other communication architectures. Figure 5.4 visualizes the relative communication rates for a random graph architecture to an ideal case in which random communication is two-way and an all to all case.

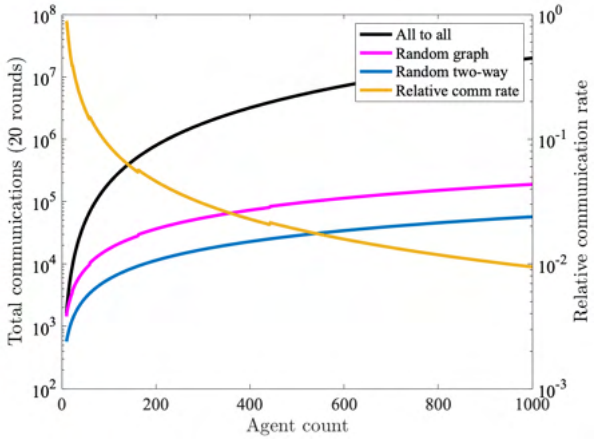


Figure 5.4: Relative rates for various communication architectures.

With communication behaviors established, it is now critical to consider the impact of communication on search convergence toward local optimality. Let the action trajectory with maximal D-UCB be sampled from each agent's search tree every τ search iterations. On sampling a new action trajectory, each agent may then communicate the actions it shall take and obtained information on other agents to all neighbors in the random graph. This behavior may lead to breakpoints, and communication rates may be determined to ensure agents are connected and reachable quickly as compared to total time allocated for search.

5.2 Decentralized Monte Carlo Tree Search

This section presents a decentralized Monte Carlo Tree Search solution modified from the approaches of Best et al.. A tree-based extension to the selection policy first presented in Garivier and Moulines is applied. Following the classic proof by Kocsis and Szepesvari, we demonstrate that the selection policy concentrates about the optimal arm for a bandit with switching payoffs following a power law. We then extend this

derivation, demonstrating concentration about the optimal arm for a search tree. Given a discount γ and t pulls, the discounted upper confidence bounds (D-UCB) policy chooses the arm that maximizes [41]

$$I_t = \underset{i \in \{1,\ldots,K\}}{\operatorname{argmax}} \{\bar{R}_{i,t}(\gamma) + c_{t_0,t_i}(\gamma)\} \tag{5.11}$$

where $\bar{R}_{i,t}$ describes a time-weighted average return and c_{t_0,t_i} describes the bias sequence for arm i. This term admits further exploration of actions estimated to be suboptimal and may be related to the typical Upper Confidence Bound [5], but the D-UCB policy further prioritizes returns from more recent episodes. To do so, nodes in the search tree are augmented to record when they are visited during search, and visit times may be utilized to express the discounted visits to a node as

$$t_j = \sum_{u=1}^{t} \gamma^{t-u} \mathbf{1}_{i_u=j}. \tag{5.12}$$

For the purpose of exploration, the discounted visits to the parent node t_0 may then be expressed as the sum of discounted visits to all child nodes,

$$t_0 = \sum_{j \in C(h)} t_j. \tag{5.13}$$

Finally, the time-weighted average return may be considered by similarly weighting recent returns, as

$$\bar{R}_{j,t}(\gamma) = \frac{1}{t_j(\gamma)} \sum_{u=1}^{t} \gamma^{t-u} R_u \mathbf{1}_{i_u=j}. \tag{5.14}$$

Returns at any time are assumed bounded as $R_u \in [-1, 1]$. With these components, the exploration term c_{t,t_i} is expressed as

$$c_{t_0,t_i} = C_p \sqrt{\frac{\log t_0(\gamma)}{t_j(\gamma)}}. \tag{5.15}$$

This form is comparable to the logarithmic bonus utilized in traditional MCTS, and future consideration could be given to the more recently studied polynomial bonus term. Several additional assumptions are

taken from Best et al. regarding switching behavior on arms, limits on expected payoff, and drift behavior of arms after switching. First, for a given arm, we may consider drift behavior, and the fact that for a given arm, the payoff distribution may change.

Assumption 1. *Fix $1 \leq i \leq K$. Let $\{\mathcal{F}_{it}\}_t$ be a filtration such that the reward process $\{R_{it}\}$ is $\{\mathcal{F}_{i,t}\}$-adapted and $R_{i,t}$ is conditionally independent of $\mathcal{F}_{i,t+1}$, $\mathcal{F}_{i,t+1}$, ... given $\mathcal{F}_{i,t-1}$. There exists an integer T_p such that for $t_i \geq T_p$ and $t < t_i$, $R_{i,t}$ is independent from $\{\mathcal{F}_{i,t}\}$ [11].*

Further context for this assumption is provided by considering how changes in the payoff distribution may occur. As agents learn online, new action trajectories may become locally optimal. We assume that agents then communicate newly optimal sequences. One agent changing behaviors may result in a sudden shift in the local optima for another agent, henceforth termed a breakpoint. Allowing an agent to communicate with other random agents every interval τ, breakpoint occurrence rates may be upper bounded in search iterations t as $\frac{t}{\tau}$. We shall further assume that the occurrence rate of breakpoints decreases over time as agents build a more detailed search tree.

Assumption 2. *Breakpoints occur as a power law in t following $\Upsilon_t = O(t^{\beta})$, where $\beta \in [0, 1)$. As such, the rate of new breakpoints trends to 0 with $\lim_{t\to\infty} \dot{\Upsilon}_t = O(t^{\beta-1}) = 0$.*

Following this assumption, the expected reward at each arm may be expected to converge to the true payoff, and the limit exists $\mu_i \lim_{t\to\infty} \forall i \in \{1, ..., K\}$. Finally, Best et al. make assumptions on the drift behavior of arms. Let drift be defined as the difference between the expected reward at time t and the limit $\delta_{i,t} = \mu_{i,t} - \mu_i$. The optimal expected payoff takes the maximal $\mu_{i,t}$. We assume that at some index $T_0(\epsilon)$, the drift becomes proportional to $\Delta_{i,t}$, the minimum difference between expected reward for the optimal arm i_u^* and the expected reward for arm i, where

$$\Delta_{i,t} = \min_{u\in\{1,...,t\}} \{\mu_{i_u^*,u} - \mu_{i,u} : \ i \neq i_u^*\}. \tag{5.16}$$

Assumption 3. *There exists a time $T_0(\epsilon)$ where, for $\epsilon > 0$ and $M_i(t) \geq T_0(\epsilon)$, $|\delta_{i,t}| \leq \epsilon\frac{\Delta_{i,t}}{2}$ and $|\delta_t^*| \leq \epsilon\frac{\Delta_{i,t}}{2}$ $\forall i$ [11].*

We now must consider the impact of breakpoints on the expected number of pulls of a suboptimal arm. Supposing the best action trajectory changes after each set of communication rounds in the worse case, the number of breakpoints is monotonically increasing and linear in search iterations at worst. This behavior would challenge any asymptotic analysis, and is resolved by Best by enforcing the number of communication intervals as a function of search iterations t. This effectively results in a cooling of resampling action trajectories from the search tree, effectively reducing application of the generated trees. Garivier and Moulines discuss bounds on selection of suboptimal arms with different breakpoint rates, including power laws as a function of selections. An upper bound in regret is applied as [41]

$$E_\gamma[\tilde{N}_t(i)] = O(\sqrt{\Upsilon_t t}\log t) \tag{5.17}$$

where Υ_t is the number of breakpoints up to time t, and i is the suboptimal arm. If $\Upsilon_t \approx O(t^\beta)$, following discussion from Garivier and Moulines and Best et al,

$$E_\gamma[\tilde{N}_t(i)] = O(t^{\frac{1+\beta}{2}}(C_p^2 \log t + T_0(\epsilon) + T_p) \tag{5.18}$$

Note that C_p is derived from the bias sequence c_{t,t_i}. The discount γ_t is carefully selected in the D-UCB formulation [41], with

$$\gamma_t = 1 - \frac{t^{\frac{\beta-1}{2}}}{4} \in [0.75, 1] \tag{5.19}$$

This result may then be utilized to consider payoff convergence toward the optimal payoff.

Lemma 3. *(Expected payoff convergence). Let*

$$\bar{R}_t = \sum_{i=1}^{|\mathcal{A}|} \frac{T_i((t)}{t}\bar{R}_{i,t} \tag{5.20}$$

With prior assumptions,

$$|E_\gamma[\bar{R}_t] - \mu^*| \leq |\delta_t^*| + O(|\mathcal{A}|t^{\frac{\beta-1}{2}}(C_p^2 \log t + T_0(\epsilon) + T_p) \tag{5.21}$$

This result is modified from Best Lemma 2 [11] for monotonically increasing breakpoints. A similar result is reached, and further analysis is provided in Kocsis et al, Theorem 3 [64].

One must next consider whether payoffs concentrate about the expected payoff. This result is given as follows, and corrects errors presented in the work of Best et al..

Lemma 4. *(Payoff concentration about the expected payoff). For* $\epsilon \in (0, 1]$*, let*

$$\Gamma_t = 9\sqrt{2\ln(2/\epsilon)}t^{\phi}$$

and

$$\frac{1+\beta}{2} < \phi < 1$$

Then, for $t \geq O(|\mathcal{A}|t^{\frac{\beta-1}{2}}\log t)$*, the following bound holds.*

$$\mathcal{P}(t|\bar{R}_t - E_\gamma[\bar{R}_t]| \geq \Gamma_t) \leq \epsilon \tag{5.22}$$

Proof: Two components are needed to apply this bound; first, an upper bound on the expected number of times a suboptimal arm is selected (Kocsis Lemma 13) [64] is required. This bound is expressed in Equation 13, and can be compared to the bound found by Kocsis of $O(\log(n))$. Second, a term related to regret is required, explicitly describing the expected total loss by pulling suboptimal arm Y rather than the optimal arm as

$$\begin{aligned} E[\text{reg}_t] &= E[\sum_{i=1}^{t} R_i] - E[S_t] \\ &= E[\sum_{i=1}^{t} R_i] - E[\sum_{i=1}^{t}(1-Z_i)R_i + Z_iY_i] \\ &= E[\sum_{i=1}^{t} Z_i(R_i - Y_i)] \leq E[2\sum_{i}^{t} Z_i] \end{aligned} \tag{5.23}$$

Note that the term S_t is comparable to the true payoff $t\bar{R}_t$, if an arbitrary arm were selected rather than a single suboptimal arm. Therefore, the expectation of regret is bounded by the number of pulls of a

suboptimal arm. Note Z_i is an indicator variable describing when the suboptimal arm is pulled and Y_i is the return for pulling arm Y at time i. Following Lemma 14 of Kocsis et al. [64], there exists some t_0 such that

$$a_t \leq \frac{\Gamma_t}{9}, \quad |R_t| \leq \frac{2\Gamma_t}{9} \tag{5.24}$$

We may bound $p = \mathcal{P}(S_t \geq E[S_t] + \Gamma_t)$ to ensure regret concentrates as

$$\begin{aligned} p \leq & \mathcal{P}\left(\sum_i^t R_i \geq E[\sum_i^t R_i] + \frac{\Gamma_t}{9}\right) \\ & + \mathcal{P}\left(2\sum_i^t Z_i \geq \frac{8\Gamma_t}{9} - R_t\right) \end{aligned} \tag{5.25}$$

Applying the Hoeffding-Azuma inequality to the first term, we have

$$\begin{aligned} \mathcal{P}\left(\sum_i^t R_i \geq E[\sum_i^t R_i] + \frac{\Gamma_t}{9}\right) &\leq \exp\left(-\frac{(\frac{\Gamma_t}{9})^2}{2t}\right) \\ &\leq \frac{\epsilon}{2}\exp\left(t^{1-2\phi}\right) \leq \frac{\epsilon}{2} \end{aligned} \tag{5.26}$$

More care must be taken for the second term. With the assumption on regret, we have

$$\begin{aligned} \mathcal{P}\left(2\sum_i^t Z_i \geq \frac{8\Gamma_t}{9} - R_t\right) &\leq \mathcal{P}\left(2\sum_i^t Z_i \geq \frac{4\Gamma_t}{9}\right) \\ &= \mathcal{P}\left(\sum_i^t Z_i \geq \frac{2\Gamma_t}{9}\right) \end{aligned} \tag{5.27}$$

By Kocsis Lemma 13,

$$\begin{aligned} \mathcal{P}\left(\sum_i^t Z_i \geq \frac{2\Gamma_t}{9}\right) &\leq -\exp\left(\frac{(\frac{2\Gamma_t}{9})^2}{8t}\right) \\ &\leq \frac{\epsilon}{2}\exp\left(t^{1-2\phi}\right) \leq \frac{\epsilon}{2}. \end{aligned} \tag{5.28}$$

Collecting terms, the total probability $p \leq \epsilon$. Note that these probabilities may be restricted even further. The looser bound is a function of the somewhat larger Γ_t as compared to Kocsis et al., which is

a result of the larger regret term and expected upper bounds on sub-optimal pulls. Nevertheless, because regret concentrates about the expectation, we guarantee polynomial concentration of the optimal return about the expectation on the order $O(t^{\phi-1})$, where $\frac{1+\beta}{2} < \phi < 1$.

Finally, we consider the claim that failure probability converges to 0.

Lemma 5. *(Convergence of failure probability). Under the given assumptions it holds that*

$$\lim_{t\to\infty} \mathcal{P}_\gamma(I_t \neq i_t^* = i^*) = 0 \tag{5.29}$$

Kocsis and Szepesvari demonstrate this result to arbitrarily small ϵ, but note that there are challenges surrounding weak bounds for the concentration of regret about the expectation [64]. Shah et al. present rich discussion on this subject alongside analysis of a polynomial form [90], and similar analysis was performed by Auger et al. [6]. However, these publications have not been peer reviewed, and the lack of trustworthy analysis of the underlying MCTS methodology remains a significant gap in the literature. We defer to such discussions and conclude by extending the analysis of a switching MAB to search trees. This problem may be considered an extension of the original scenario in which a series of bandits with hierarchical relationships may be considered. Further introduction is provided by Best et al.

Theorem 1. *(Bias convergence for trees). Consider D-UCB applied to tree $\mathcal{T}$ with depth d and branching factor $|\mathcal{A}|$ The reward distributions of the leaf nodes are i.i.d and nonstationary. Breakpoints follow power law assumptions, and a discount factor is applied as previously outlined. With $M_{i_0}(t) \geq T_p$ and $M_{i_0}(t) \geq T_0$, the bias in the optimal action-value estimate at the root node may be expressed as*

$$|\bar{F}_{i_0,t_{i_0}} - \mu_{i_0}^*| = O(|\mathcal{A}|Dt^{\frac{\beta-1}{2}}\log(t)) \tag{5.30}$$

and given infinite time failure probability asymptotically trends to zero.

The proof of this theorem is equivalent to that presented by Best et al. [11], and the key result of this section is the extension of analysis to less restrictive breakpoint sequences.

5.3 Implementation

Insofar, methods for communication between agents and analyses of the underlying MCTS methods have been outlined. It remains to unite these methods and consider how they may be applied to a Markov game. A brief visualization is given in Algorithm 3. Locally, an agent alternates between generating a search tree with Dec-MCTS and communicating results of that search tree to other random agents. The agent communicates its current best action trajectory, and in addition, it also propagates any other actions sets it has received from other agents through the graph. With timestamped trajectories, each agent then has downstream access to recent likely actions for all other agents with the assumption of a fully connected graph.

During simulation, an agent applies the actions specified for other agents to inform expected rewards; if no action trajectory is available for another agent, an action is randomly sampled. Progressive widening [25] is applied for large action spaces to simulate new actions within the search tree. A random rollout is applied if a new action is desired, and domain knowledge may be applied to generate new actions. Simulation proceeds until a time or computation budget is exhausted, at which point the D-UCB maximizing action is applied.

Algorithm 3 Dec-MCTS over random graphs.

Require: belief b, objective g, t_{max}, τ, d
Ensure: optimal action for local agent i
  $\mathcal{T} \leftarrow$ initialize D-UCT tree
  $t \leftarrow 0$
  $a_{(i)} \leftarrow []$ ▷ Received action trajectories for other agents
  $a_i \leftarrow []$ ▷ Optimal action trajectory
  $comm \leftarrow []$ ▷ outward edges for communication
  while $t < t_{max}$ **do**
    for τ iterations **do**
      $\mathcal{T} \leftarrow \text{simulate}(\mathcal{T}, b, d, a_{(i)}, g, t)$
      $t \leftarrow t + 1$
    $a_i \leftarrow I_t(\mathcal{T})$ ▷ D-UCB applied recursively
    if increased outdegree needed **then**
      $comm \leftarrow [comm, rand([1, ..., K] \setminus (i \cup comm))]$
    transmit$(comm, a_i \cup a_{(i)})$
    return a_i

Algorithm 4 Dec-MCTS Simulation Loop.

function SIMULATE($\mathcal{T}, b, d, a_{(i)}, g, t$)
  if $d = 0$ **then return** $\mathcal{T}, 0$
  $a \leftarrow$WIDEN($\mathcal{T}$)
  $a' \leftarrow$DRAW($a_{(i)}, |\mathcal{A}|$) ▷ Take received or random action for other agents
  $b' \leftarrow T(b, a \cup a')$
  $r \leftarrow g(b, a \cup a', b') - g(b, a', T(b, a'))$
  if $a \notin C(\mathcal{T})$ **then**
    $C(\mathcal{T}) \leftarrow \{a, b', r\}$
    $\mathcal{T}a, r_\gamma \leftarrow$ROLLOUT($\mathcal{T}a, b', d-1, a_{(i)}, g, t$)
    $R(t) \leftarrow r + \gamma r_\gamma$
  else
    $\mathcal{T}a, r_\gamma \leftarrow$SIMULATE($\mathcal{T}a, b', d-1, a_{(i)}, g, t$)
    $R(t) \leftarrow r + \gamma r_\gamma$
  $N(\mathcal{T}) \leftarrow N(\mathcal{T}) + 1$
  visits($\mathcal{T}$) $\leftarrow$visits($\mathcal{T}$) $\cup\, t$
  $Q(a) \leftarrow \bar{X}_{a,t_a}(\gamma)$ ▷ Discounted empirical average reward **return** $\mathcal{T}, Q(a)$

function WIDEN($\mathcal{T}$)
  if $C(\mathcal{T}) \leq N(\mathcal{T})^\alpha$ **then**
    $a \leftarrow$NEW(b, $C(\mathcal{T})$) ▷ Simulate new action
  else
    $a \leftarrow I_t$ ▷ Take maximized D-UCT arm.
  return a

5.4 Decentralized Decision Making in Simulation

The decentralized methodology may now be applied to a variety of sensor tasking problems. In each case, we consider the following problem design.

- K: the number of agents considered in the problem.
- $\mathcal{S}$: An ensemble of Gaussian space object state estimates, with objects labeled $1, ..., N$.
- $\{\mathcal{A}^i\}$: $[1, ..., N]$, describing the object an agent should observe at a given epoch
- $\mathcal{P} : \mathcal{S} \times \mathcal{A}$: Kalman updates are performed for each observation an agent makes, then nonlinear propagation is performed using the underlying dynamical system.
- R^i: a scalar reward function for the ith agent. Generally, this is reduction in state uncertainties across the catalog.
- $\gamma_i \in [0, 1]$: the discount factor over time for agent i.

We first consider a week-long simulation utilizing two sensors in Boulder, Colorado. The sensors are tasked in a decentralized manner to track approximately 200 geostationary objects, maintaining state estimates throughout the simulation. This simulation supports the application of MCTS to hardware at the University of Colorado at Boulder Vision, Autonomy, and Decision Research (VADeR) observatory.

5.4.1 Decentralized Geostationary Sensor Tasking

In Table 5.1, a brief overview of the sensors utilized in the simulation is provided. These models are representative of the real sensors utilized in the VADeR observatory, and each sensor offers the capability to detect and track geostationary objects, with limiting magnitudes of approximately 18 and 15, respectively. Sensors are tasked to observe an object every 30 seconds, and in this study, other objects that may lie in the field of view are not considered. Sensors are tasked in a decentralized manner, and allowed to communicate every 500 search iterations. In the allocated search time, approximately 1500 iterations generally occurred, resulting in an effective communication rate between sensors of 0.1 Hz. Because there are only two agents in

this problem, all information is immediately available to agents on communication, greatly simplifying any analysis of connectivity.

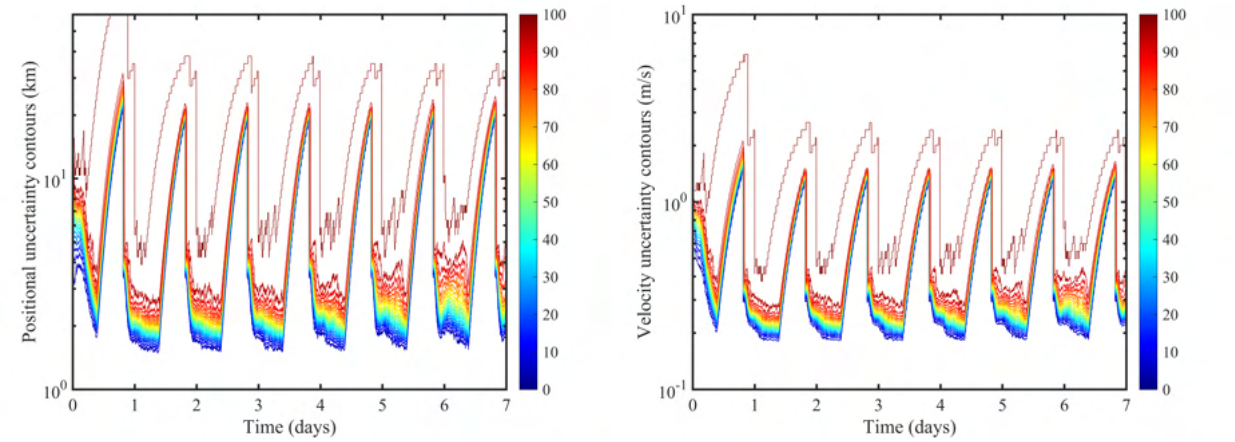
A rollout policy is applied sampling new actions within tree search as a function of projected state uncertainties into the agent field of regard. Sampling weights are scaled according to the trace of uncertainty, with the assumption that observation of objects with larger projected covariance traces leads to greater information gain. Reduction of state uncertainty is applied as a reward.

Specification	**SITH**	**PANOPTICON**
Aperture (m)	0.6	0.2
f-number	6.5	3
Pixel pitch (μm)	9	9
QE	0.74	0.74
Read noise (pix/s)	3.7	3.7
Optical Transmission	0.921	0.921
Atmospheric Transmission	0.7	0.7

Table 5.1: Sensor portfolio for the Vision, Autonomy, and Decision Research observatory.

Operation is assumed to proceed with no weather impacts over a one week period. In Figure 5.5, we first consider positional and velocity $3-\sigma$ uncertainties over time across the catalog. A relatively large amount of process noise is incorporated, leading to consistent growth in velocity uncertainties during daytime measurement gaps. Generally, velocity uncertainties remain on the order of a meter per second at the beginning of nightly observation, and quickly reduce to approximately 25 centimeters per second after observation. Approximately 20 kilometers a day of growth in positional uncertainties is a direct result of transformed velocity uncertainties over time. Otherwise, positional uncertainties are consistently on the order of 2 kilometers $3-\sigma$ during nightly observation.

It is also useful to consider the structure of state uncertainties in the field of regard of the observers. Figure 5.6 visualizes the projection of catalog uncertainties into measurement space alongside a custody bound of 10 arcminutes $3-\sigma$. The entire catalog is maintained to this custody bound throughout the simulation, and during observation, the catalog projected uncertainty consistently lies below 10 arcseconds $3-\sigma$. This is consistent with an expected standard deviation in observations on the order of several arcseconds for the SITH telescope, with slightly larger uncertainties using the PANOPTICON sensor. There



(a) Percentage of $3-\sigma$ positional uncertainties below contours over time.

(b) Percentage of $3-\sigma$ velocity uncertainties below contours over time.

Figure 5.5: Catalog uncertainty contours over a 1 week simulation using the VADeR observatory.

is very little variation between objects, because each object behaves in much the same manner throughout the simulation, and is effectively stationary from the perspective of each observer.

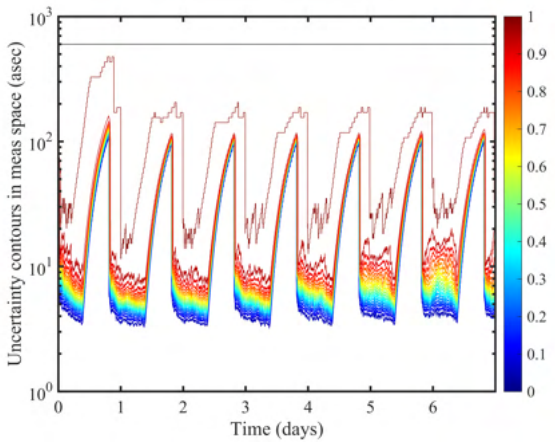
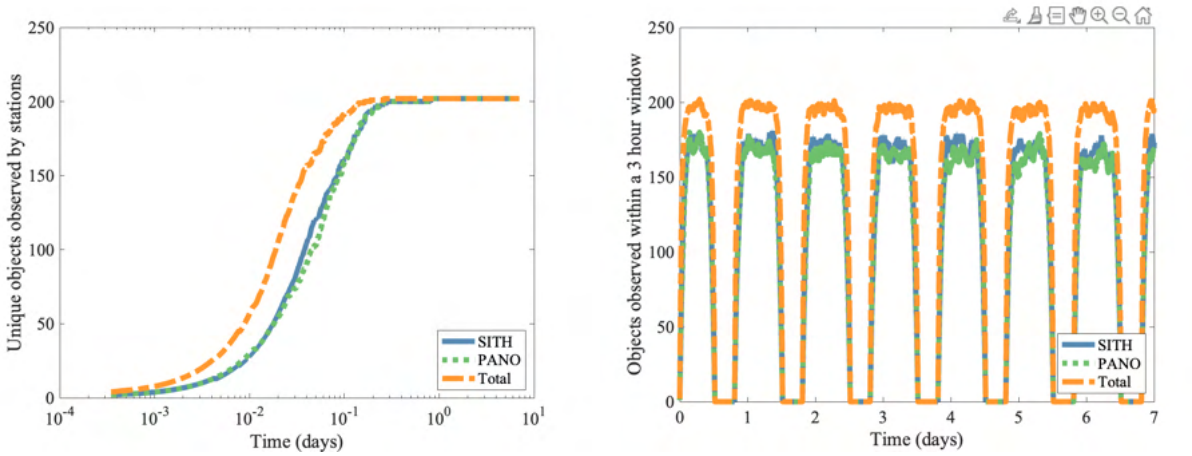


Figure 5.6: $3-\sigma$ uncertainty projections in the SITH field of regard.

Finally, we consider how objects across the catalog were studied. Figure 5.7 presents several interesting data features on object observation. All objects in the study were observed by approximately 2 and a half hours into simulation, and each sensor tracks new objects at an equivalent rate. The populations each sensor studies have some overlap, suggesting there was initial interest in further constraint of a subset of objects in the catalog. Likely, these were objects with comparatively large initial uncertainties. Proceeding further

into the simulation, sliding window visualizations of objects observed by each sensor are presented. Over a moving 3 hour period, each sensor observes approximately 170 objects, and the entire catalog is maintained between each sensor throughout each night. In addition to this information, Figure 5.7b also visualizes the day-night structure of ground-based observation, with gaps visible during each period in which no observation is possible.



(a) Unique objects observed and time to full observation of the catalog.

(b) Windowed observations of unique objects by instrument.

Figure 5.7: Data features in objects tracked by the VADeR observatory.

These results offer initial insight into the geostationary tracking capabilities with the VADeR observatory using decentralized MCTS, and support the extension of this methodology to more complex problems. It is also worthwhile to further probe the robustness of decentralized MCTS to loss of observers or lapses in communication, a challenge further addressed in the next section.

5.4.2 Cislunar Decentralized Decision Making

We now consider application of the decentralized MCTS methodology to a tasking problem in cislunar space akin to prior cislunar catalog maintenance simulations [32]. Here, we narrow the focus of the problem to a set of 100 SOs placed in Halo orbits about the Earth-Moon L1 and L2 Lagrange points. Each object intermittently performs stationkeeping maneuvers that are utilized in the truth trajectories for these states, and four agents are tasked to maintain state estimates on all objects in the catalog. Observers may make

Specification	Lunar	Space-based
Aperture (m)	0.2	0.5
f-number	3	7
Pixel pitch (μm)	5	5
QE	0.8	0.9
Read noise (pix/s)	2.0	2.0
Optical Transmission	0.756	0.756
Atmospheric Transmission	1.0	1.0

Table 5.2: Space and ground-based sensor specifications for decentralized cislunar catalog maintenance.

detections every 120 seconds, and a combination of two lunar surface observers and a space-based observer placed in a L2 Northern Halo orbit are utilized. In addition, the sensing architecture is augmented with a space-based observer placed in a L1 Northern Halo that is also 1:1 resonant with the Earth-Moon synodic period. The L2 Halo observer is instantiated at perilune, while the L1 Halo observer is instantiated at apolune, such that each observer occupies a largely distinct subset of state space throughout simulation, since observer periods are equivalent. The lunar observers are placed at the lunar north and south poles. The simulation is approximately instantiated at new moon, with an epoch Julian date of 2459153.5. The sensor models are outlined in Table 5.2.

We first consider the scenario with all sensors operating in a decentralized manner, with sensor communication every 500 tree search iterations. A MCTS rollout heuristic is applied for each sensor that incorporates state uncertainty alongside detected maneuver information and dynamical knowledge of maneuver potential [34]. Specifically, the trace of projected state uncertainty into the observer field of regard is utilized as a measure. If the trace of the projection is significantly larger than that of measurement uncertainty, the observation offers significant information gain. In addition, information from the Cauchy-Green Stress Tensor is incorporated to inform maneuver utility, and the resultant sample weight is rescaled if the studied object previously maneuvered [34]. Each component weight may be expressed as

$$\omega_{y,i} = \frac{\text{tr}(H_i P_i H_i^T)}{\sum_j \text{tr}(H_j P_j H_j^T)} \tag{5.31}$$

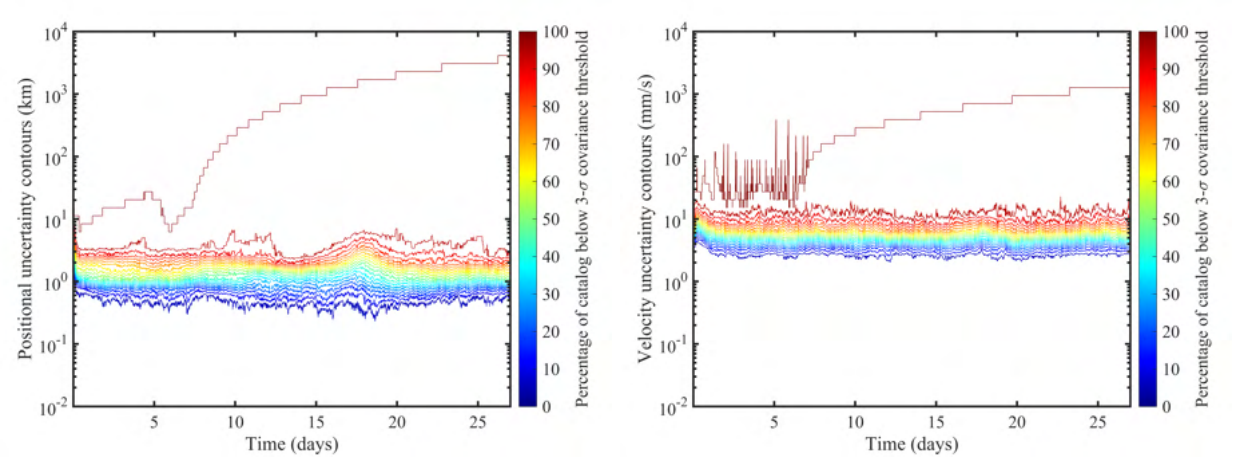
$$\omega_{\rho,i} = \frac{\rho_i(t + T_i, t)}{\sum_j \rho_j(t + T_j, t)} \tag{5.32}$$

with the full sampling weight

$$\omega_i = \xi_i \left(\nu\omega_{y,i} + \eta\omega_{\rho,i}\right) \tag{5.33}$$

$$\xi_i = \begin{cases} 1 & t - t_{M,i} > \tau \\ \omega_M & t - t_{M,i} \leq \tau. \end{cases} \tag{5.34}$$

Reduction of space object covariance traces is applied as a reward. In the presented results, a depth $d = 20$ is used for a search time t_s of 4 seconds for each observer. A discount $\gamma = 0.99$ is used. Positional and velocity covariance traces across the catalog are first presented in Figure 5.8. We find that the ensemble of sensors utilized are sufficient to maintain state estimates across the catalog of SOs, with approximate median $3-\sigma$ positional uncertainties of 2 kilometers and velocity uncertainties of 10 millimeters per second across 27 days of simulation. Note that one object is lost after a large maneuver at approximately 17 days into simulation.



(a) Percentage of $3-\sigma$ positional uncertainties below contours over time in cislunar space.

(b) Percentage of $3-\sigma$ velocity uncertainties below contours over time in cislunar space.

Figure 5.8: Catalog uncertainty contours over a month of observation with a suite of four cislunar observers.

It is also worthwhile to discuss the structure of the velocity uncertainty contours presented in the Figure. At any point in the simulation, the largest uncertainty contours evolve on the order of a meter per second; these spikes correspond with maneuver epochs for each epoch, and the largest velocity uncertainties in the catalog at a given time are generally associated with objects that have recently maneuvered. These structures are also clearly visible in Figure 5.9, in which state uncertainties are projected into the field of regard of the lunar north pole observer. Generally, these uncertainties evolve on the order of 10 arcseconds

$3-\sigma$, but clear spikes are visible on the order of several arcminutes as maneuvers occur. These structures are a direct result of the adaptive process noise methods that enable correction of state estimates as measurement residuals increase in norm.

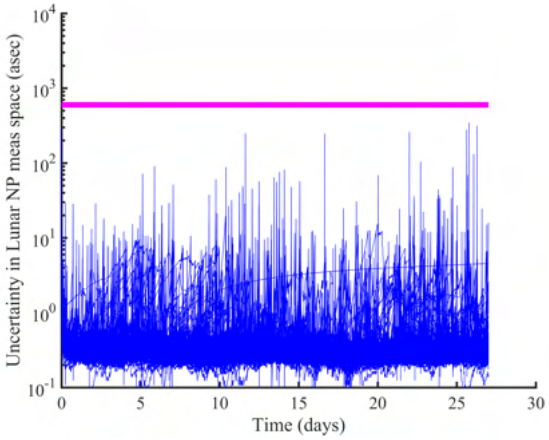
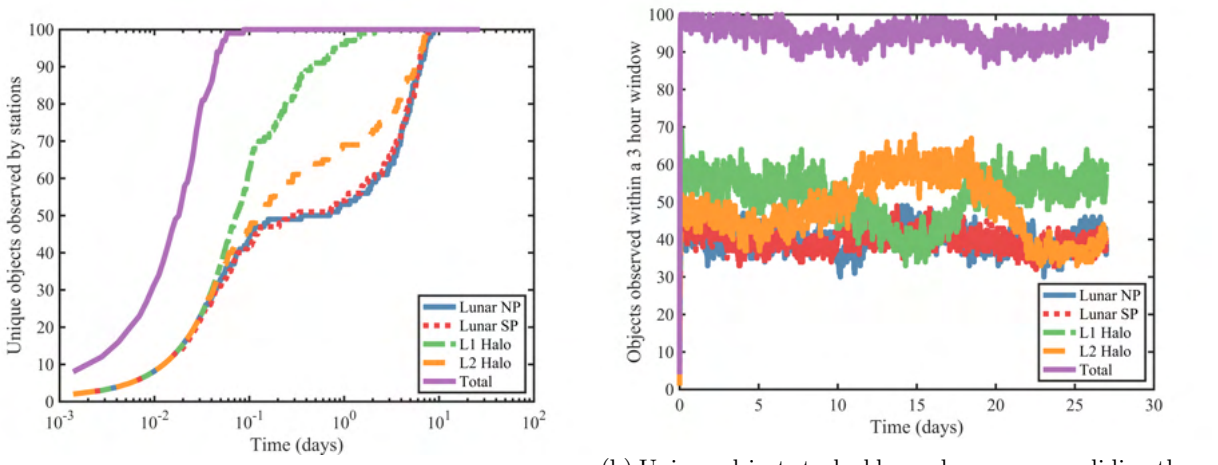


Figure 5.9: State uncertainties projected into the field of regard of an observer at the lunar north pole.

Very interesting structures may be noted when studying the objects tracked by each sensor in greater detail. Figure 5.10a first visualizes the number of unique objects that each sensor tracks. The full catalog is observed by approximately 3 hours of real time, and the L1 Northern Halo observer is found to be most impactful in the initial periods of the simulation. This is a direct result of the solar phase angle leading to little lunar illumination from the perspective of the L1 observer. Because the simulation is instantiated at new moon, the moon is quite illuminated from the perspective of the L2 observer, and thus, that observer struggles to observe many objects that evolve about L1. On the other hand, as the simulation nears full moon at approximately 14 days into simulation, the L1 Northern Halo observer is challenged to observe objects evolving about L2. This structure is clearly represented in Figure 5.10b, in which the percentages of the total population the space-based observers detect over a sliding window switch near full moon. It is also worthwhile to briefly discuss the lunar surface observers placed at the lunar north and south poles. Both observers consistently track a subset of the population, and each detects all initially visible objects relatively quickly, on the order of several hours. Other objects then slowly evolve into each sensor's field of regard, and all objects are observed by each sensor by the maximal orbital period in the catalog, on the order of 8



(a) Unique objects tasked by each sensor in cislunar space.

(b) Unique objects tasked by each sensor over sliding three hour windows.

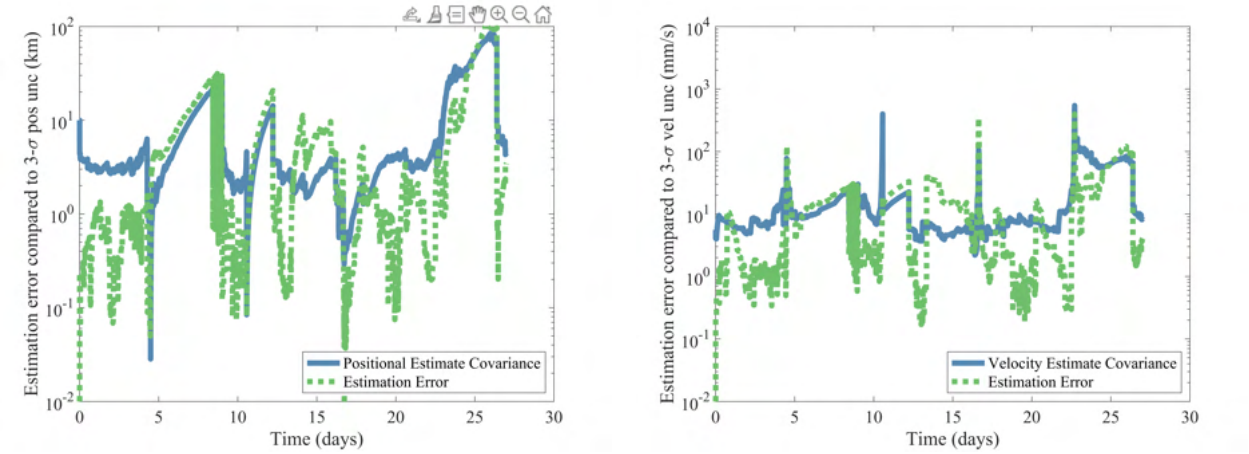
Figure 5.10: Tasking data features tracking maneuvering objects in cislunar space.

days.

Comparing the presented results to prior research [34] studying cislunar space objects through maneuvers, the combination of lunar surface and space-based observers admits an increase in positional resolution, while velocity uncertainties are relatively comparable between cases. This suggests that increasing the number of observers utilized is less of a factor in the reduction of velocity uncertainties; an alternative observer feature that may be more successful in this regard would be the utilization of a sensor with much greater pixel resolution. If such a sensor were incorporated, subsequent observations would offer greatly increased velocity information, leading to a greater reduction in velocity uncertainties. Another critical factor when considering velocity uncertainties is the presence of many maneuvers. Velocity uncertainties are greatly increased around maneuver epochs, likely leading to long-term challenges in reduction of uncertainties in velocity space.

Finally, it is useful to briefly demonstrate that estimators remain consistent in this scenario. Figure 5.11 visualizes state uncertainties alongside estimation error for a random object in the catalog, demonstrating successful state estimation. Note that some estimation error may be expected, since the object must be observed to initiate the correction process. This briefly occurs between 14 and 16 days, but is corrected after a maneuver at approximately 17 days into the simulation. It it also interesting to note several periods

during which it appears the object was weakly observable or unobservable. Most prominently, no observation appears to occur between 5 and 9 and between 11 and 12 days into the simulation. Maneuvers during the sample case largely appear to be accounted for.



(a) Positional state estimates for a L2 Southern Halo tracked using decentralized MCTS.

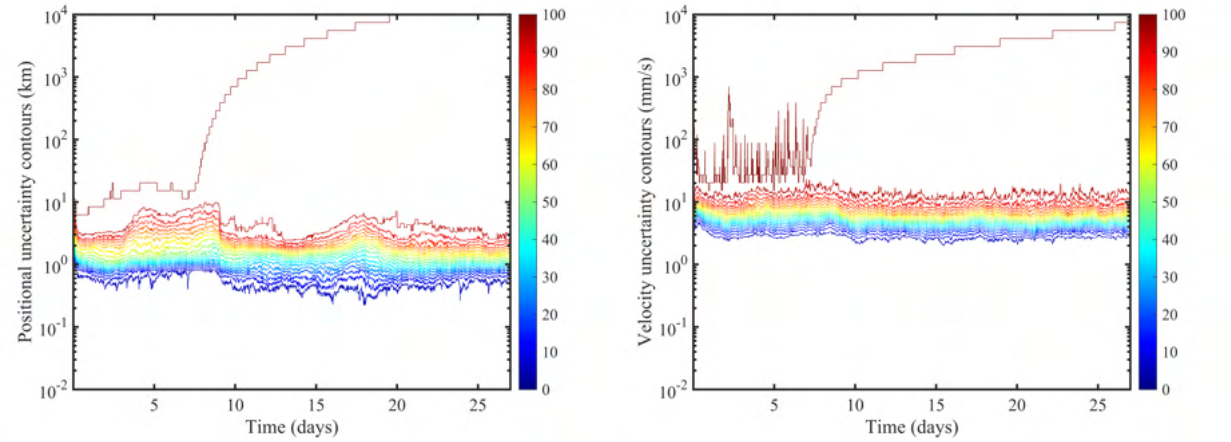
(b) Velocity state estimates for a L2 Southern Halo tracked using decentralized MCTS.

Figure 5.11: Successful state estimation with decentralized sensor tasking.

5.4.3 Robustness to Communication Failures

We now additionally consider a modification to the prior scenario in which communication failures occur over a subset of the simulation. In this case, no communication is possible with the L2 Northern Halo observer between 3 and 9 days into the simulation; during this period, the L2 observer must operate in isolation, while the lunar and L1 observers coordinate without the isolated observer. It is first useful to characterize the catalog uncertainties as a whole across the simulation. Between 3 and 9 days, the connected observers and the isolated observer have a differing notion of state estimates across the catalog, and when communication with the isolated observer is resolved, observations from all observers are processed sequentially over the period of isolation to form a common catalog between all observers. A visualization is given of the catalog used by connected observers between days 3 and 9 in Figure 5.12 and the isolated observer in Figure 5.13.

First considering Figure 5.12, we observe catalog uncertainties that are quite similar to those presented



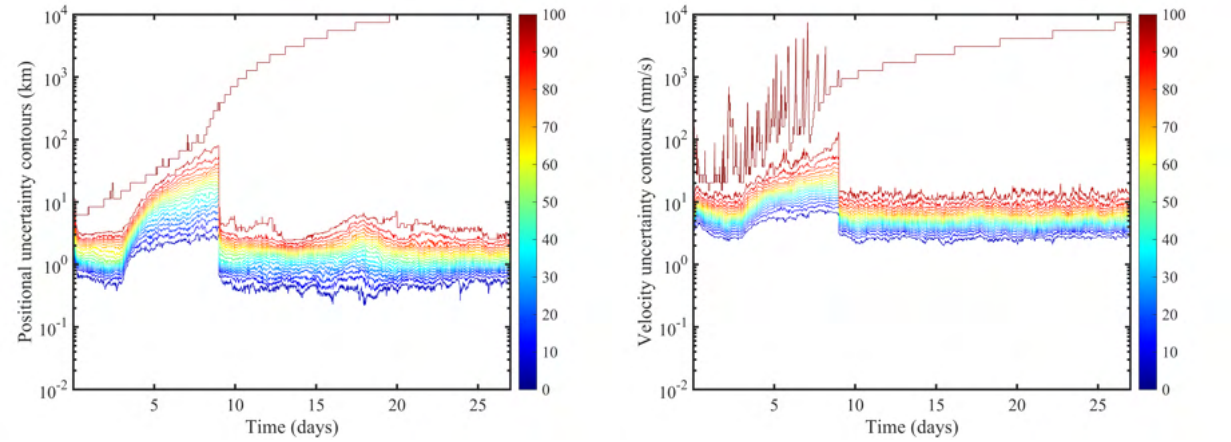
(a) Percentage of $3-\sigma$ positional uncertainties below contours over time in cislunar space.

(b) Percentage of $3-\sigma$ velocity uncertainties below contours over time in cislunar space.

Figure 5.12: Catalog uncertainty contours over a month of observation with a suite of four cislunar observers.

in the prior section. In this case, a single object is lost earlier in the simulation, while the remainder of the catalog maintained to approximate median uncertainties of 2 kilometers $3-\sigma$ in position and 10 millimeters per second $3-\sigma$ in velocity. During the period in which the L2 Northern Halo observer is isolated from the remainder of the sensing architecture, slight changes from the original case may be noted. Between 3 and 9 days, a slight increase in positional and velocity uncertainties across the catalog occurs, with median uncertainties reaching approximately 5 kilometers in position by 9 days. When information on the tasking history of the isolated observer suddenly becomes available at this time, the catalog immediately shifts back to covariances levels that are quite comparable to the fully connected case. Interestingly, there is little shift in velocity uncertainties, a feature that suggests it would be quite likely that lunar sensors and a L1 Northern Halo observer were sufficient to maintain the catalog over the period of isolation. This result further demonstrates the robustness of decentralized MCTS to challenges in communication, even in scenarios where space objects are maneuvering.

We next may consider structures in the catalog uncertainties from the perspective of the isolated observer. Figure 5.13 visualizes these uncertainties, and it is especially critical to study the Figure between 3 and 9 days into simulation. One may note that uncertainties increase somewhat quickly over this gap, with median positional uncertainties for the isolated observer reaching approximately 30 kilometers $3-\sigma$



(a) Percentage of $3-\sigma$ positional uncertainties below contours over time in cislunar space.

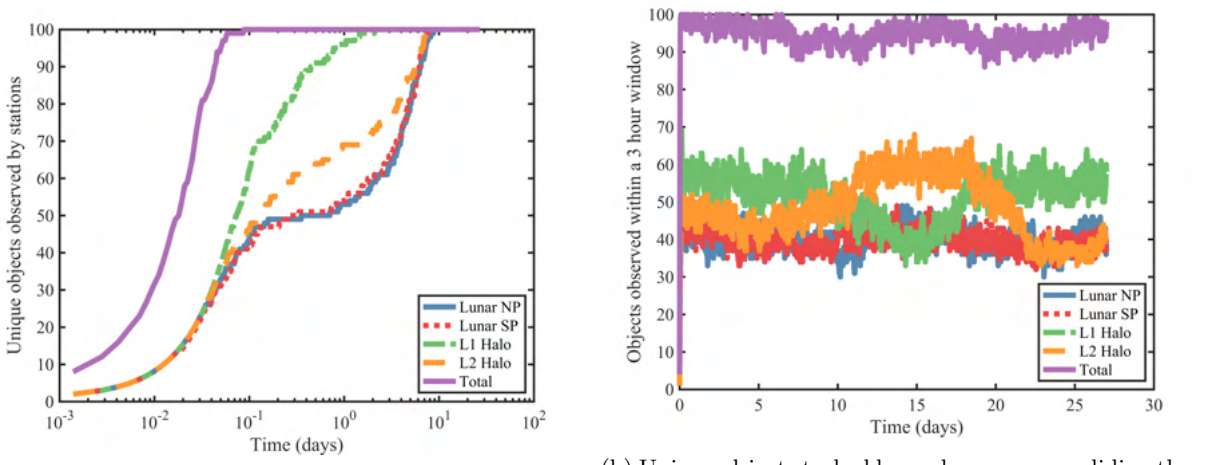
(b) Percentage of $3-\sigma$ velocity uncertainties below contours over time in cislunar space.

Figure 5.13: Catalog uncertainty contours over a month of observation with a suite of four cislunar observers.

by 9 days. While the observer continues to detect objects during the period of isolation, the observations it makes are not sufficient to maintain the catalog, though a subset of the catalog appears to stabilize at approximately 6 days into simulation, a feature that is more apparent in the velocity contours. The behavior of the isolated observer during this period may also be compared to the prior case in which the L2 Halo agent was never isolated. Figure 5.14 visualizes trends in objects tasked during the isolated case, and the structures present here are almost identical to those presented in Figure 5.10. As such, the isolated observer continues to visit all objects that it is able to observe, again supporting the robustness of the MCTS methodology.

It is briefly worth revisiting structures in estimation error for this scenario, and a random object is selected for visualization. Figure 5.15 visualizes estimation error for an object following a L1 Northern Halo orbit with three prominent maneuvers at 6, 14, and 22 days into simulation. The estimator is largely consistent throughout the simulation, with a brief period during which positional errors exceed $3-\sigma$ covariances. Smoothing results were note logged in this scenario, so maneuver estimates are not visualized, but covariances successfully account for maneuvers in each case and grow to approximately 0.2 meters during each maneuver epoch. Because of a diversity of viewing geometries, few structures in uncertainties or estimation error as a function of orbital phase are observed.

Finally, it is useful to consider whether uncertainty projections into measurement space remain small



(a) Unique objects tasked by each sensor in cislunar space.

(b) Unique objects tasked by each sensor over sliding three hour windows.

Figure 5.14: Tasking data features tracking maneuvering objects in cislunar space.

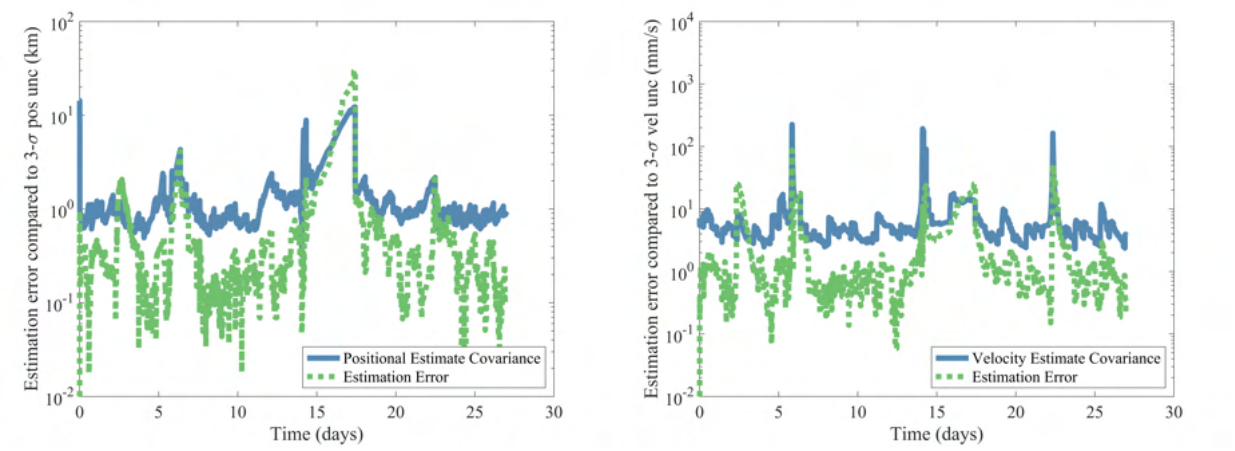
enough that custody is maintained for all objects. Figure 5.16 visualizes measurement space projections, again for the lunar north pole observer, and while spikes occur around maneuver epochs, uncertainties are largely maintained to several arcseconds $3-\sigma$ in measurement space. Compared to Figure 5.9, a small increase in projected uncertainties occurs between 3 and 9 days as the L2 Northern Halo is isolated, with uncertainties increasing to 10 arcseconds on average during this period.

The results of this scenario establish robustness of decentralized MCTS to failures in communication, and safe operation over large gaps in communication is an ideal feature for space domain awareness in the cislunar regime.

5.5 Discussion and Conclusions

This chapter establishes a decentralized methodology for MCTS sensor tasking, increasing scalability of MCTS to many-agent problems that are expected in SDA, especially in the cislunar regime. Several components of this contribution are worth revisiting in further detail.

First, the development of a communication paradigm using random graphs ensures that the MCTS methodology remains unexploitable and offers guarantees in communication time, connectivity, and robust connectivity. As visualized in Figure 5.4, this methodology reduces communication rates as compared to



(a) Positional estimation error for a L1 Northern Halo. (b) Velocity estimation error for a L1 Northern Halo.

Figure 5.15: Estimation error structures with decentralized MCTS tasking and communication failures during observation.

a tasking scheme in which all agents communicate directly with each other by orders of magnitude. The random graph communication scheme ensures that lines of communication aren't known to non-operators, a distinct advantage compared to a centralized spoke-hub paradigm. Further communication architectures may also be explored alongside the MCTS methodology, especially for many-agent problems, and it would be interesting to compare MCTS returns if an all to all architecture is utilized. This comparison is not made in the presented results, since there are so few agents, but this comparison could easily be made in the context of problems such as UAV patrolling.

The effects of discrete communication are then analyzed within the context of MCTS, and further analysis is feasible for this problem. An interesting avenue of further research would be application and analysis of the polynomial form of MCTS, as is analyzed in Chapter 3, to the decentralized problem. Further analysis would also be beneficial to support Chapter 5, Theorem 1, as the extension of the analysis of nonstationary multi-armed bandits to D-UCB trees is nontrivial, due to the nonstationary effects of leaf nodes on the parent nodes in the search tree. One avenue for better demonstrating this result could be incorporation of the effects of leaf nodes themselves as breakpoints, a method of analysis that may be expected to reduce the convergence rate of the presented algorithm. In addition to analysis of the algorithm, a visualization of methods for implementation of the decentralized methodology is presented.

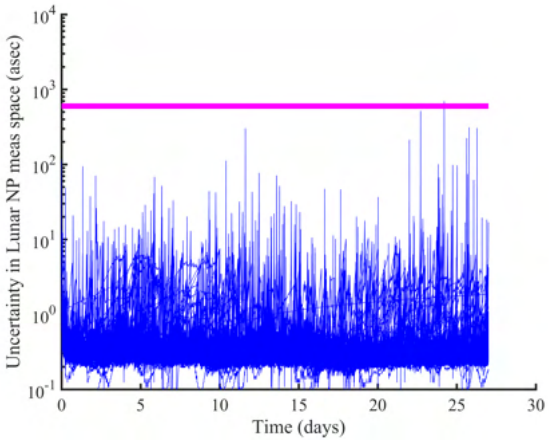


Figure 5.16: State uncertainties using decentralized MCTS with communication failures projected into the field of regard of an observer at the lunar north pole.

Finally, we present results utilizing the decentralized MCTS methodology in several scenarios. First, decentralized MCTS is used to support live sensor tasking with the VADeR observatory, and results in this case support the conclusion that the observatory may successfully maintain state estimates on a large catalog of geostationary objects. The methodology is then extended to the cislunar regime, demonstrating decentralized sensor techniques for a challenging tracking problem and uniting the variety of methods presented throughout this dissertation. The decentralized methodology is further shown to be robust to loss of communication between observers, a key feature for successful operation of autonomous systems in space domain awareness.

Chapter 6

Optical Sensor Tasking with the VADeR Observatory

As a final contribution to this thesis, the sensor tasking methods presented are applied to the Vision, Autonomy, and Decision Research (VADeR) observatory. The observatory, installed in July 2021, fields an array of optical sensors that may be autonomously directed. The primary telescope in the observatory, "SITH", is a 0.6 meter f/6.5 corrected Dall-Kirkham optical tube with an approximate diagonal field of view of 44 arcminutes that operates in the visible spectrum using an array of Sloan filters. Alongside the primary telescope, the observatory also operates a coaligned array, "PANOPTICON", of four 0.2 meter f/3 Riccardi-Honders astrographs of approximate diagonal field of view of 2.82 degrees. Each optical tube on the coaligned telescope utilizes a different supporting sensor, and the observatory may use this system for observation with Sloan filters in the visual spectrum, optical spectroscopy, observation in the infrared spectrum, and rate observation using an event sensor. This instrument may also be reconfigured to stack imagery in the visual spectrum or yield a mosaiced field of view of 5 degrees. All instruments in the observatory are operated using Planewave L600 mounts, allowing for slews of up to 25 degrees per second and accurate rate tracking of space objects.

6.1 Distributed Observatory Operations

A variety of developments were required to ensure safe autonomous operation over a variety of systems. A visualization of the observatory architecture is provided in Figure 6.1. In order to safely utilize observatory sensors, constraints are placed on mount control and dome operation. In order to open the VADeR dome, weather APIs are utilized to ensure cloud cover, precipitation probabilities, wind speeds, and humidity

remain below desired thresholds. In the future, automated visual inspection of any snow cover using security cameras will also be applied. To autonomously operate observatory mounts, the VADeR dome must also be opened.

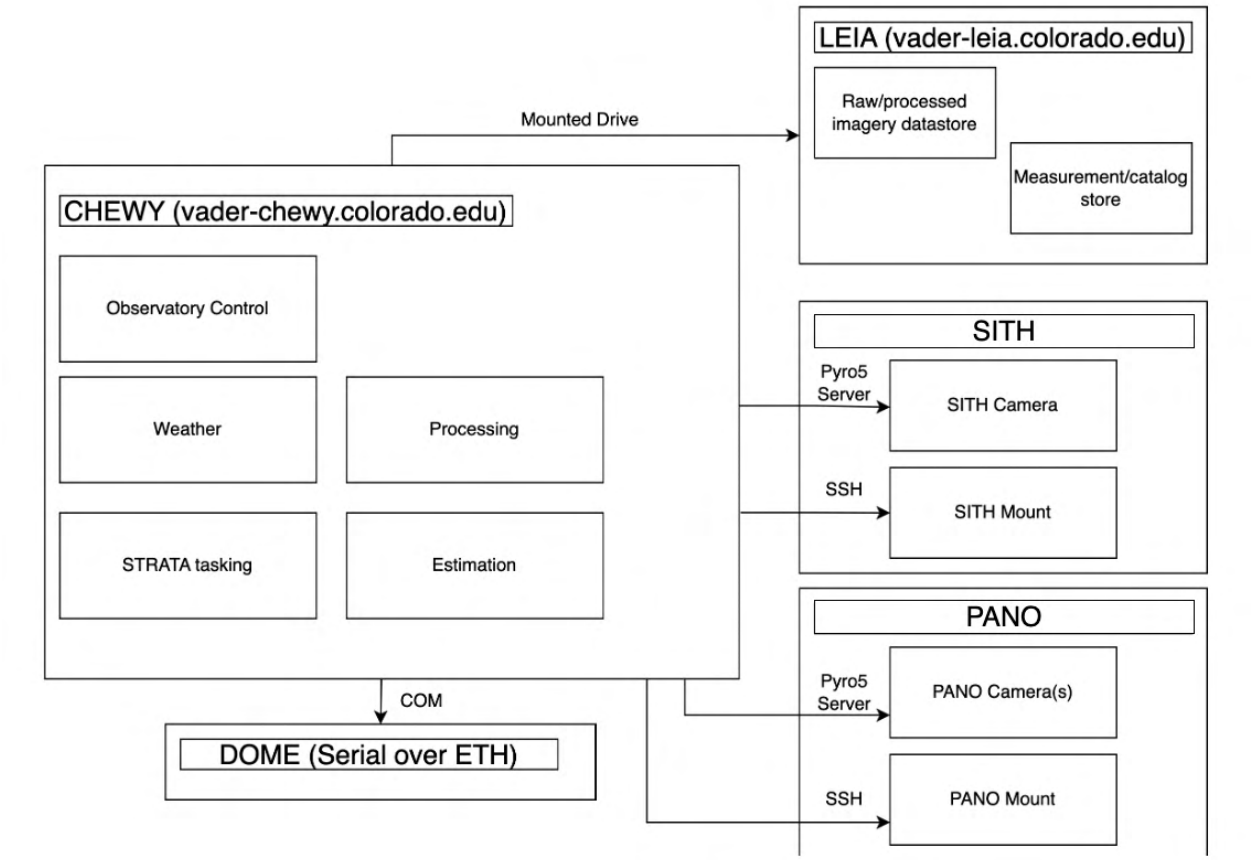


Figure 6.1: Operations diagram for the VADeR observatory

Autonomous operation of the VADeR observatory also requires coordination between a variety of systems. The VADeR CHEWY server acts as the central node in autonomous operation, managing any control input, image processing, tasking, and estimation threads. The server stores generated data over a mounted drive and interfaces remotely with telescope control systems. A client-server architecture is utilized for imaging with each telescope camera, and both the SITH and PANOPTICON mounts are tasked via SSH. Finally, the VADeR dome is operated over a networked serial interface.

6.2 Data Processing and Imaging

This section outlines the processing and source extraction pipelines utilized within the observatory. Each processing step is abstracted such that custom methodologies may be applied, and methods are exemplified specifically for the sCMOS sensors utilized within the observatory for visual spectrum applications. A visualization of the processing pipeline as a whole is presented in Figure 6.2.

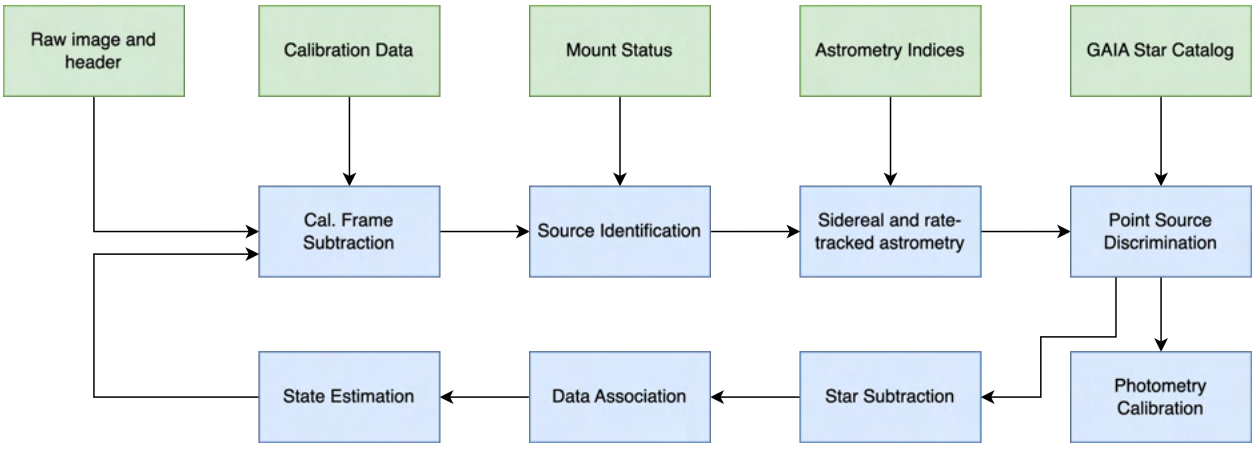


Figure 6.2: Image Processing pipeline for the VADeR observatory

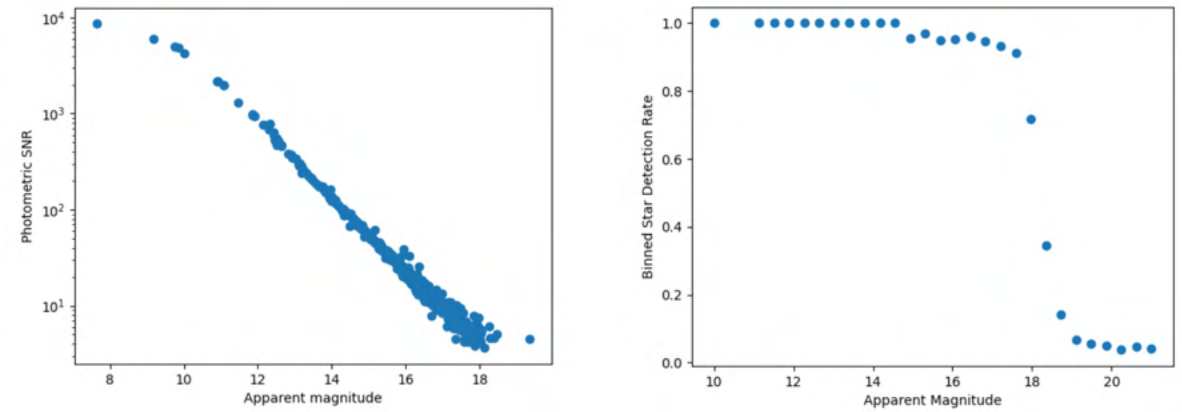
Within the observatory, bias and dark frames consistent with the exposure length are captured each campaign prior to observation. Because the sCMOS sensors used are relatively nonlinear in exposure time, dark frames at the exposure time are necessary for accurate subtraction of dark current. In addition to these frames, flat field images are necessary for correction of vignetting within images, a feature that is especially prominent for the 0.2 meter optical systems.

A variety of methods may utilized for subtraction of background noise from imagery, and a mix of open source and custom methodologies may be applied, including iterative sigma clipping and mode estimation [99], sigma clipping and box interpolation, and iterative polynomial fitting [66]. The resultant background-subtracted image may then be expected to be zero-mean and adequately prepared for source identification. Optionally, additional procedures such as registration of hot pixels and correction of edge pixels may be applied.

Largely, the source identification process makes use of the well-documented Source Extractor library

[10]. Within Source Extractor, a pyramidal convolution is applied to threshold and merge sources above a specified photometric signal to noise ratio. For point sources within the image, this methodology may be used to accurately identify the point source centroid, and for streaking sources in an image, the methodology may be used to identify streak centroids corresponding to the location in the image of the streaking object at the midpoint of exposure. Streaks act as a somewhat challenging problem for deblending potentially separate objects, but Source Extractor is often used for such processes with use cases for objects such as extended galaxies. Generally, saddle points are utilized to make deblending decisions, and since one may expect that a bright streak is resolved in an image as a multivariate combination of error functions [113], Source Extractor most commonly identifies streaks as a single source.

The resultant pixel locations of source centroids are then applied to generate an astrometric solution for the telescope field at the midpoint of exposure. This result is generated using the Astrometry.net library [65] and the 2MASS star catalog [95]. With information sourced from mount statuses on approximate pointing and a priori knowledge of approximate sensor field of view, this solution is generated on the order of milliseconds. As a result, pixel information at the midpoint of exposure can be translated into angular information on the celestial sphere. Augmenting this process with accurate timing and high-frequency information on mount slew rates, one may then describe pointing throughout the exposure.



(a) Photometric SNR of detections against star apparent magnitudes.

(b) Binned star detection rate by star apparent magnitudes.

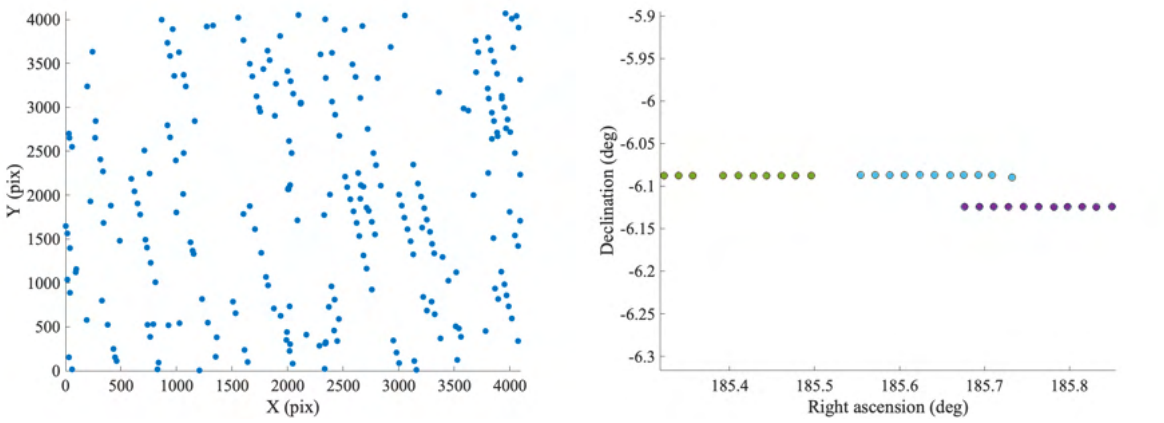
Figure 6.3: Apparent magnitudes plotted against photometric SNR and detection rates for the SITH telescope.

Because the 2MASS catalog generally only includes stars with apparent magnitude of 15 or brighter, the GAIA DR3 catalog is applied for star subtraction processes and photometry estimation. Currently, the data release consists of known apparent magnitudes for approximately 1.8 billion sources with a limiting magnitude of 21 [104]. Sources in the catalog are matched using k-d trees, and a regression between known apparent magnitude and image photometric SNR may be formed. The resultant regression may then be used to estimate the apparent magnitude of detected space objects. Additionally, matched objects may then be subtracted from the image. Subtracting stars from imagery is relatively straightforward when stars are realized as point sources, but is challenging when stars streak through the sensor field of view. In such cases, many artifact detections can be left in imagery, necessitating multi-target tracking methods for data association. The star registration process is visualized in Figure 6.3 and used to demonstrate an approximate limiting magnitude of 18 for a 30 second exposure using the SITH telescope.

A source may then be associated with angular and photometric information, and it remains to associate observations with prior state estimates or new tracks. For this problem, a variety of methods may be incorporated, including joint probabilistic data association, [98], multiple hypothesis tracking [13], or finite set statistics [107]. Currently, for rate-tracked imagery, a Gaussian Mixture Probability Hypothesis Density Filter is used to confirm tracks from sequential imagery [107]. Detections may then be gated for a specific track, and the resultant measurement set can then be associated with a catalog object. The effects of this process are visualized in Figure 6.4, where 11 images are taken centered on the ECHOSTAR 11 satellite with two other geostationary objects in the field of view. The dynamical evolution of tracks and measurement updates in the full state space may then be considered in a variety of manners, most commonly using the Unscented Kalman filter [59]. Orekit is utilized within this context for state and uncertainty propagation [75].

6.3 Observatory Tasking

The tasking methods previously in this paper may be applied asynchronously with observatory control, imaging, and processing pipelines with the overall goal of maintenance of a large catalog of space objects. A Python interface to the underlying C++ models outlined in the prior chapters of this thesis was devel-



(a) Initial satellite and clutter detections across 11 images of ECHOSTAR 11.

(b) Final geostationary tracks plotted in right ascension and declination.

Figure 6.4: Successful data association using Gaussian Mixture PHD filters.

oped using SWIG [8] in order to interface tasking libraries with the observatory. The tasking procedure is performed over a receding horizon while data collection for the prior tasking decision is performed. Tasking decisions are assumed over a longer horizon of two minutes while shorter exposures are taken within the observatory. This admits redundancy for cases in which poor imagery is collected and allows for observation of space objects over longer orbital arcs. The developed Python interface extends all capabilities previously presented in this thesis into the observatory environment, including catalog maintenance, follow-up observation, tasking for maneuvering targets, and decentralized sensor tasking. While the underlying C++ models do not support significant data association and multi-target tracking capabilities, these tasks may be performed outside of the tree search loop; tracks with high probability of existence are passed into tree search catalogs and labeled if association with NORAD-cataloged space objects is possible.

An example campaign is performed with the decentralized tasking methods presented in Chapter 5, studying the local geostationary environment. Tracks are maintained for a total of 141 space objects following near-geostationary orbits, with data collected between the dates of May 1st, 2023 and May 9th, 2023. The catalog is visualized as a set ground tracks in Figure 6.5. As is visible, the catalog consists of a mix of active, publicly known geostationary objects and defunct objects operating in graveyard orbits about the geostationary belt. The catalog is instantiated on May 1st, 2023, using Two-Line Elements and

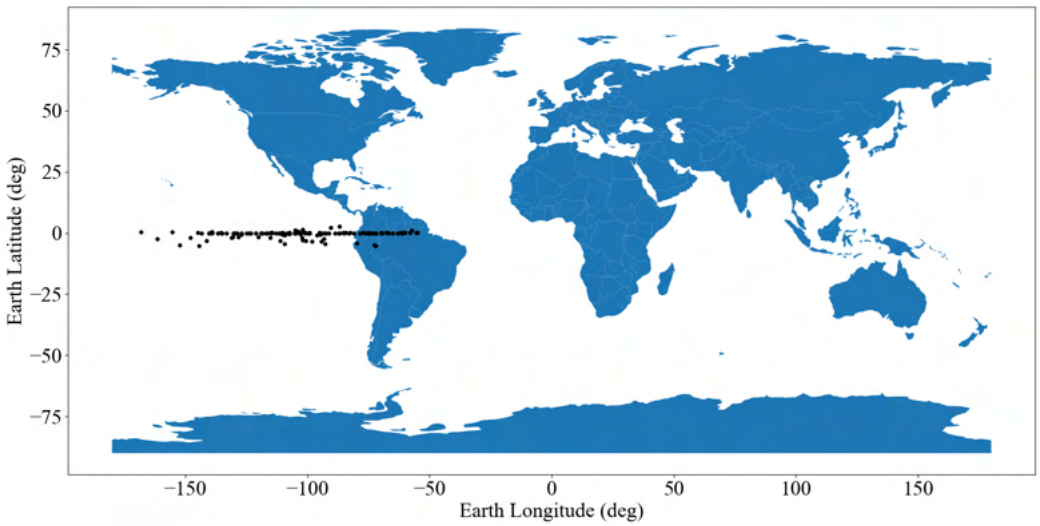
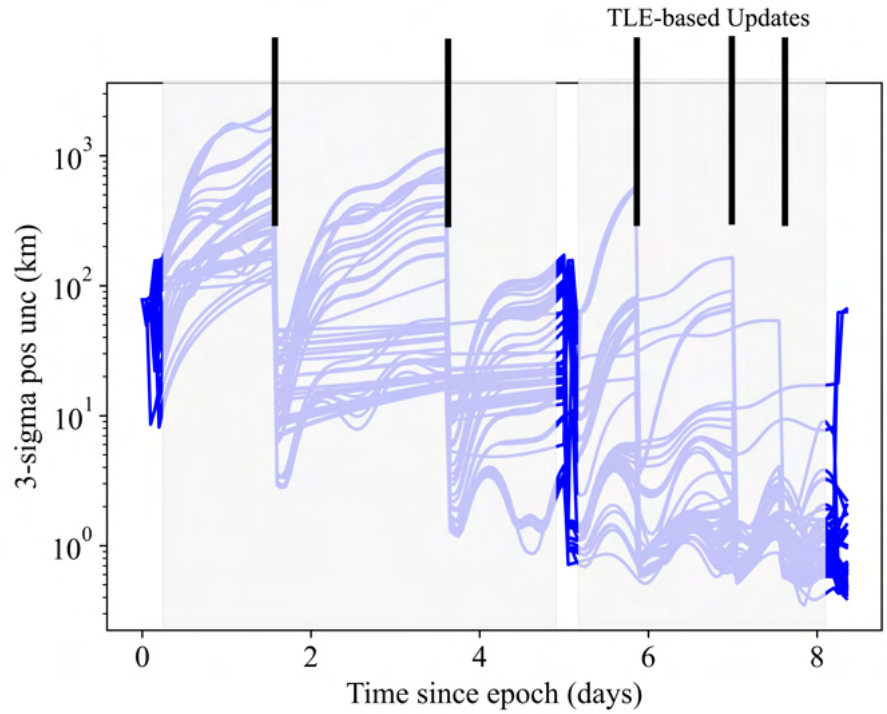


Figure 6.5: A geostationary catalog projected onto the surface of the Earth.

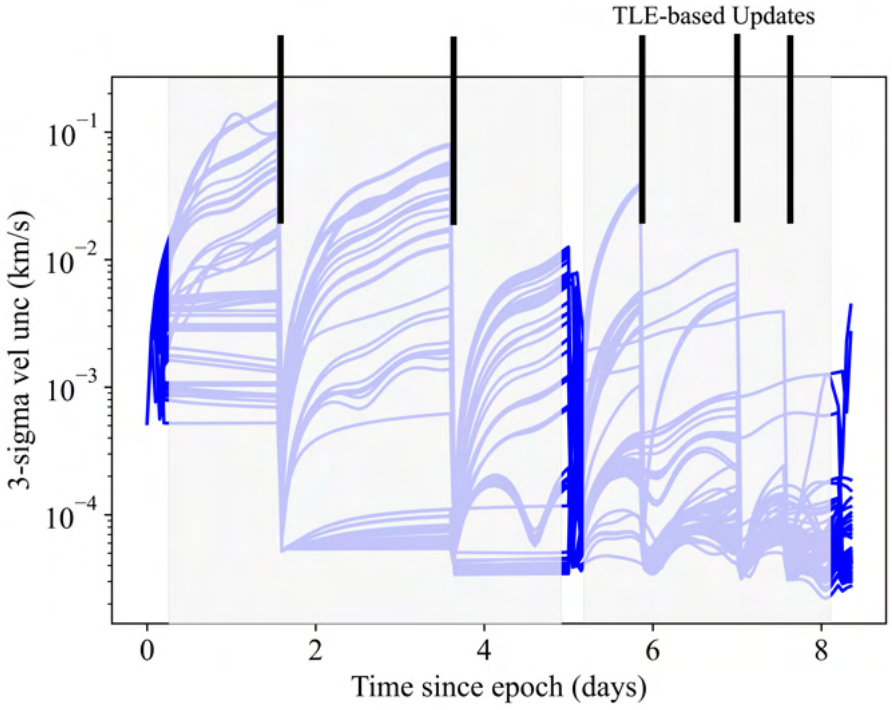
conservative uncertainties.

In Figure 6.6, we first present catalog uncertainties with tracking using a combination of optical observations and intermittent full state updates sourced from Two-Line Elements. Three data tasking periods are incorporated in the study; the first occurs during the first four hours of observation, the second occurs for four hours at 5 days into the study, and the last occurs at 8 days into the study. Significant reductions in state uncertainties may be noted during the periods of observation, and each period is framed in the Figure, while periods without observation are made opaque. Two-Line Element-based updates are also performed at five noted points in the study, at approximately 1.5, 3.5, 6, 7, and 7.5 days into the test scenario. These updates ensure that state uncertainties don't grow so large that custody is lost, a feature that is especially relevant because of initially conservative $3-\sigma$ uncertainties of 45 kilometers in each position axis and 300 millimeters per second in each velocity axis.

The majority of the catalog is eventually maintained with positional uncertainties on the order of 1 kilometer for the vast majority of the catalog and velocity uncertainties on the order of 100 millimeters per



(a) 3-σ positional uncertainties using the VADeR observatory and TLE updates.



(b) 3-σ velocity uncertainties using the VADeR observatory and TLE updates.

Figure 6.6: Catalog uncertainty traces over three nights of observation using the VADeR observatory.

second $3-\sigma$. This is consistent with the results presented in Chapter 5, but because of weather constraints, it was challenging to consistently observe the catalog. This motivates further data collection in Summer 2023, with the goal of nightly observation over at least a one week period. Additionally, the catalog should have been instantiated with much less conservative initial state estimates. It is also worth noting that several objects needed to be reinstantiated because of filter divergence. This was likely a result of small maneuvers that are expected to be quite common for geostationary objects, and it remains to extend the estimation methods presented in Chapter 4 of this dissertation to a Pythonic environment for use with the observatory, as well as for integration with industry standards for propagation such as Orekit. Such a filter would better account for scenarios in which measurement residuals quickly become large as a result of maneuvers. Another potential source of these errors could be uncertainty in the astrometric solutions found using Astrometry.net, as well as inconsistent rates at which astrometric solutions were found for imagery. Because stars are resolved in imagery as small streaks, it is likely that incorporation of a matched filter using the expected profile of stars from mount information could improve pointing estimation.

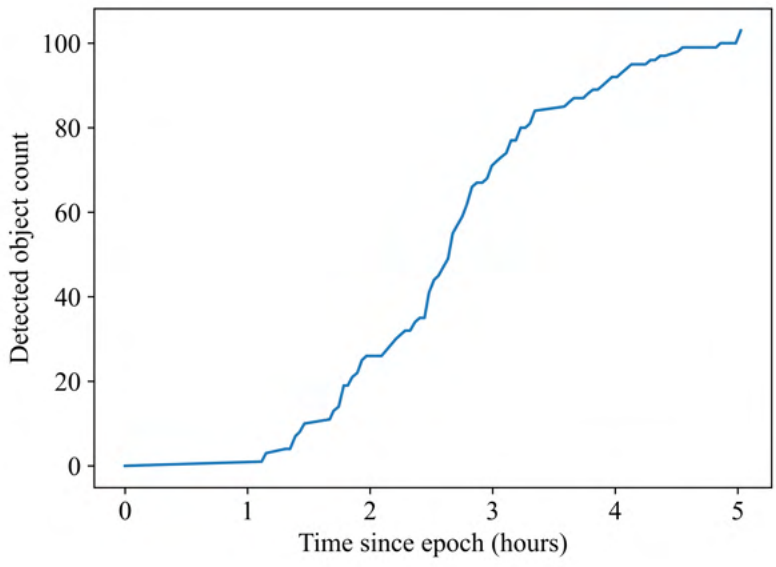


Figure 6.7: Unique objects tracked by the observatory on May 1st, 2023.

It is also useful to visualize detection capabilities during periods where observatory tasking is per-

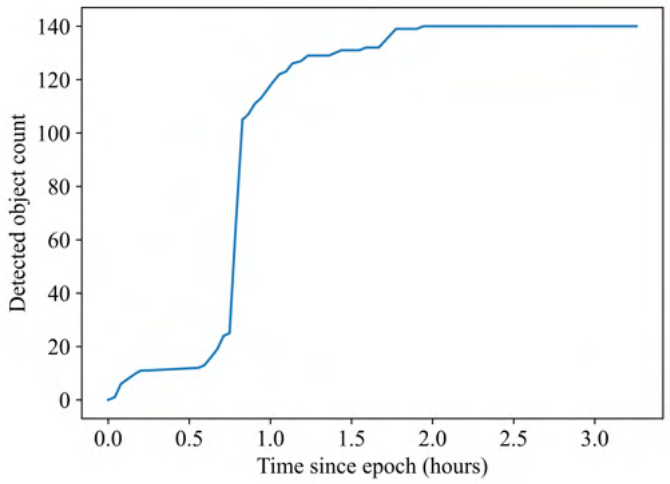


Figure 6.8: Unique objects tracked by the observatory on May 6th, 2023.

formed. The total number of objects tracked on two nights of observation are visualized in Figures 6.7 and 6.8. Note a discrete spike in the second case, where a TLE update was performed midway through the tasking period. In each case, approximately 120 of the 141 space objects were visited. More detailed analysis could be performed considering visits by sensor, since one may expect more objects to be detected by the PANOPTICON sensor, because of a larger field of regard, while the SITH sensor offers more detailed measurement updates.

These results demonstrate the utility of the methods presented in this thesis for online and decentralized sensor tasking in a near-optimal manner. Physical results validate the underlying algorithms and motivate use of Monte Carlo Tree Search in more complex sensor tasking scenarios. MCTS shall continue to be applied to the VADeR observatory, considering the long-term impact of autonomous sensor tasking on observatory operations.

6.4 Discussion and Conclusions

This chapter presents contributions that greatly augment the autonomous capabilities of the VADeR observatory. We first present pipelines developed for safe operation, autonomous imaging, image processing, and data association. These tools are then utilized alongside a Pythonic interface to MCTS libraries to

autonomously task the VADeR observatory. Several nights of autonomous observation were performed demonstrating MCTS as a viable algorithm for autonomous space domain awareness in a real scenario. State estimates are maintained for the public catalog of geostationary objects visible from Boulder, Colorado, with further studies planned in the coming months.

Several avenues of research may augment the capabilities of the VADeR observatory. Of particular interest is the incorporation of novel estimators presented in Chapter 4 to inform autonomous maneuver detection and estimation. Such capabilities are necessary for real scenarios in which objects are expected to maneuver often. In addition, incorporation of search and follow-up observation to autonomous operation is a clear next step. Currently, catalogs are instantiated using two line elements, but an incredibly useful demonstration of the MCTS methodology would be to autonomously instantiate a catalog via correlated tracks while cataloged state estimates are maintained. Much of the research presented in Chapter 3 on follow-up observation may be applied within this process, and search over subsets of the celestial sphere such as the visible portion of the geostationary belt may be applied to instantiate admissible sets on which follow-up tasking may be performed.

Chapter 7

Conclusions and Future Work

This thesis has presented a series of methodologies introducing Monte Carlo Tree Search to the sensor tasking problem, augmenting search techniques for agile targets, and decentralizing decision making for many-agent problems. The results illustrated alongside these theoretic developments outline the feasibility of Monte Carlo Tree Search for a variety of sensing regimes in both simulation and real scenarios.

7.1 Research Review

This section provides a brief review of the major contributions in each chapter, providing context for final conclusions of the dissertation.

7.1.1 Sensor Tasking for Catalog Maintenance

We first revisit the introduction of MCTS-based sensor tasking to space domain awareness, where specific focus was given to the catalog maintenance problem. This primary contribution represents advances in both decision making and space domain awareness literature. Analysis was outlined for a polynomial form of MCTS, and extensive numerical analysis of reward structures was performed. This analysis is a beneficial addition to MCTS literature that supports recent advances that apply polynomial MCTS [6, 90]. Additionally, a variety of rollout methodologies are introduced, incorporating space domain awareness-based and information-theoretic knowledge into MCTS. Such techniques increase MCTS convergence and ensure observations that offer useful information are prioritized.

The results presented in this chapter represent the first known scenario in which successful catalog

maintenance is demonstrated for space object populations in the cislunar regime. In these results, this chapter initiates discussion common throughout this dissertation, contemplating observational needs for the emerging focus of space domain awareness. The presented results augment knowledge of necessary components for successful space object tracking and act as preliminaries for the more complex tracking scenarios presented in later chapters.

7.1.2 Sensor Tasking for Follow-up Observation

Monte Carlo Tree Search is next applied to an alternate tasking scenario, in which follow-up observation of a single uncertain object is desired, rather than coordinated observation of an ensemble of objects. This contribution advances both decision-making and estimation within the broader context of space domain awareness. It also demonstrates that MCTS is not limited in application to a subset of critical problems in SDA, but rather, may act as a general tool for scenarios that may be formulated as sequential decision-making problems.

To augment the decision-making process, several rollout heuristics are first developed that consider how an uncertain object may dynamically evolve in the field of regard of an observer. This information is utilized to explore actions in which an observer searches over a subset of measurement space. While the search process progresses, consideration is also given to the underlying estimate of the studied object. A novel methodology for incorporating negative information as an update on a Gaussian sum filter is presented, increasing knowledge of object states when no detection is made. This contribution is critical for scenarios in which limited observations are feasible or in which follow-up observation and orbit determination for an object is of utmost priority. The presented techniques may be extended to pure search, or incorporated into multi-objective optimization problems.

7.1.3 Sensor Tasking for Maneuver Detection and Estimation

The ensuing chapter then extends Monte Carlo Tree Search to a series of more challenging tracking problems, first considering application of the sensor tasking methodology when maneuvering space objects are tracked. Such scenarios require more complex estimators for successful autonomous operation, and

MCTS may be augmented to further consider maneuver impact and potential.

First, a novel unscented smoother is presented for maneuver detection and estimation. The smoother extends the Optimal Control-Based Estimator [74], and detections of maneuvers above a certain threshold are utilized to cue priority sampling for maneuvering objects. Additionally, local dynamical systems analysis is performed using the Cauchy-Green Stress Tensor, informing MCTS on the feasibility of maneuvers in terms of long term impact of a maneuver on the perturbation from a nominal trajectory. Both the unscented OCBE smoother and CGT analysis are applied alongside covariance-based metrics to augment MCTS rollout heuristics for maneuvering targets.

MCTS is then applied for a scenario in which 100 stationkeeping objects in Halo orbits are observed over an Earth-Moon synodic periods. MCTS is demonstrated to yield consistent and high-resolution state estimates, and the developed estimator is robust to large maneuvers and observation gaps. The estimator is further studied for challenging scenarios tracking transfer trajectories between L1 and L2 Northern Halo orbits.

7.1.4 Decentralized Sensor Tasking

The final theoretic contribution in this thesis considered extension of MCTS to a decentralized paradigm in which each agent generates its own search tree in isolation and intermittently shares promising actions it may take with other agents. This contribution increases the realism of any analysis performed using MCTS-based tasking, and it is desirable for a space-based observer to make tasking decisions onboard, rather than via a centralized operator.

An initial contribution in this chapter considers a communication protocol for sharing information between agents using random graphs. This methodology offers guarantees on communication time between agents by studying the diameter of the resultant graph. It ensures communication is robust to lines of failure with probabilistic guarantees on r-connectivity. These features are incredibly important for many-agent problems and support future observational needs as space object populations greatly increase. The effects of discrete communication between agents are then analyzed in the context of decentralized MCTS. Guarantees are derived with more general assumptions on the breakpoint effects of communication, supporting claims

for MCTS as an asymptotically convergent methodology.

A variety of results are then presented using the developed decentralized MCTS methodology. A simulated study utilizing the VADeR observatory to track geostationary objects is first outlined, and promising results motivate further extension of decentralized MCTS to physical operation of the observatory. Decentralized MCTS is then demonstrated for a many-agent cislunar tracking problem, and shown to be robust to large gaps in communication. These results unite the contributions presented throughout this dissertation, demonstrating consistent decentralized tasking and estimation for agile objects in nonlinear domains.

7.2 Future Work

Several avenues of further research may be considered as a result of this thesis. First, the developed methodologies offer a unique means of characterizing the effectiveness of observing architectures. Some discussion of such subjects is presented throughout this dissertation, but there remains great opportunity to study observational efficacy across a range of multi-agent architectures, especially in the cislunar regime. This analysis differs from recent approaches to the architecture design problem [105, 61] in that both observational accessibility and covariance analysis may be considered. The analysis remains highly relevant as costly future observing assets are considered for the cislunar regime, and information-based metrics of architecture utility need to be considered in addition to observational accessibility. Further analysis could also be performed in the catalog maintenance context applying changes in attitude for space-based sensors as a cost in search rewards. It would be incredibly interesting to leverage observer attitude rates to benefit tasking, using prior rotation of an observing spacecraft to traverse through measurement space.

Another major research direction to be considered in future work is the incorporation of multiple tasking objectives. Many methods for multi-objective optimization exist within the reinforcement learning literature. Commonly, one may combine these objectives into a single objective function, a process known as scalarization [50]. This approach is problematic in that the resultant solution may be usable, but only considers a subset of all scalarizations. In a multi-objective context, it is much more useful, especially for human operators, to consider the trade space of objectives, well-known as the Pareto front. The POMDP problem may be reformulated for this purpose transitioning from a scalar reward to a vector-valued reward

$\vec{R}$. The value function associated with some policy is then also vector-valued. One may express the Pareto front in this context as the undominated set of policies, within which there exists no other policy with greater value in all objectives [50]. Generally, the Monte Carlo Tree Search-based extensions to multi-objective reinforcement learning consider expansion based on either hypervolume of the reward vector [109] or Pareto dominance [22]. It would be quite interesting to apply such methodologies to the SDA sensor tasking problem, but several objectives may be challenging to formulate within this context.

Many of the rewards discussed in this dissertation remain applicable. A successful multi-objective MCTS algorithm shall detect new objects via pure search and cue sensors for follow-up observation, all the while maintaining existing estimates. Rewards for each objective are likely at separate scales, but the discussed Monte Carlo Tree Search-based extensions are agnostic to scaling, unlike techniques that rely on a specific scalarization. Notably, establishing a reward scheme for pure search is likely to be challenging, and decision-theoretic pure search for SDA in itself is an avenue of research that could benefit from further study.

The decentralized SDA sensor tasking problem could also be further explored, especially in the context of communication failures. This thesis considers the effects of large communication gaps on decision making, but an equally interesting problem is to study the robustness of decentralized MCTS to increasing failure rates of communication. This is an issue that is much more likely to arise than large communication gaps, and MCTS may be expected to perform well in such scenarios.

Finally, much more work is needed to fully demonstrate MCTS as a tasking solution in real scenarios. Initially, MCTS will be further applied to the VADeR observatory, but in the long term, it would be beneficial to apply MCTS as a collaboration between multiple observing sites, or even utilizing a real space-based sensor.

7.3 Research Impact

The major impact of this research lies in the integration of game-theoretic planning methodologies to space domain awareness. Prior to this thesis, sensor tasking methodologies largely applied dynamic programming techniques, without much consideration for the wealth of recent literature in the sequential

decision making and reinforcement learning communities. The research presented in this thesis outlines means for formulating SDA sensor tasking as a sequential decision making problem in a variety of scenarios, and presents a variety of methods for exploring high-value actions within Monte Carlo Tree Search.

This research is also incredibly impactful in its extension of tasking methodologies to challenging scenarios not previously studied in the literature. Detailed analysis of sensor tasking in the cislunar regime is performed, and few previous methodologies consider catalog maintenance in the context of maneuvering space objects. Extension of novel methodologies to such nonlinear contexts further supports MCTS as a state of the art tasking methodology.

Finally, this dissertation presents an effective solution to the decentralized SDA sensor tasking problem. Decentralization is an emerging need as space object populations quickly grow, yet little literature exists on the subject. This advance ensures that decentralized decision making shall be studied further, and decentralized methods are demonstrated to be incredibly effective for exploration of the high-dimensional solution space inherent to many-agent sensor tasking.

Combined, these contributions extend SDA sensor tasking to approximate optimal solutions over long horizons in a variety of contexts. MCTS-based sensor tasking supports operations in nonlinear, decentralized environments, from catalog maintenance to initial orbit determination. This methodology fundamentally improves the efficiency of high-value observing assets and augments autonomous SDA capabilities.

Bibliography

[1] Louigi Addario-Berry, Borja Balle, and Guillem Perarnau. Diameter and stationary distribution of random r-out digraphs. Electronic Journal of Combinatorics, 27(3):1–41, 2020.

[2] John Africano, Thomas Schildknecht, Mark Matney, Paul Kervin, Eugene Stansbery, and Walter Flury. A Geosynchronous Orbit Search Strategy. Journal of Allergy and Clinical Immunology, 1(2):357–369, 2004.

[3] Christopher Amato and Frans A. Oliehoek. Scalable planning and learning for multiagent POMDPs. Proceedings of the National Conference on Artificial Intelligence, 3:1995–2002, 2015.

[4] Arroyo Parejo Carlos Álvaro, Ortiz Noelia Sánchez, González Raúl Domínguez, and Space Deimos. Effect of Mega Constellations on Collision Risk in Space. 8th European Conference on Space Debris, (April):20–23, 2021.

[5] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. Machine Learning, 47:235–256, 2002.

[6] David Auger, Adrien Couëtoux, and Olivier Teytaud. Continuous Upper Confidence Trees with Polynomial Exploration – Consistency. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pages 194–209, 2013.

[7] Nicola Baresi, Zubin P. Olikara, and Daniel J. Scheeres. Fully Numerical Methods for Continuing Families of Quasi-Periodic Invariant Tori in Astrodynamics. Journal of the Astronautical Sciences, 65(2):157–182, 2018.

[8] David M. Beazley. SWIG: An easy to use tool for integrating scripting languages with C and C++. 4th Annual USENIX Tcl/Tk Workshop 1996, TCL/TK 1996, (July):1–16, 1996.

[9] Richard Bellman. A Markovian Decision Process. Journal of Mathematics and Mechanics, 6(5), 1957.

[10] E. Bertin and S. Arnouts. SExtractor: Software for source extraction. Astronomy and Astrophysics Supplement Series, 117(2):393–404, 1996.

[11] Graeme Best, Oliver M. Cliff, Timothy Patten, Ramgopal R. Mettu, and Robert Fitch. Dec-MCTS: Decentralized planning for multi-robot active perception. International Journal of Robotics Research, 38(2-3):316–337, 2019.

[12] Graeme Best, Oliver M. Cliff, Timothy Patten, Ramgopal R. Mettu, and Robert Fitch. Decentralised Monte Carlo Tree Search for Active Perception. Springer Proceedings in Advanced Robotics, 13:864–879, 2020.

[13] Samuel S. Blackman. Multiple hypothesis tracking for multiple target tracking. IEEE Aerospace and Electronic Systems Magazine, 19(1 II):5–18, 2004.

[14] Marianne R Bobskill. The Role of Cis-lunar Space in Future Global Space Exploration. Global Space Exploration Conference, (1):1–15, 2012.

[15] B Bollobas and W. Fernandez de la Vega. The Diameter of Random Regular Graphs. Combinatorica, 2(September 1981):125–134, 1982.

[16] Stephanne Boucheron, Gabor Lugosi, and Olivier Bousquet. Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford Scholarship Online, 2013.

[17] J. Breakwell and J. Brown. The 'halo' family of 3-dimensional periodic orbits in the restricted3-body problem. 20:389–404, 1976.

[18] Han Cai, Steve Gehly, Yang Yang, Reza Hoseinnezhad, Robert Norman, and Kefei Zhang. Multisensor tasking using analytical Rényi divergence in labeled multi-Bernoulli filtering. Journal of Guidance, Control, and Dynamics, 42(9):2078–2085, 2019.

[19] R Cardin, D Burchett, and H G Reed. SNARE (Sensor Network Autonomous Resilient Extensible): Decentralized Sensor Tasking Improves SDA Tactical Relevance. 2021.

[20] Y. T. Chan, A. G.C. Hu, and J. B. Plant. A Kalman Filter Based Tracking Scheme with Input Estimation. IEEE Transactions on Aerospace and Electronic Systems, AES-15(2):237–244, 1979.

[21] Hyeong Soo Chang, Michael C. Fu, Jiaqiao Hu, and Steven I. Marcus. An adaptive sampling algorithm for solving Markov decision processes. Operations Research, 53(1):126–139, 2005.

[22] Weizhe Chen and Lantao Liu. Pareto Monte Carlo Tree Search for Multi-Objective Informative Planning. 2019.

[23] Shushman Choudhury, Jayesh K. Gupta, Peter Morales, and Mykel J. Kochenderfer. Scalable Online Planning for Multi-Agent MDPs. Journal of Artificial Intelligence Research, 73:821–846, 2022.

[24] Ryan D. Coder and Marcus J. Holzinger. Multi-objective design of optical systems for space situational awareness. Acta Astronautica, 128:669–684, 2016.

[25] Adrien Couetoux, Jean-baptiste Hoock, Nataliya Sokolovska, and Olivier Teytaud. Continuous Upper Confidence Trees. Learning and Intelligent Optimization, (section 3):433–445, 2011.

[26] Rémi Coulom. Efficient Selectivity and Backup Operators in Monte-Carlo Tree Search. 5th International Conference on Computer and Games, 2006.

[27] John P. Cunningham, Philipp Hennig, and Simon Lacoste-Julien. Gaussian Probabilities and Expectation Propagation. 2:1–56, 2011.

[28] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2):182–197, 2002.

[29] Kyle J. DeMars and Moriba K. Jah. Probabilistic initial orbit determination using Gaussian mixture models. Journal of Guidance, Control, and Dynamics, 36(5):1324–1335, 2013.

[30] R. Scott Erwin, Paul Albuquerque, Sudharman K. Jayaweera, and Islam Hussein. Dynamic sensor tasking for space situational awareness. Proceedings of the 2010 American Control Conference, ACC 2010, pages 1153–1158, 2010.

[31] ESA Space Debris Office. ESA's Annual Space Environment Report. (July):1–78, 2019.

[32] Samuel Fedeler, Marcus Holzinger, and William Whitacre. Sensor tasking in the cislunar regime using Monte Carlo Tree Search. Advances in Space Research, (xxxx):1–19, 2022.

[33] Samuel J Fedeler, Marcus J Holzinger, and William Whitacre. Optimality and Application of Tree Search Methods for POMDP-based Sensor Tasking. In Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference, number 1, 2020.

[34] Samuel J Fedeler, Marcus J. Holzinger, and William Whitacre. Cislunar Space Object Tracking Considering Maneuver Estimation and Maneuver Utility. Journal of Guidance, Control, and Dynamics (submitted), pages 1–25, 2023.

[35] Jakob N. Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, pages 2974–2982, 2018.

[36] D. Fränken, M. Schmidt, and M. Ulmke. "Spooky action at a distance" in the cardinalized probability hypothesis density filter. IEEE Transactions on Aerospace and Electronic Systems, 45(4):1657–1664, 2009.

[37] Alan Frieze and Michał Karoński. Introduction to Random Graphs. Introduction to Random Graphs, 2015.

[38] Carolin Frueh. Sensor Tasking for Multi-Sensor Space Object Surveillance. 7th European Conference on Space Debris, Darmstadt, 7(533):1–8, 2017.

[39] Carolin Frueh, Hauke Fielder, and Johannes Herzog. Heuristic and optimized sensor tasking observation strategies with exemplification for geosynchronous objects. Journal of Guidance, Control, and Dynamics, 41(5):1036–1048, 2018.

[40] K. Fujimoto and K. T. Alfriend. Optical short-arc association hypothesis gating via angle-rate information. Journal of Guidance, Control, and Dynamics, 38(9):1602–1613, 2015.

[41] Aurélien Garivier and Eric Moulines. On upper-confidence bound policies for switching bandit problems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6925 LNAI:174–188, 2011.

[42] Steven Gehly, Brandon A. Jones, and Penina Axelrad. Search-Detect-Track Sensor Allocation for Geosynchronous Space Objects. IEEE Transactions on Aerospace and Electronic Systems, 54(6):2788–2808, 2018.

[43] Alan Genz and Giang Trinh. Numerical computation of multivariate normal probabilities using bivariate conditioning. Journal of Computational and Graphical Statistics, (1):141–149, 1992.

[44] Gary M. Goff, Jonathan T. Black, Joseph A. Beck, and Joshuah Hess. A dynamic sensor tasking strategy for tracking maneuvering spacecraft using multiple models. 2016 AIAA Guidance, Navigation, and Control Conference, (January), 2016.

[45] Stephen Gorove. The Geostationary Orbit: Issues of Law and Policy. American Journal of International Law, 73(3):444–461, 1979.

[46] Jesse A. Greaves and Daniel J. Scheeres. Observation and Maneuver Detection for Cislunar Vehicles: Using Optical Measurements and the Optimal Control Based Estimator. Journal of the Astronautical Sciences, 68(4):826–854, 2021.

[47] Matthew J. Gualdoni and Kyle J. DeMars. Impartial Sensor Tasking via Forecasted Information Content Quantification. Journal of Guidance, Control, and Dynamics, 43(11):1–15, 2020.

[48] Davide Guzzetti, Emily M. Zimovan, Kathleen C. Howell, and Diane C. Davis. Stationkeeping analysis for spacecraft in lunar near rectilinear halo orbits. Advances in the Astronautical Sciences, 160:3199–3218, 2017.

[49] Frank Havlak and Mark Campbell. Discrete and continuous, probabilistic anticipation for autonomous robots in Urban environments. IEEE Transactions on Robotics, 30(2):461–474, 2014.

[50] Conor F. Hayes, Roxana Rădulescu, Eugenio Bargiacchi, Johan Källström, Matthew Macfarlane, Mathieu Reymond, Timothy Verstraeten, Luisa M. Zintgraf, Richard Dazeley, Fredrik Heintz, Enda Howley, Athirai A. Irissappane, Patrick Mannion, Ann Nowé, Gabriel Ramos, Marcello Restelli, Peter Vamplew, and Diederik M. Roijers. A Practical Guide to Multi-Objective Reinforcement Learning and Planning. 2021.

[51] Keric Hill, Paul Sydney, Randy Cortez, Kris Hamada, Daron Nishimoto, Pacific Defense Solutions, N Holopono St, Kim Luu, and Paul W Schumacher. Dynamic Tasking of Networked Sensors Using Covariance Information. Advanced for Maui Optical and Space Surveillance Technologies Conference, (Scenario 1), 2010.

[52] Keric Hill, Paul Sydney, Kris Hamada, Randy Cortez, Kim Luu, Moriba Jah, Paul W. Schumacher, Michael Coulman, Jeff Houchard, and Dale Naho'olewa. Covariance-based network tasking of optical sensors. Advances in the Astronautical Sciences, 136:769–786, 2010.

[53] Tyler A. Hobson and I. Vaughan L. Clarkson. A particle-based search strategy for improved Space Situational Awareness. Conference Record - Asilomar Conference on Signals, Systems and Computers, pages 898–902, 2013.

[54] M J Holzinger, C C Chow, and P Garretson. A Primer on Cislunar Space. pages 1–23, 2021.

[55] M. J. Holzinger and M. K. Jah. Challenges and potential in space domain awareness. Journal of Guidance, Control, and Dynamics, 41(1):15–18, 2018.

[56] Marcus J. Holzinger, Daniel J. Scheeres, and Kyle T. Alfriend. Object correlation, maneuver detection, and characterization using control-distance metrics. Journal of Guidance, Control, and Dynamics, 35(4):1312–1325, 2012.

[57] Junling Hu and Michael P. Wellman. Multiagent Reinforcement Learning: Theoretical Framework and an Algorithm. ICML, 98:242–250, 1998.

[58] Andris D. Jaunzemis, Marcus J. Holzinger, and K. Kim Luu. Sensor tasking for spacecraft custody maintenance and anomaly detection using evidential reasoning. Journal of Aerospace Information Systems, 15(3):131–156, 2018.

[59] Simon J Julier and Jeffrey K Uhlmann. New extension of the Kalman filter to nonlinear systems. Proceedings of SPIE, (7), 1997.

[60] Woojun Kim, Myungsik Cho, and Youngchul Sung. Message-Dropout: An Efficient Training Method for Multi-Agent Deep Reinforcement Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01):6079–6086, 2019.

[61] Michael Klonowski, Marcus J. Holzinger, and Naomi Owens Fahrner. Optimal Cislunar Architecture Design Using Monte Carlo Tree Search Methods Michael Klonowski The University of Colorado at Boulder Marcus J . Holzinger The University of Colorado at Boulder Naomi Owens Fahrner. In Advanced for Maui Optical and Space Surveillance Technologies Conference, 2022.

[62] Mykel J. Kochenderfer, Christopher Amato, Girish Chowdhary, Jonathan P. How, Hayley J. Davison Reynolds, Jason R. Thornton, Pedro A. Torres-Carrasquillo, N. Kemal Üre, and John Vian. Decision Making Under Uncertainty: Theory and Application. page 352, 2015.

[63] Levente Kocsis and Csaba Szepesvari. Bandit based Monte-Carlo Planning. Lecture Notes in Computer Science, 4212:282–293, 2006.

[64] Levente Kocsis, Csaba Szepesvári, and Jan Willemson. Improved Monte-Carlo Search. White paper, (1):22, 2006.

[65] Dustin Lang, David W. Hogg, Keir Mierle, Michael Blanton, and Sam Roweis. Astrometry.net: Blind astrometric calibration of arbitrary astronomical images. Astronomical Journal, 139(5):1782–1800, 2010.

[66] M. Levesque and S. Buteau. Image processing technique for automatic detection of satellite streaks. (February):60, 2007.

[67] Michael H. Lim, Claire J. Tomlin, and Zachary N. Sunberg. Sparse tree search optimality guarantees in POMDPs with continuous observation spaces. 2019.

[68] Yixuan Lin, Zhuoran Yang, Kaiqing Zhang, Zhaoran Wang, Tamer Basar, Romeil Sandhu, and Ji Liu. A Communication-Efficient Multi-Agent Actor-Critic Algorithm for Distributed Reinforcement Learning. In IEEE Conference on Decision and Control, 2019.

[69] Richard Linares and Roberto Furfaro. Dynamic Sensor Tasking for Space Situational Awareness via Reinforcement Learning. Advanced Maui Optical and Space Surveillance Technologies Conference, pages 1–10, 2016.

[70] Richard Linares and Roberto Furfaro. An Autonomous Sensor Tasking Approach for Large Scale Space Object Cataloging. Advanced Maui Optical and Space Surveillance Technologies Conference, pages 1–17, 2017.

[71] Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. Morgan Kaufmann Publishers, Inc., 1994.

[72] Michael L. Littman, Anthony R. Cassandra, and Leslie Pack Kaelbling. Learning policies for partially observable environments: Scaling up. Machine Learning Proceedings 1995, pages 362–370, 1995.

[73] D P Lubey and D J Scheeres. Towards Real-Time Maneuver Detection: Automatic State and Dynamics Estimation with the Adaptive Optimal Control Based Estimator. Amos, 2015.

[74] Daniel P. Lubey and Daniel J. Scheeres. Identifying and estimating mismodeled dynamics via optimal control policies and distance metrics. Journal of Guidance, Control, and Dynamics, 37(5):1512–1523, 2014.

[75] Luc Maisonobe, Véronique Pommier-Maurussane, and Pascal Parraud. Orekit: an Open-source Library for Operational Flight Dynamics Applications. 4th International Conference on Astrodynamics Tools and Techniques, (February 2021):5, 2010.

[76] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan. Interacting Multiple Model Methods in Target Tracking : A Survey. IEEE Transactions on Aerospace and Electronic Systems, 34(1), 1998.

[77] W. J. Merline and Steve B. Howell. A realistic model for point-sources imaged on array detectors: The model and initial results. Experimental Astronomy, 6(1-2):163–210, 1995.

[78] A. Milani, M. E. Sansaturio, G. Tommei, O. Arratia, and S. R. Chesley. Multiple solutions for asteroid orbits: Computational procedure and applications. Astronomy & Astrophysics, 431(2):729–746, 2005.

[79] Vivek Muralidharan and Kathleen C. Howell. Leveraging stretching directions for stationkeeping in Earth-Moon halo orbits, 2022.

[80] Timothy S. Murphy and Marcus J. Holzinger. Generalized Minimum-Time Follow-up Approaches Applied to Tasking Electro-Optical Sensor Tasking. Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference, XX(X):1–33, 2017.

[81] Timothy S. Murphy and Marcus J. Holzinger. Generalized Minimum-Time Follow-up Approaches Applied to Tasking Electro-Optical Sensor Tasking. Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference, 2017.

[82] Joelle Pineau, Geoff Gordon, and Sebastian Thrun. Point-based value iteration: An anytime algorithm for POMDPs. IJCAI International Joint Conference on Artificial Intelligence, pages 1025–1030, 2003.

[83] James Rowland, Darren Mcknight, Bonnie Prado Pino, Benedikt Reihs, and Matthew A Stevenson. A Worldwide Network of Radars for Space Domain Awareness in Low Earth Orbit. Amos, 2021.

[84] Andrew R. Runnalls. Kullback-Leibler approach to Gaussian mixture reduction. IEEE Transactions on Aerospace and Electronic Systems, 43(3):989–999, 2007.

[85] Paul A. Samuelson. How Deviant Can You Be? Journal of the American Statistical Association, 63(324):1522–1525, 1968.

[86] Simo Särkkä. Unscented Rauch-Tung-Striebel smoother. IEEE Transactions on Automatic Control, 53(3):845–849, 2008.

[87] Hanspeter Schaub and John Junkins. Analytical Mechanics of Space Systems. 2009.

[88] Thomas Schildknecht. Optical surveys for space debris. Astronomy and Astrophysics Review, 14(1):41–111, 2007.

[89] Paul W. Schumacher, John A. Gaebler, Christopher W.T. Roscoe, Matthew P. Wilkins, and Penina Axelrad. Parallel initial orbit determination using angles-only observation pairs. Celestial Mechanics and Dynamical Astronomy, 130(9):1–20, 2018.

[90] Devavrat Shah, Qiaomin Xie, and Zhi Xu. Non-Asymptotic Analysis of Monte Carlo Tree Search. Operations Research, 70(6):3234–3260, 2022.

[91] Peng Mun Siew, Daniel Jang, Thomas G. Roberts, and Richard Linares. Space-Based Sensor Tasking Using Deep Reinforcement Learning, volume 69. Springer US, 2022.

[92] David Silver and Joel Veness. Monte-Carlo Planning in Large POMDPs. Advances in neural information processing systems (NIPS)., pages 1–9, 2010.

[93] C. Simó, G. Gómez, J. Llibre, R. Martínez, and J. Rodríguez. On the optimal station keeping control of halo orbits. Acta Astronautica, 15(6-7):391–397, 1987.

[94] Dan Simon. Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches. 2006.

[95] M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock. The Two Micron All Sky Survey (2MASS). The Astronomical Journal, 131(2):1163–1183, 2006.

[96] Marshall Smith, Douglas Craig, Nicole Herrmann, Erin Mahoney, Jonathan Krezel, Nate McIntyre, and Kandyce Goodliff. The Artemis Program: An Overview of NASA's Activities to Return Humans to the Moon. IEEE Aerospace Conference Proceedings, pages 1–10, 2020.

[97] R. Sridharan and Antonio F. Pensa. U.S. Space Surveillance Network Capabilities. SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, 3434(November 1998):88–100, 1998.

[98] Jason Stauch, Travis Bessell, Mark Rutten, Jason Baldwin, Moriba Jah, and Keric Hill. Joint Probabilistic Data Association and Smoothing Applied to Multiple Space Object Tracking. Journal of Guidance, Control, and Dynamics, 41(1):1–15, 2017.

[99] Peter B. Stetson. Daophot: a Computer Program for Crowded-Field Stellar Photometry. Publications of the Astronomical Society of the Pacific, 99:191–22, 1987.

[100] Zachary Sunberg, Suman Chakravorty, Richard Scott Erwin, and Senior Member. Information Space Receding Horizon Control for Multisensor Tasking Problems. IEEE Transactions on Cybernetics, 46(6):1325–1336, 2016.

[101] Zachary Sunberg and Mykel J. Kochenderfer. Online algorithms for POMDPs with continuous state, action, and observation spaces. In Twenty-Eighth International Conference on Automated Planning and Scheduling, 2018.

[102] Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. 2018.

[103] Byron D Tapley. Statistical Orbit Determination Theory. 1973.

[104] A Vallenari, A G A Brown, and T Prusti. Gaia Data Release 3: Summary of the content and survey properties. Astronomy & Astrophysics, pages 1–23, 2022.

[105] Jacob K. Vendl and Marcus J. Holzinger. Cislunar Periodic Orbit Analysis for Persistent Space Object Detection Capability. Journal of Spacecraft and Rockets, pages 1–12, 2021.

[106] Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander S. Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom L. Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature, 575(7782):350–354, 2019.

[107] Ba Ngu Vo and Wing Kin Ma. The Gaussian mixture probability hypothesis density filter. IEEE Transactions on Signal Processing, 54(11):4091–4104, 2006.

[108] Eric A Wan, Rudolph Van Der Merwe, and N W Walker Rd. The Unscented Kalman Filter for Nonlinear Estimation.

[109] Weijia Wang and Michèle Sebag. Multi-objective monte-carlo tree search. Journal of Machine Learning Research, 25:507–522, 2012.

[110] Yizao Wang, Jean Yves Audibert, and Rémi Munos. Algorithms for infinitely many-armed bandits. Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference, pages 1729–1736, 2009.

[111] Patrick S. Williams, David B. Spencer, and Richard S. Erwin. Coupling of estimation and sensor tasking applied to satellite tracking. Journal of Guidance, Control, and Dynamics, 36(4):993–1007, 2013.

[112] Sam Wishnek, Marcus J Holzinger, and Patrick Handley. Robust Cislunar Initial Orbit Determination. In Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference, 2021.

[113] Samuel Wishnek and Marcus J. Holzinger. Astrometry and time-resolved photometry from streaks using calibrated ultra-wide field of view cameras. AIAA Scitech 2020 Forum, 1 PartF(January):1–16, 2020.

[114] N.C. Wormald. Models of random regular graphs (lecture notes). London Mathematical Society Lecture Notes, pages 1–60, 1999.

[115] Johnny L. Worthy and Marcus J. Holzinger. Incorporating Uncertainty in Admissible Regions for Uncorrelated Detections. Journal of Guidance, Control, and Dynamics, 38(9):1673–1689, 2015.

[116] Johnny L. Worthy and Marcus J. Holzinger. Use of uninformative priors to initialize state estimation for dynamical systems. Advances in Space Research, 60(7):1373–1388, 2017.

[117] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms. Studies in Systems, Decision and Control, 325:321–384, 2021.

[118] David Zuehlke, Taylor Yow, Daniel Posada, Joseph Nicolich, Christopher W. Hays, Aryslan Malik, and Troy Henderson. Initial Orbit Determination for the CR3BP using Particle Swarm Optimization. pages 1–20, 2022.

Appendix A

Asymptotic Analysis of MCTS with Double Progressive Widening and Polynomial Exploration

A.1 Deterministic Observation Node Likelihood Sampling

Consider a general observation likelihood ω_i, the joint probability of a set of independent measurements Y_i given a prior state estimate X such that

$$\omega_i = p(Y_i, X) = \prod_{j=1}^{|Y_i|} p(\vec{y_j}, X)$$

When observation widening does not occur, one must traverse to a previously generated observation node associated with that action when an action is selected. First consider a general sampling scheme assuming that whenever a traversal is taken from a decision node w to an observation node i, random sampling occurs in proportion to observation likelihood

$$p(i)_0 = \frac{\omega_i}{\sum_{j=1}^{|w|} \omega_j}.$$

Because of widening methods, this sampling methodology leads to a bias towards measurements generated early in the simulation process. As such, a modified form may be introduced such that

$$p(i)_{corr} = p(i)_0 \frac{p(i)_0 N}{N(i)},$$

where N describes the number of visits to the parent decision node and observation node i, respectively. Normalizing this result across all child nodes leads to

$$p(i) = \frac{p(i)_{corr}}{\sum_{j=1}^{|w|} p(j)_{corr}} = \frac{\frac{\omega_i^2 N}{(\sum_{j=1}^{|w|} \omega_j)^2 N(i)}}{\sum_{j=1}^{|w|} \frac{\omega_j^2 N}{(\sum_{k=1}^{|w|} \omega_k)^2 N(j)}} = \frac{\frac{\omega_i^2}{N(i)}}{\sum_{j=1}^{|w|} \frac{\omega_j^2}{N(j)}}.$$

Applying this weighting scheme, sampling will converge to a likelihood-proportional result given infinite simulation, but further refinement is possible. As observational nodes are sampled in this manner, the ratio between the weight and the number of visits to a child observation node converges as

$$\frac{\omega_i}{N(i)} \approx \frac{1}{N}$$

If an observation node is undersampled, this ratio becomes larger than $\frac{1}{N}$. Exploiting this behavior, observation node selection can be made purely deterministic, where the next sampled observation maximizes the criterion

$$o = \underset{i}{\operatorname{argmax}} \frac{\omega_i}{N(i)}.$$

Using deterministic sampling, the number of visits to each child node may then be upper and lower bounded by

$$N(i) \geq \frac{\omega_i N^2}{N + |w| - 1},$$

$$N(i) \leq \omega_i N + 1.$$

A.2 Consistency of Observation Nodes

The key goal in evaluating the consistency of the value function estimate from observation nodes to decision nodes is determination of the desired widening parameter α_o. Assuming this error is composed by error at the child nodes, estimation error at the parent decision node w may be formulated as a function of the associated observation nodes

$$|V(w) - V^*(w)| = \left| \sum_{i=1}^{|w|} \frac{N(i)}{N} V(i) - V^*(w) \right|. \tag{A.1}$$

Applying the triangle inequality, Equation A.1 may be related to estimation error at each observation node as

$$|V(w) - V^*(w)| \leq \left| \sum_{i=1}^{|w|} \frac{N(i)}{N} (V(i) - V^*(i)) \right| + \left| \sum_{i=1}^{|w|} \frac{N(i)}{N} (V^*(i) - V^*(w)) \right|. \tag{A.2}$$

Considering the first term of this result, note that the number of visits $N(i)$ to node i is lower bounded by Equation 3.4, and upper bounded by Equation 3.5. The consistency definition of Equation 3.2 also applies. As such an upper bound to this term can be expressed as

$$\begin{aligned}\left|\sum_{i=1}^{|w|}\frac{N(i)}{N}(V(i)-V^*(i))\right| &\leq \left|(\sum_{i=1}^{|w|}\omega_i+\frac{1}{N})(C_dN(i)^{-\mu_d})\right| \\ &\leq \left|\sum_{i=1}^{|w|}(\omega_i+\frac{1}{N})C_d(\frac{\omega_iN^2}{N+|w|-1})^{-\mu_d})\right|\end{aligned}$$

When the number of child nodes at w is small relative to N, this bound is reduced to

$$\left|\sum_{i=1}^{|w|}\frac{N(i)}{N}(V(i)-V^*(i))\right| \leq \left|\sum_{i=1}^{|w|}(\omega_i+\frac{1}{N})C_d(\omega_iN)^{-\mu_d})\right|. \tag{A.3}$$

To express this result as a function of the desired widening parameter α_o, we consider the expected error is solely a function of a single random variable, the expected weight

$$E[\omega_i]=\frac{E[p(Y_i,X)]}{\sum_{j=1}^{|w|}E[p(Y_j,X)]}=\frac{1}{|w|}=\frac{1}{\lfloor N^\alpha\rfloor}.$$

To further bound the error expressions, it is also critical to consider the variance of expected error, and therefore, the second-order behavior of observation node weights. We demonstrate this evaluation assuming that the joint probabilities for state estimates and measurements are Gaussian random variables associated with a single measurement, but the structure of these arguments holds for an arbitrary distribution. With these assumptions, the second moment of observation node likelihoods is expressed as

$$\begin{aligned}E[\omega_i^2] &= \frac{E[p(Y_i,X)^2]}{|w|E[p(Y_j,X)^2]+(|w|^2-|w|)E[p(Y_j,X)]^2} \\ &= \frac{1}{|w|}\frac{(2\pi)^{-k}|\Sigma|^{-1}}{(2\pi)^{-k}|\Sigma|^{-1}+3^{\frac{k}{2}}(|w|-1)(2^{-2k}\pi^{-k}|\Sigma|^{-1})}\end{aligned}$$

$$E[\omega_i^2]=\frac{1}{|w|}\frac{1}{1+3^{\frac{k}{2}}(|w|-1)2^{-k}}. \tag{A.4}$$

Note that this result is reached using the expectation operator on

$$E[p(Y_i, X)] = \int_{-\infty}^{\infty} \left(\frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (Y_i - \vec{h}_i(X))^T \Sigma^{-1} (Y_i - \vec{h}_i(X)) \right) \right)^2 dX,$$

$$E[p(Y_i, X)^2] = \int_{-\infty}^{\infty} \left(\frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (Y_i - \vec{h}_i(X))^T \Sigma^{-1} (Y_i - \vec{h}_i(X)) \right) \right)^3 dX,$$

$$\Sigma \approx H_i P_i H_i^T + R_i.$$

These expressions are unnormalized Gaussians with an adjusted variance such that

$$\begin{aligned} E[p(Y_i, X)] &= \frac{1}{(2\pi)^k |\Sigma|} \int_{-\infty}^{\infty} \exp\left(-(Y_i - \vec{h}_i(X))^T \Sigma^{-1} (Y_i - \vec{h}_i(X)) \right) dX \\ &= 2^{-k} \pi^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} \end{aligned}$$

and similarly

$$\begin{aligned} &E[p(Y_i, X)^2] = \\ &\frac{1}{(2\pi)^{\frac{3k}{2}} |\Sigma|^{\frac{3}{2}}} \int_{-\infty}^{\infty} \exp\left(-\frac{3}{2} (Y_i - \vec{h}_i(X))^T \Sigma^{-1} (Y_i - \vec{h}_i(X)) \right) dX \\ &= (2\pi)^{-k} 3^{-\frac{k}{2}} |\Sigma|^{-1} \end{aligned}$$

Equation A.4 is expressed from these results, and likelihood sample variance can then be computed as

$$var(\omega_i) = E[\omega_i^2] - E[\omega_i]^2 = \frac{1}{|w|} \frac{1}{1 + 3^{\frac{k}{2}} (|w| - 1) 2^{-k}} - \frac{1}{|w|^2} \tag{A.5}$$

Note that variance follows as $O(|w|^{-2})$ as the number of child nodes grows large. Variance in observation weights is not a function of properties of the decision made, but simply the number of samples generated. That is, these results are explicitly independent of the covariance of the joint distribution Σ. As the sample standard deviation decreases with $\frac{1}{|w|}$ in the same order as the expectation, the following arguments on boundedness hold at the same rates to arbitrary variance.

Returning to Equation A.3, we may express the result as

$$E\left[\left| \sum_{i=1}^{|w|} (\omega_i + \frac{1}{N}) C_d (\omega_i N)^{-\mu_d} \right| \right] = N^{\alpha_o} \left| (E[\omega_i] + \frac{1}{N}) C_d (E[\omega_i] N)^{-\mu_d} \right|$$

$$\approx C_d(1+N^{\alpha_o-1})N^{-(1-\alpha_o)\mu_d} = O(N^{-(1-\alpha_o)\mu_d} + N^{-(1-\alpha_o)(1+\mu_d)}) \tag{A.6}$$

Since $\alpha_o \in (0,1),\ \mu_d \in (0,1)$, error in the first term then decreases with the dominating term $O(N^{-(1-\alpha_o)\mu_d})$.

Considering the second term in Equation A.2, using Equation 3.5 note the result is upper bounded by

$$\left|\sum_{i=1}^{|w|} \frac{N(i)}{N}\left(V^*(i) - V^*(w)\right)\right| \leq \left|\sum_{i=1}^{w} |(\omega_i + \frac{1}{N})\left(V^*(i) - V^*(w)\right)\right|$$

Again taking the expectation, we can apply Hoeffding's inequality, finding

$$E\left[\left|\sum_{i=1}^{|w|}(\omega_i + \frac{1}{N})(V^*(i) - V^*(w))\right|\right] = \tag{A.7}$$

$$\left|\sum_{i=1}^{|w|}(N^{-\alpha_o} + N^{-1})\left(V^*(i) - V^*(w)\right)\right| \leq t \tag{A.8}$$

with probability at least

$$1 - 2\exp\left(-\frac{2t^2}{N^{\alpha_o}(N^{-\alpha_o})^2}\right).$$

This result must also be constrained as strictly less than or equal to the dominating term such that $t \leq N^{-(1-\alpha)\mu_d}$. One may then also assume this result is exponentially sure in t with $\eta = 1$, such that this probability grows as a function in t. With the defined convergence rate η, one then finds

$$1 - 2\exp\left(-\frac{2t^2}{N^{\alpha_o}(N^{-\alpha_o})^2}\right) = 1 - 2\exp\left(-2t^{-1}\right),$$

$$t^{-3} = N^{\alpha_o},$$

$$N^{3(1-\alpha_o)\mu_d} = N^{\alpha_o},$$

$$\alpha_o = \frac{3\mu_d}{1+3\mu_d}. \tag{A.9}$$

This result may then be substituted back into the dominating term $O(N^{-(1-\alpha_o)\mu_d})$. Using $\alpha_o = \frac{3\mu_d}{1+3\mu_d}$ a minimal exponent

$$\mu_{d-\frac{1}{2}} = \frac{\mu_d}{1+3\mu_d} \tag{A.10}$$

is found, and the nodes are recursively consistent. Note that this result for convergence is equivalent to that for random Markov decision process transitions in [6].

A.3 Alternate proof for Observation Nodes with Guarantees

In many cases, this result may be sufficient, but it does rely on assumptions on the variance of the random variables considered. As the likelihood weights that challenge this derivation can be considered as a set of realizations, the application of Samuelson's inequality [85]. can also be considered, in which upper and lower bounds for a set of samples can be expressed as

$$\bar{X} - \sigma\sqrt{N-1} \leq X_j \leq \bar{X} + \sigma\sqrt{N-1}\ \forall X_j \in X \tag{A.11}$$

where $\bar{X}$ is the sample mean, σ is the sample standard deviation, and N samples are taken.

Applied to the likelihood weights, we find

$$\begin{aligned}\omega_i &\leq \frac{1}{|w|} + \sqrt{|w|-1}var(\omega_i) \\ &\leq \frac{1}{|w|} + \frac{|w|-1}{|w|}(\frac{1-a}{1-a+a|w|})^{\frac{1}{2}} \approx \frac{1}{|w|} + (\frac{1-a}{a|w|})^{\frac{1}{2}}\end{aligned}$$

as $a|w| >> 1-a$, where $a = 3^{\frac{k}{2}}2^{-k}$ and k is the state space dimension. Similarly applying the lower bound

$$\frac{1}{|w|} - (\frac{1-a}{a|w|})^{\frac{1}{2}} \leq \omega_i \leq \frac{1}{|w|} + (\frac{1-a}{a|w|})^{\frac{1}{2}}$$

Applying these new guaranteed bounds, we return to Equation A.3 with

$$\left|\sum_{i=1}^{|w|}\frac{N(i)}{N}(V(i)-V^*(i))\right| \leq$$
$$\left|\sum_{i=1}^{|w|}\left(\frac{1}{|w|} + (\frac{1-a}{a|w|})^{\frac{1}{2}} + \frac{1}{N}\right)C_d\left((\frac{1}{|w|} - (\frac{1-a}{a|w|})^{\frac{1}{2}})N\right)^{-\mu_d}\right|$$

Noting $|w| = \lfloor N^{\alpha_o} \rfloor$, the result then has runtimes on the order

$$\begin{aligned}O(&N^{-\mu_d(1-\alpha_o)} + N^{-\mu_d(1-\frac{\alpha_o}{2})} + N^{\frac{\alpha_o}{2}-\mu_d(1-\alpha_o)} \\ &+ N^{\frac{\alpha_o}{2}-\mu_d(1-\frac{\alpha_o}{2})} + N^{-(1+\mu_d)(1-\alpha_o)} + N^{-1-\alpha_o-\mu_d(1-\frac{\alpha_o}{2})})\end{aligned} \tag{A.12}$$

With $\alpha_o, \mu_d \in [0,1)$, this term is then dominated with runtime $O(N^{\frac{\alpha_o}{2}-\mu_d(1-\alpha_o)})$. Restricting this exponent to be negative, we find

$$\alpha_o < \frac{\mu_d}{\frac{1}{2}+\mu_d} \tag{A.13}$$

Now taking the second term of Equation A.2, we then reapply Hoeffding's inequality.

$$\left|\sum_{i=1}^{|w|}\frac{N(i)}{N}(V^*(i)-V^*(w))\right| \leq \left|\sum_{i=1}^{|w|}(\omega_i+\frac{1}{N})(V^*(i)-V^*(w))\right|$$
$$\leq N^{\alpha_o}\left|(\frac{1}{|w|}+(\frac{1-a}{a|w|})^{\frac{1}{2}}+\frac{1}{N})(V^*(i)-V^*(w))\right| \leq t$$

with probability at least (dropping the small N^{-1} term)

$$1-2\exp\left(-2t^2N^{-\alpha_o}(N^{-\alpha_o}+bN^{-\frac{\alpha_o}{2}})^{-2}\right)$$
$$\approx 1-2\exp\left(-2\frac{t^2}{b^2+2bN^{-\frac{\alpha_o}{2}}}\right)$$

for $b=(\frac{1-a}{a})^{\frac{1}{2}}$. One may then force this probability to converge exponentially surely as $O(Ct^{-1})$. It is also critical to ensure this probability converges at least as fast as the dominating term with $t^{-1} \leq N^{\frac{\alpha_o}{2}-\mu_d(1-\alpha_o)}$. Then, by matching terms in t and N, we find the relation

$$1-2\exp\left(-2\frac{t^2}{b^2+2bN^{-\frac{\alpha_o}{2}}}\right)=1-2\exp\left(Ct^{-1}\right)$$
$$t^3=N^{\frac{\alpha_o}{2}}$$
$$N^{-\frac{3\alpha_o}{2}+3\mu_d(1-\alpha_o)}=N^{\frac{\alpha_o}{2}}$$
$$\alpha_o(2+3\mu_d)=3\mu_d$$

$$\alpha_o=\frac{3\mu_d}{2+3\mu_d} \tag{A.14}$$

Reapplying this to the maximum exponent, we find convergence rates of

$$\mu_{d-\frac{1}{2}}=-\left(\frac{\alpha}{2}-\mu_d(1-\alpha)\right)$$

$$\mu_{d-\frac{1}{2}}=\frac{\mu_d}{4+6\mu_d} \tag{A.15}$$

on guaranteed bounds for error in estimating the value function across observational nodes.

A.4 Consistency of Decision Nodes

First, this proof relies on establishing an exploration function f that ensures decision nodes are selected infinitely often given infinite simulation. Lemma 3 of [6] holds. For an arbitrary non-decreasing map f from

$\Re^1$ to $\Re^1$, a score function may be computed at observation node z for child decision node i as

$$sc_n(i) = V_n(i) + \sqrt{\frac{f(N)}{N(i)}}. \tag{A.16}$$

All children must be selected infinitely often provided that $\lim_{+\infty} f = +\infty$. In particular, bounding behavior on visits can be defined as

$$N(i) \geq \frac{1}{4} \min\left(f(N^{1-\alpha_d}), N^{1-\alpha_d}\right). \tag{A.17}$$

Next, the use of polynomial exploration as in [6] is justified, as compared to methodologies used in [25, 63]. Consider a polynomial exploration function

$$f(N) = N^e \tag{A.18}$$

where $e \in (0, 1)$.

An upper bound on estimation error is determined as in [6] using careful parameter selection

$$V(z) - V(z) \leq (1 + C_{d-\frac{1}{2}}) N^{-\mu_{d-\frac{1}{2}} \frac{1-\alpha_d}{1+\mu_{d-\frac{1}{2}}}} \tag{A.19}$$

for observation node z. One must then determine a fixed coefficient μ_{d-1} such that all child decision nodes w of z verify exponentially surely

$$|V(z) - V^*(z)| \leq C_{d-1} N^{-\mu_{d-1}}$$

In order to find a lower bound, the assumptions of Equation 3.3 are followed. An expression for the lower bound is desired as a function of the minimal number of visits described in Equation A.17. As such, we choose the bound Δ to scale in proportion to the minimal number of visits

$$\Delta = \left(\frac{1}{4} min(f(N), N)\right)^{-\mu_{d-\frac{1}{2}}} \tag{A.20}$$

Now consider a time $N^{\xi(1-\alpha_d)}$, applying some positive, bounded coefficient ξ. At this step, knowing the widening coefficient, the number of children of observation node z is at least $\lfloor N^{\xi(1-\alpha_d)\alpha_d} \rfloor$. Using Definition 3, assuming the exploration function is strictly less than N, the probability not a single child node lies within the bound Δ is

$$p_n = (1-\theta\Delta^p)^{\lfloor N^{\xi(1-\alpha_d)\alpha_d} \rfloor}$$

$$\log p_n \approx N^{\xi(1-\alpha_d)\alpha_d} \log(1-\theta\Delta^p)$$

When the exponentially sure component is small ($\theta\Delta^p << 1$) a Taylor series expansion can be applied and

$$\log p_n \approx -4^{\mu_{d-\frac{1}{2}p}} N^{\xi(1-\alpha_d)\alpha_d} \theta f(N^{\xi(1-\alpha_d)})^{-\mu_{d-\frac{1}{2}p}} \tag{A.21}$$

For the proof to proceed, this result must monotonically decrease in N such that

$$p_n(N \to \infty) \to 0$$

It also is not desired for this result to be a function of the undetermined regularity constant p. For this to be the case, f must be a polynomial function in N; applying Equation A.18,

$$\log p_n \approx -4^{\mu_{d-\frac{1}{2}p}} N^{\xi(1-\alpha_d)(\alpha_d - e\mu_{d-\frac{1}{2}p})}$$

The quantity $\alpha_d - e\mu_{d-\frac{1}{2}p}$ is then restricted such that it is only a function of α_d

$$\alpha_d - e\mu_{d-\frac{1}{2}p} > 0$$

$$e = \frac{c\alpha_d}{\mu_{d-\frac{1}{2}p}}$$

$$\alpha_d - e\mu_{d-\frac{1}{2}p} = (1-c)\alpha_d$$

To determine the arbitrary constant c, the log probability not a single node's value function is estimated to be within Δ of the Bellman optimal value is constrained to $O(-C\Delta^{-1})$. Returning to Equation A.21,

$$\Delta^{-1} = N^{(1-c)\xi\alpha_d(1-\alpha_d)}$$

$$N^{e\mu_{d-\frac{1}{2}p\xi(1-\alpha_d)}} = N^{c\xi\alpha_d(1-\alpha_d)} = N^{(1-c)\xi\alpha_d(1-\alpha_d)}$$

$$c = 0.5$$

Then, $e = \frac{\alpha_d}{2\mu_{d-\frac{1}{2}p}}$.

So long as ξ is lower bounded by a constant in the domain $(0,1)$, the estimation error is then lower-bounded

by the terms Δ, $N^{\xi-1}$, and $\frac{N^{\alpha_d+e-1}}{\Delta^2}$ [6].

ξ is a chosen quantity; therefore, the second terms may be bounded as $O(\Delta)$ as follows

$$N^{\xi-1} = C\Delta = C4^{\mu_{d-\frac{1}{2}}} N^{-\mu_{d-\frac{1}{2}}\xi e(1-\alpha)}$$

$$\xi - 1 = -\mu_{d-\frac{1}{2}}\xi e(1-\alpha_d)$$

$$\xi = \frac{1}{1+\mu_{d-\frac{1}{2}}e(1-\alpha_d)} = \frac{p}{p+0.5\alpha_d(1-\alpha_d)} < 1$$

$$N^{\xi-1} = 4^{-\mu_{d-\frac{1}{2}}}\Delta < \Delta = O(\Delta)$$

We now wish to evaluate the third term $\frac{N^{\alpha_d+e-1}}{\Delta^2}$, desiring $N^{\alpha_d+e-1} \leq O(\Delta^3)$. Considering this, term, first ensure that the numerator converges exponentially surely with the constraint

$$\alpha_d + e - 1 = \left(1 + \frac{1}{2p\mu_{d-\frac{1}{2}}}\right)\alpha_d - 1 \leq -\frac{1}{2}$$

$$\alpha_d \leq \frac{p\mu_{d-\frac{1}{2}}}{1+2p\mu_{d-\frac{1}{2}}}$$

This additionally ensures a relation between the widening term α_d and the score function for selection. To reduce constants let $p = 2$ and

$$\alpha_d = \frac{2\mu_{d-\frac{1}{2}}}{1+4\mu_{d-\frac{1}{2}}} \leq \frac{2}{5} \tag{A.22}$$

is found to satisfy the constraint. Then,

$$\log(\Delta^3) = 6\xi e(1-\alpha_d)\mu_{d-\frac{1}{2}} = \frac{3\alpha_d(1-\alpha_d)\mu_{d-\frac{1}{2}}}{2+\frac{1}{2}\alpha_d(1-\alpha_d)} \leq \frac{18}{53}$$

and

$$O(\Delta^3) \geq O(N^{-\frac{18}{53}}) > O(N^{-\frac{1}{2}}) \geq O(N^{\alpha_d+e-1}) \tag{A.23}$$

and the third term must converge faster than $O(\Delta)$. Therefore, the term $O(\Delta)$ lower bounds the estimation error. Substituting into a recursive form,

$$V(z) - V(z) \geq CN^{-\frac{4\alpha_d(1-\alpha_d)}{4+\alpha_d(1-\alpha_d)}}$$

and

$$\mu_{d-1} = \frac{4\alpha_d(1-\alpha_d)}{4+\alpha_d(1-\alpha_d)}. \tag{A.24}$$

Appendix B

Admissible Regions

An admissible region (AR) may be formed as a compact subset of the range range-rate half plane. With no initial assumptions on feasible range range-rate pairs, dynamical constraints, generally with Keplerian assumptions, may be incorporated to reduce the feasible subset in which the target lies. First consider the net information from 1 or more observations $\vec{y}$ of a state $\vec{x}$ from observer state $\vec{o}$. The probabilistic admissible region admits measurement uncertainty [115], but bounds here are established using the observation to outline the determinable subset of the state $\vec{x}_d$, as

$$\vec{y} = \vec{h}\left(\vec{x}; \vec{o}, t\right) \tag{B.1}$$

$$\vec{x}_d = \vec{h}^{-1}\left(\vec{y}; \vec{x}, \vec{o}, t\right) \tag{B.2}$$

$\vec{x}_u$ is denoted the unobservable subset and the goal of this problem is to define constraints on the unobservable subset. These constraints may be dynamical or physical limitations, such as an orbit that is bound to Earth, a radius of periapsis that ensures an object won't collide with Earth, or minimal or maximal eccentricities. Constraints may be generically expressed as

$$g_i\left(\vec{x}_d, \vec{x}_u; \vec{u}, t\right) \leq 0 \tag{B.3}$$

Then, the admissible region may be expressed as the set intersection of the admissible regions formed by each constraint, with

$$\mathcal{AR}_i \in \mathcal{R}^u \tag{B.4}$$

$$\mathcal{AR}_i = \{\vec{x}_u \in \mathcal{R}^u |\ g_i\left(\vec{x}_d, \vec{x}_u; \vec{u}, t\right) \leq 0\} \tag{B.5}$$

$$\mathcal{AR} = \bigcap_{i=1}^{n} \mathcal{AR}_i \tag{B.6}$$

B.1 Admissible Region Constraints

In general, constraints must be expressed in terms of the unobservable subset. For optical observers, the unobservable subset is range and range rate, and it is useful to apply several definitions. First, the full state may be expressed in terms of the observer state and relative range vector $r\vec{h}o$, with

$$\vec{r} = \vec{o} + \vec{\rho} \tag{B.7}$$

$$\dot{\vec{r}} = \dot{\vec{o}} + \dot{\vec{\rho}} \tag{B.8}$$

$$\vec{\rho} = \rho\hat{l} \tag{B.9}$$

$$\dot{\vec{\rho}} = \dot{\rho}\hat{l} + \rho\dot{\hat{l}} \tag{B.10}$$

The line of sight vector may be expressed in terms of measurement information, with

$$\vec{x}_d = \begin{bmatrix} \alpha & \delta & \dot{\alpha} & \dot{\delta} \end{bmatrix}^T \tag{B.11}$$

Then, line of sight rates may be evaluated, with the following definitions.

$$\dot{\hat{l}} = \frac{d\hat{l}}{dt} = \dot{\alpha}\frac{d\hat{l}}{d\alpha} + \dot{\delta}\frac{d\hat{l}}{d\delta} = \dot{\alpha}\hat{l}_\alpha + \dot{\delta}\hat{l}_\delta \tag{B.12}$$

$$\hat{l} = \begin{bmatrix} \cos\delta\cos\alpha \\ \cos\delta\sin\alpha \\ \sin\delta \end{bmatrix} \quad \hat{l}_\alpha = \begin{bmatrix} -\cos\delta\sin\alpha \\ \cos\delta\cos\alpha \\ 0 \end{bmatrix} \quad \hat{l}_\delta = \begin{bmatrix} -\sin\delta\cos\alpha \\ -\sin\delta\sin\alpha \\ \cos\delta \end{bmatrix} \tag{B.13}$$

B.1.1 Energy

Energy is expressed in terms of position and velocity as

$$\mathcal{E} = \frac{\dot{\vec{r}} \cdot \dot{\vec{r}}}{2} - \frac{\mu}{(\vec{r} \cdot \vec{r})^{\frac{1}{2}}} = -\frac{\mu}{2a} \tag{B.14}$$

Given prior definitions, the position and velocity vectors may be transformed into functions of $\vec{x}_u$. First,

$$\vec{r} \cdot \vec{r} = \vec{o} \cdot \vec{o} + \rho^2 \hat{l} \cdot \hat{l} + 2\rho \vec{o} \cdot \hat{l} = \rho^2 + \omega_0 \rho + \omega_1. \tag{B.15}$$

where

$$\omega_0 = 2\vec{o} \cdot \hat{l} \tag{B.16}$$

$$\omega_1 = \vec{o} \cdot \vec{o} \tag{B.17}$$

It is useful to note that this resultant form is quadratic in range. The inner product of the velocity vector may similarly be evaluated as

$$\dot{\vec{r}} \cdot \dot{\vec{r}} = \dot{\vec{o}} \cdot \dot{\vec{o}} + \dot{\rho}^2 \hat{l} \cdot \hat{l} + \rho^2 \dot{\hat{l}} \cdot \dot{\hat{l}} + 2\left(\dot{\rho} \dot{\vec{o}} \cdot \hat{l} + \rho \dot{\vec{o}} \cdot \dot{\hat{l}} + \rho \dot{\rho} \hat{l} \cdot \dot{\hat{l}}\right) \tag{B.18}$$

$$= \dot{\rho}^2 + \omega_2 \dot{\rho} + \omega_3 \rho^2 + \omega_4 \rho + \omega_5 \tag{B.19}$$

with terms in ω defined as

$$\omega_2 = 2\dot{\vec{o}} \cdot \hat{l} \tag{B.20}$$

$$\omega_3 = \dot{\hat{l}} \cdot \dot{\hat{l}} \tag{B.21}$$

$$\omega_4 = 2\dot{\vec{o}} \cdot \dot{\hat{l}} \tag{B.22}$$

$$\omega_5 = \dot{\vec{o}} \cdot \dot{\vec{o}} \tag{B.23}$$

Note that ω_3 is not unit, because

$$\dot{\hat{l}} \cdot \dot{\hat{l}} = \dot{\alpha}^2 \cos^2 \delta + \dot{\delta}^2 \tag{B.24}$$

Substituting these expressions into Equation B.14, a quadratic equation in range rate may be expressed, with

$$\dot{\rho}^2 + \omega_2 \dot{\rho} + F(\rho) - 2\mathcal{E} = 0 \tag{B.25}$$

$$F(\rho) = \omega_3 \rho^2 + \omega_4 \rho + \omega_5 - \frac{2\mu}{\sqrt{\rho^2 + \omega_0 \rho + \omega_1}} \tag{B.26}$$

Therefore, for any given energy and range, there must be 0, 1, or 2 range rate solutions that form bounds for the admissible region of range range rate pairs. With a given energy, ranges may be varied, and range rates can be computed to numerically form a boundary that satisfies all orbits where $\mathcal{E} = \mathcal{E}_{max}$. At any given range,

$$\dot{\rho} = -\frac{\omega_2}{2} \pm \sqrt{\left(\frac{\omega_2}{2}\right)^2 - F(\rho) - 2\mathcal{E}_{max}} \tag{B.27}$$

and the maximal range is evaluated as

$$\left(\frac{\omega_2}{2}\right)^2 - F(\rho) - 2\mathcal{E}_{max} = 0 \tag{B.28}$$

B.1.2 Radius of Periapsis and Eccentricity

Admissible regions may similarly be formed from other Keplerian properties. First consider a maximum eccentricity, with eccentricity defined as

$$e = \sqrt{1 + \frac{2\mathcal{E}\vec{h} \cdot \vec{h}}{\mu^2}} \tag{B.29}$$

This can be expressed in energy and angular momentum as

$$2\mathcal{E}\vec{h} \cdot \vec{h} \leq \mu^2 (e_{max}^2 - 1) \tag{B.30}$$

A minimum radius of periapsis may also be applied using the angular momentum vector. Here,

$$r_p = a(1 - e) = \frac{\mu}{2\mathcal{E}} (e - 1) \geq r_{p,min} \tag{B.31}$$

Substituting for eccentricity, a bound is established as

$$\vec{h} \cdot \vec{h} - r_{p,min} \left(2\mu + 2\mathcal{E}\right) \geq 0 \tag{B.32}$$

B.1.3 Angular Momentum-based Admissible Region Constraints

There is a clear need to express the angular momentum vector in terms of the unobservable range and range rate. Explicitly,

$$\vec{h} = \vec{r} \times \dot{\vec{r}} \tag{B.33}$$

$$= \vec{o} \times \dot{\vec{o}} + \dot{\rho}\vec{o} \times \hat{l} + \rho\vec{o} \times \dot{\hat{l}} + \rho\hat{l} \times \dot{\vec{o}} + \rho^2 \hat{l} \times \dot{\hat{l}} \tag{B.34}$$

and the result may be simplified as

$$\vec{h} = \vec{h}_1\dot{\rho} + \vec{h}_2\rho^2 + \vec{h}_3\rho + \vec{h}_4 \tag{B.35}$$

$$\vec{h}_1 = \vec{o} \times \hat{l} \quad \vec{h}_2 = \hat{l} \times \dot{\hat{l}} \quad \vec{h}_3 = \hat{l} \times \dot{\vec{o}} + \vec{o} \times \dot{\hat{l}} \quad \vec{h}_4 = \vec{o} \times \dot{\vec{o}} \tag{B.36}$$

The inner product of the angular momentum vector may also be expressed as quadratic in range rate, with

$$\vec{h} \cdot \vec{h} = c_0\dot{\rho}^2 + P(\rho)\dot{\rho} + U(\rho) \tag{B.37}$$

$$P(\rho) = c_1\rho^2 + c_2\rho + c_3 \tag{B.38}$$

$$U(\rho) = c_4\rho^4 + c_5\rho^3 + c_6\rho^2 + c_7\rho + c_8 \tag{B.39}$$

Each constant can be defined using the components of the angular momentum vector, with

$$c_0 = \vec{h}_1 \cdot \vec{h}_1 \quad c_1 = 2\vec{h}_1 \cdot \vec{h}_2 \quad c_2 = 2\vec{h}_1 \cdot \vec{h}_3 \quad c_3 = 2\vec{h}_1 \cdot \vec{h}_4 \quad c_4 = \vec{h}_2 \cdot \vec{h}_2$$

$$c_5 = 2\vec{h_2} \cdot \vec{h}_3 \quad c_6 = 2\vec{h_2} \cdot \vec{h}_4 + \vec{h}_3 \cdot \vec{h}_3 \quad c_7 = 2\vec{h}_3 \cdot \vec{h}_4 \quad c_8 = \vec{h}_4 \cdot \vec{h}_4$$

B.1.4 Eccentricity

Returning to the established constraint for maximal eccentricity, it is clear that the result can be expected to be quartic in range rate. Substituting the derived expressions, a bound is expressed as

$$a_4\dot{\rho}^4 + a_3\dot{\rho}^3 + a_2\dot{\rho}^2 + a_1\dot{\rho} + a_0 \leq 0 \tag{B.40}$$

where

$$a_4 = c_0 \quad a_3 = P(\rho) + \omega_1 c_0 \quad a_2 = c_0 F(\rho) + \omega_1 P(\rho) + U(\rho)$$

$$a_1 = F(\rho)P(\rho) + \omega_1 U(\rho) \quad a_0 = F(\rho)U(\rho) + \mu^2(1 - e_{max}^2).$$

At any range ρ, this quartic expression may be solved to determine eccentricity bounds. While as many as four solutions may be found, in practice, there are usually 0, 1, or 2. However, this result implies that while the resultant admissible region is a compact set, it is not necessarily a connected set.

B.1.5 Radius of Periapsis

Similar substitutions may be applied for radius of periapsis, but unlike the eccentricity bound, the resulting expression remains quadratic, with

$$q_2\dot{\rho}^2 + q_1\dot{\rho} + q_0 \geq 0 \tag{B.41}$$

$$q_0 = U(\rho) - F(\rho)r_{p,min}^2 - 2\mu r_{p,min} \tag{B.42}$$

$$q_1 = P(\rho) - \omega_1 r_{p,min}^2 \tag{B.43}$$

$$q_2 = (c_0 - r_{p,min}^2). \tag{B.44}$$

B.2 Representing Admissible Regions as Estimates

With established derivations of admissible region bounds, it remains to instantiate the uniform admissible region as a state estimate. A common methodology for this process is utilization of a Gaussian mixture

[29]. Such methods are often computationally practical, especially if large sets of admissible regions are being generated, propagated, and compared. The general strategy is to first consider the problem of approximating a univariate uniform prior with a set of Gaussian mixands, then extend this result into multiple dimensions. An optimization may be formed, and it is first useful to define the L_2 distance norm for two distributions $p(x)$ and $q(x)$ as

$$L_2[p||q] = \int_{\mathcal{R}} \left(p(x) - q(x)\right)^2 dx \tag{B.45}$$

Assuming an L_2 norm cost, and L mixands, the objective is then

$$\min J = L_2[p||q] \text{ subject to } \omega_i > 0 \ \forall \ i \text{ and } \sum_{i=1}^{L} \omega_i = 1. \tag{B.46}$$

The L_2 norm is separable with the assumption that p is uniform over the domain $[a, b]$ as

$$L_2[p||q] = \frac{1}{b-a} + \int_{-\infty}^{\infty} q^2(x)dx - \frac{2}{b-a}\int_a^b q(x)dx. \tag{B.47}$$

Applying multiplicative properties of Gaussian pdfs and Gaussian integrals, the other terms of the L_2 norm are evalauted as

$$\int_{-\infty}^{\infty} q^2(x) = \sum_{i=1}^{L}\sum_{j=1}^{L} \omega_i\omega_j\Gamma(\mu_i, \mu_j, P_i, P_j) \tag{B.48}$$

$$\int_a^b q(x)dx = \frac{1}{2}\sum_{i=1}^{L} \omega_i \left[\text{erf}\left(\frac{b-\mu_i}{\sqrt{2P_i}}\right) - \text{erf}\left(\frac{a-\mu_i}{\sqrt{2P_i}}\right)\right] \tag{B.49}$$

The resultant form still requres $L \times (\omega, \mu, P) = 3L$ parameters to optimize, so several assumptions may be made. First, each mixand is given an equivalent weight. Mixand means are then equally distributed across the support $[a, b]$, and mixands are assumed homeoscedastic, such that the only variable to optimize is covariance σ^2. Then,

$$L_2[p||q] = \frac{1}{b-a} + \frac{\omega^2}{2\sqrt{\pi}\sigma}\sum_{i=1}^{L}\sum_{j=1}^{L}\exp\left(-\frac{1}{4}\left(\frac{\mu_i-\mu_j}{\sigma}\right)^2\right) - \frac{\omega}{b-a}\sum_{i=1}^{L}\left[\text{erf } B_i - \text{erf } A_i\right] \tag{B.50}$$

$$A_i = \left(\frac{a-\mu_i}{\sqrt{2}\sigma}\right), \quad B_i = \left(\frac{b-\mu_i}{\sqrt{2}\sigma}\right) \tag{B.51}$$

A variety of root finding algorithms may now be used with the gradient of the L_2 norm expressed as

$$\frac{dL_2[p||q]}{d\sigma} = \frac{\omega^2}{2\sqrt{\pi}\sigma^2}\sum_{i=1}^{L}\sum_{j=1}^{L}\left(\frac{1}{2}\left(\frac{\mu_i-\mu_j}{\sigma}\right)^2 - 1\right)\exp\left(\frac{1}{4}\left(\frac{\mu_i-\mu_j}{\sigma}\right)^2\right) \tag{B.52}$$

$$-\frac{2\omega}{(b-a)\sqrt{\pi}\sigma}\sum_{i=1}^{L}\left[A_i\exp(-A_i^2) - B_i\exp(-B_i^2)\right]$$

Note that the weights and means are given as

$$\omega = \frac{1}{L}, \quad \mu_i = a + \frac{(b-a)i}{L+1} \quad \forall\ i \in [1, 2, ..., L] \tag{B.53}$$

Unfortunately, the range-marginal PDF is decidedly not uniform, and

$$p_\rho(\rho) = \int_{-\infty}^{\infty} p_{\rho,\dot{\rho}}(\rho,\nu)d\nu \tag{B.54}$$

$$p_\rho(\rho) \propto \dot{\rho}_{max}(\rho) - \dot{\rho}_{min}(\rho) \tag{B.55}$$

To correct this issue, it is useful to first recognize that the established PDF is linear in weights, where

$$q(x) = \vec{h}^T(x)\vec{\omega} \tag{B.56}$$

$$h_i(x) = p_g(x;\mu_i,\sigma^2) \tag{B.57}$$

As such, weights may be updated in a least squares manner. Consider taking M samples from the range-marginal PDF. Least squares may then be used to minimizing the cost function

$$\min J = ||\vec{p} - H\vec{\omega}|| \text{ subject to } \vec{\omega} \geq \vec{0} \text{ and } \vec{1}^T\vec{\omega} = 1 \tag{B.58}$$

Note $H_{i,j}$ is a $M \times L$ matrix where

$$H_{i,j} = p_g(\rho_i; \mu_j, \sigma^2) \tag{B.59}$$

After the least squares update, the weight vector may simply be normalized to satisfy the unit constraint and an updated range-marginal PDF is established as

$$p_\rho(\rho) \approx \sum_{i=1}^{L_\rho} \omega_{\rho,i} p_g(\rho; \mu_{\rho,i}, \sigma_\rho^2) \tag{B.60}$$

Assuming a total of $L_\rho \times L_{\dot{\rho}}$ mixands are desired, for each mixand in the range marginal PDF, extreme range rate values (a_i, b_i) may be computed at the mixand mean range . The uniform range rate GMM may then be separately approximated with a set of $L_{\dot{\rho}}$ mixands, where

$$\omega_{\dot{\rho}} = \frac{1}{L_{\dot{\rho}}} \quad \mu_{\dot{\rho},j} = a_i + \frac{(b_i - a_i)j}{L_{\dot{\rho}} + 1} \tag{B.61}$$

and an appropriate variance $\sigma_{\dot{\rho},i}^2$ is found during the approximation.

The uniform PDF should have total weight $\omega_{\rho,i}$, and therefore,

$$\omega_{\rho_i,\dot{\rho}} = \omega_{\rho,i}\omega_{\dot{\rho}} \tag{B.62}$$

Mixand means and covariances may now be trivially defined over the full distribution as

$$\vec{\mu}_{\rho_i,\dot{\rho}_j} = \begin{bmatrix} \mu_{\rho_i} \\ \\ \mu_{\dot{\rho}_j}(\rho_i) \end{bmatrix} \tag{B.63}$$

$$P_{\rho_i,\dot{\rho}_j} = \begin{bmatrix} \sigma_\rho^2 & 0 \\ \\ 0 & \sigma_{\dot{\rho}(\rho_i)}^2 \end{bmatrix} \tag{B.64}$$

$$p(\rho,\dot{\rho}) = \sum_{i=1}^{L_\rho}\sum_{j=1}^{L_{\dot{\rho}}} \omega_{\rho_i,\dot{\rho}} p_g \left(\begin{bmatrix} \rho \\ \dot{\rho} \end{bmatrix} ; \vec{\mu}_{\rho_i,\dot{\rho}_j}, P_{\rho_i,\dot{\rho}_j} \right) \tag{B.65}$$