

**INITIALIZATION OF SEQUENTIAL ESTIMATION FOR UNOBSERVABLE
DYNAMICAL SYSTEMS USING PARTIAL INFORMATION IN THE PRESENCE
OF SYSTEMIC UNCERTAINTY**

A Dissertation
Presented to
The Academic Faculty

By

Johnny Lee Worthy III

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Aerospace Engineering

Georgia Institute of Technology

May 2017

Copyright © Johnny Lee Worthy III 2017

**INITIALIZATION OF SEQUENTIAL ESTIMATION FOR UNOBSERVABLE
DYNAMICAL SYSTEMS USING PARTIAL INFORMATION IN THE PRESENCE
OF SYSTEMIC UNCERTAINTY**

Approved by:

Dr. Marcus J. Holzinger, Advisor
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Glenn Lightsey
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Mark F. Costello
School of Aerospace Science
Georgia Institute of Technology

Dr. Daniel Scheeres
Department of Aerospace Engineer-
ing Sciences
University of Colorado Boulder

Dr. Travis Blake (Lt. Col., Ret.
USAF
Senior Manager for Space Domain
Awareness
Lockheed Martin Space Systems

Date Approved: April 5, 2017

To Lauren. None of this would have happened without your love, support, and insistence that I not give up. Thank you for helping me get through the occasional dark days of this PhD process.

ACKNOWLEDGEMENTS

I would like to thank Marcus J. Holzinger for being the best advisor a student could ask for. They say a lot of what makes or breaks your experience in grad school is your advisor, and I don't think I could have gotten a better one. From life to research to academics, the advice and guidance you've given me throughout this journey will continue to lead me to great success in the future.

I would also like to thank Dr. Pucha, my undergraduate research advisor, and Dr. Atiq Bhuiyan, then Ph.D. candidate, for their wonderful introduction into the world of research. I know I would not have ended up where I am without the interactions with Dr. Pucha and his research group and his early academic guidance. Working with him and Atiq gave me the chance to experience first hand what life in grad school would be like and helped me to develop the skills to which I attribute my success thus far.

Last, but not least, I would like to thank Holzinger's entire research group. I could not have had better colleagues to go through this process with. From incredibly useful research discussions to much needed board game time on Fridays, you guys really made this whole thing worth it.

TABLE OF CONTENTS

Acknowledgments	iv
List of Tables	x
List of Figures	xii
Summary	xv
Chapter 1: Introduction and Background	1
1.1 Space Situational Awareness	2
1.1.1 The Admissible Region Method	8
1.1.2 The Statistical Representation of the Admissible Region	9
1.1.3 Evidential Reasoning Theory Applied to the Admissible Region	10
1.1.4 Association of observations via the admissible region	12
1.2 Summary of contributions and relevant literature	14
1.3 List of Publications	17
Chapter 2: Analytic Consideration of Systemic Uncertainties in the Admissible Region Approach	19
2.1 Admissible States	19
2.2 Uncertainties in the Admissible States	22

2.3	Joint Admissible Region for Multiple Constraints	28
2.4	Track Correlation	30
2.5	Summary	31
2.6	Application	31
2.6.1	Angles-Only Initial Orbit Determination Initialization	32
2.6.2	Example Constraints for Initial Orbit Determination	34
2.6.3	Accounting for Uncertainty	35
2.7	Validation of the Approximation	39
2.7.1	True PDF	40
2.7.2	Kullback-Leibler divergence	42
2.8	Joint probability of set membership with Two Constraints	51
2.9	Joint Probability for Dependent Constraints	53
2.10	Conclusions	60
	Chapter 3: Probabilistic interpretation of the admissible region	61
3.1	General Probability Transformations	61
3.2	Admissible Region Transformations	63
3.2.1	Observability of Admissible States	63
3.2.2	Defining the Admissible Region	72
3.2.3	The Admissible Region Prior	73
3.2.4	Transformation of the Admissible Region Prior	74
3.2.5	Priors Using Uncertain Admissible Region Constraints	82
3.2.6	The Observability Condition	84

3.3	Additional Transformations	85
3.3.1	Linear Transformations	86
3.3.2	Sigma Point Transformations	86
3.3.3	Transformations Over Time	87
3.4	Discussion	88
3.5	Simulation and Results	89
3.6	Conclusions	94
Chapter 4: Evidential reasoning theory applied to the admissible region		98
4.1	Dempster-Shafer theory	98
4.2	Belief Functions on Real Numbers	101
4.3	Combination of Evidence	104
4.4	The Admissible Region BBA	105
4.5	Observability and Ignorance	112
4.6	Application to Sequential Estimation	114
4.7	Reduction to Bayesian Inference	115
4.8	Simulation and Results	117
4.8.1	Scenario 1	118
4.8.2	Scenario 2	123
4.8.3	Scenario 3	127
4.8.4	Scenario 4	132
4.9	Conclusions	134
Chapter 5: Association of observations via the admissible region		137

5.1	Association of Admissible Regions with Uncertainty	137
5.1.1	Optimization Approach for Set Intersection	138
5.1.2	Simplified Optimization Approach for Set Intersection	142
5.1.3	Simplified Optimization Approach for General Phenomenologies	147
5.1.4	Hypothesis Test for Observation Association	149
5.1.5	Solution at the Intersection of Admissible Regions	153
5.1.6	Approximation of the Bayesian Update using Gaussian Mixtures	156
5.1.7	The Observability Condition	159
5.2	Simulation Results	160
5.2.1	Case 1: Single Solution	163
5.2.2	Case 2: Ambiguous Solutions	167
5.2.3	Case 3: GPS Satellite Observation	171
5.2.4	Case 4: GEO Object Observation (Semi-Unobservable)	176
5.3	Conclusions	181
Chapter 6: Future Work		183
6.1	Resampling DS Particle Filter	183
6.2	Evidence Based Association and Tracking	184
6.3	Characterization and Tracking	185
6.4	High Area to Mass Ratio Objects	185
6.5	Characterization of Dependent Errors	186
Chapter 7: Conclusions		187

Appendix A: Constraint Derivatives	191
References	203

LIST OF TABLES

1.1	Overview of contributions, existing literature and publications	15
2.1	GEO Optical Observation Measurements	36
2.2	LEO Optical Observation Measurements	36
2.3	Observer Parameters (Atlanta, GA) and True Object States	36
2.4	Measurement Error and Parameter Uncertainty	37
2.5	Computer Specs	41
2.6	Sensitivities to Systemic Errors	50
4.1	Parameters, Measurement Error and Parameter Uncertainty	118
4.2	True Orbit	119
4.3	True Orbit for Case 3	127
5.1	Measurement Error and Parameter Uncertainty	161
5.2	True Orbit for Case 1	163
5.3	Estimated Orbit for Case 1	164
5.4	Case 1 Probability of association and optimization solutions	166
5.5	Case 2 Probability of association and optimization solutions	169
5.6	Estimated Orbit for Case 2 - Solution 1	170
5.7	Estimated Orbit for Case 2 - Solution 2	170

5.8	Estimated Orbit for Case 2 - Solution 3	170
5.9	True Orbit for Case 3	171
5.10	Case 3 Probability of association and optimization solutions	174
5.11	Estimated Orbit for Case 3	175
5.12	CPU Run Time Comparison	175
5.13	SDSS Star IDs	178

LIST OF FIGURES

1.1	Summary of the contributions	16
2.1	Determining the constraints for $\delta\mathbf{X}_u$	24
2.2	Admissible region for the Case 1 GEO observation	37
2.3	Admissible region for the Case 1 LEO observation	37
2.4	Analytic probabilities with uncertainty	39
2.5	Approximate analytical probability $\pm 3\sigma$ bounds	40
2.6	Monte Carlo numerically obtained probabilities	43
2.7	Error between analytically and numerically computed probability	44
2.8	% Error focused on regions of interesting behavior.	45
2.9	Two points along L used to calculate ΔL_j	45
2.10	D_{KL} as a function of % error for a standard Normal distribution	47
2.11	Plots of D_{KL} comparing the analytical and Monte Carlo simulation	47
2.12	D_{KL} over the well behaved region of the line s plotted along with % Error	48
2.13	Validation of approximate analytical methodology	51
2.14	Analytical probabilities for simulated cases	54
2.15	Analytic joint probability	55
2.16	Monte Carlo joint probability	56
2.17	Percent error for LEO case	57

2.18	Revised analytic joint probability	59
2.19	Revised percent error between analytical and numerical methods	59
3.1	A set of 5000 points sampled uniformly from the admissible region.	91
3.2	Values of $ \partial\mathbf{x}/\partial\tilde{\mathbf{x}} $ evaluated for each particle $\mathbf{x}(t)$	92
3.3	Outline of the two approaches for generating the PDF in cartesian coordinates	95
3.4	Condition number of $\mathbf{P}(t, t_0, \mathbf{x}(t))$ computed for each particle $\mathbf{x}_u(t)$	95
3.5	Admissible region after 30 minutes	95
3.6	Difference in PDFs after 30 minutes	96
3.7	Difference in PDFs after 10 minutes	96
3.8	Difference in PDFs after 60 minutes	97
4.1	Initial distributions of filters for Scenario 1	121
4.2	Belief, Plausibility, and Probability updates for $t = 40s$	123
4.3	Belief, Plausibility, and Probability updates for $t = 120s$	124
4.4	Belief, Plausibility, and Probability updates for $t = 480s$	125
4.5	The belief mass assigned to the nonsingletons of Ω	126
4.6	Asymptotic behavior of Ignorance for Scenario 1.	126
4.7	Belief, Plausibility, and Probability updates for $t = 540s$	127
4.8	Incorrect hypothesis belief mass	128
4.9	Asymptotic behavior of Ignorance for Scenario 2.	128
4.10	Scenario 3 initial distributions	130
4.11	Belief, Plausibility, and Probability updates for $t = 2017-03-04T09:40:04$	131
4.12	The belief mass assignments to the propositions in Θ	131

4.13	Images of the tumbling COSMOS 1247 object.	133
4.14	Initial distributions for Scenario 4	134
4.15	Belief, Plausibility, and Probability updates for $t = 160s$	135
4.16	Results for how the belief assigned to $\bar{\mathcal{X}}$ changes with constraint strictness .	136
5.1	The distance metric d defined on $\mathcal{X}_i(t)$ and $\mathcal{X}_j(t)$	141
5.2	The distance metric defined on $\mathcal{X}_{d,i}(t)$ and $\mathcal{X}_{d,j}(t)$	143
5.3	Simplified optimization problem	145
5.4	An notional example where there are two local minima.	147
5.5	Illustration of uncertain set $\mathcal{X}_{d,i}$	149
5.6	A representation of the how f_0 is constructed.	151
5.7	Binary Hypothesis Testing Setup	153
5.8	Notional multi-intersection case	154
5.9	Ambiguous case due to uncertainties	154
5.10	Optimal solutions for Case 1	165
5.11	Comparison of approximate and numerical PDFs	166
5.12	Illustration of three potential local minima solutions found by the optimizer	168
5.13	Optimal solutions for Case 2	169
5.14	Comparison of PDFs for Case 2	170
5.15	Real images containing GPS BIIR-2	172
5.16	Optimal solutions for Case 3	174
5.17	PDF comparison for Case 3	175
5.18	Real image containing an uncorrelated streak	176

5.19	Second uncorrelated streak	177
5.20	Closeup image of the streaks of the GEO object.	178
5.21	Optimal solutions for Case 4	180
5.22	Distance metric for Case 4	180
5.23	Uninformative prior update for Case 4	181
6.1	Three satellites in the ANIK-F1 Cluster at GEO	184

SUMMARY

The primary goal of space situational awareness and space domain awareness is to acquire and maintain a full characterization of the space environment. The electro-optical and radar sensors used to observe space objects are not able to fully characterize the state of an object when either a single or a short, unobservable sequence of measurements is obtained. The admissible region method is often used as a bootstrap method to initialize estimation in these unobservable cases, however there are several areas of improvement identified and addressed in this thesis. First, an approximate analytic probability of set membership function is defined to account for systemic uncertainties in the generation of the admissible region set. Then, a rigorous application of probability theory proves the admissible region to be an uninformative prior and defines the necessary conditions for when this uninformative prior becomes a true PDF. Belief and plausibility functions, derived from Dempster-Shafer theory, are then applied to remove any assumptions on prior probabilities and enable testing of the constraint hypotheses used to construct admissible regions. Finally, an optimization based method is defined which uses the admissible regions from sets of uncorrelated tracks and determines the probability of association. The efficacies of the proposed methodologies are demonstrated using both simulated and real observation data from the Georgia Tech Space Object Research Telescope.

CHAPTER 1

INTRODUCTION AND BACKGROUND

State estimation is a critical component of most modern dynamical systems. In certain cases a sensor can essentially provide full state information. For example, radar ranging devices operated at high frequencies can essentially provide information on both position and velocity states. However, in more challenging cases a given sensor may not be able to fully provide a state estimate. This may be due to the fact that the observations are only available sporadically, only a subset of the state space can be measured, or the system is inherently unobservable. Initializing an estimation scheme in these situations is then complicated by the fact that due to this unobservability a well-defined prior distribution may be impossible to obtain [1, 2]. Uninformative priors are used as a way to address this problem in Bayesian estimation schemes, however they inherently violate axioms of probability [3, 4]. The challenge and goal is to provide an improved way to initialize a state estimation scheme conditioned on the known or estimated uncertainties in the system without violating probability theory.

In a broader sense, this particular problem of state estimation can be generalized to the field of data fusion, since each sensor provides information about a system which is ‘fused’ in some sense to yield a better understanding about the state of the system. Furthermore, thinking of state estimation in this way opens up additional methodologies which can be utilized to fuse data. While in many engineering systems Bayesian estimation is typically the primary tool, there are many other approaches to fuse data which are starting to permeate into the engineering disciplines. Khaleghi et. al. offers a comprehensive review of the recent developments in data fusion and the different fusion rules [5]. Probabilistic fusion handles uncertainty well and includes the classic Kalman Filter and its derivatives as well as sequential monte carlo methods and the particle filter [6, 7, 8, 9]. Evidential reasoning is

essentially a relaxation of probability theory which enables imprecision of data in addition to uncertainty [10, 11]. Fuzzy set theory further relaxes evidential reasoning by utilizing partial set membership, providing a way to quantify vagueness of data [12]. While there are several other data fusion approaches discussed by Khaleghi, concepts from each of these three types can often be used in a complimentary fashion and the contributions of this work will explore the use of a hybrid of these approaches to address some current problems with initialization of state estimation using uninformative priors.

Khaleghi et. al. also offers several driving motivations for data fusion based on the properties of the systems often being estimated and the current state of the art research in data fusion systems. Some of the relevant challenges for state estimation in dynamical systems are posed by the imperfection and, more importantly, the imprecision of the data, the association of disparate data, operational timing considerations, and potentially conflicting data or spurious data. Many estimation schemes are incapable of accounting for or addressing all of these challenges and problems, however recent developments and advances in random finite set statistics for data fusion, initially proposed by Mahler, appear to offer a potential solution [13]. Yet many of these methods are computationally expensive and it is desired to find alternative hybrid approaches which extend the utility of existing estimation tools even when tasked with these challenges. Many of these problems can be related directly to the challenges posed by Space Situational Awareness (SSA) where some aspect of each of these are present in the systems used to detect, track, and characterize space objects. The next sections introduce SSA, the current state estimation methodologies for space objects, problems with these approaches, and the proposed contributions which utilize various concepts of data fusion to address these problems.

1.1 Space Situational Awareness

Space situational awareness (SSA) and space domain awareness (SDA) are both national and global priorities as space is a critical strategic domain for military applications, a high-

value domain for commercial telecommunications, and a unique environment for Earth and space science [14, 15]. SDA lends from concepts of both Maritime Domain Awareness and Air Domain Awareness with the goal of achieving a comprehensive understanding of all aspects of the space environment that could impact the security, safety, or economy of the United States [16, 17]. SSA, on the other hand, is the ability to maintain a comprehensive understanding of the state of key elements of the space environment within a specific volume of time and how their states might change in the near future [18]. The current priority for the United States Joint Space Operations Center (JSpOC) is to fully characterize the space environment in Earth orbit for the protection of current and future space assets, the fundamental task supporting SSA goals [19, 20]. Research in improvements for the state estimation capabilities of the current SSA architecture is increasingly active in support of a comprehensive SDA program [21]. These systems are limited primarily by the inability of the current radar and optical systems to individually observe full state information on space objects over a short period of observation time or given a single observation. Yet these sensors must still be able to, given sufficient observation time, provide full state estimates. Sufficient observation time is not always given and there are vastly more objects in orbit than sensors, which leads to spurious observations of objects which may not be easily associated with other observations to provide a useful state estimate. This problem is often seen in data deprived systems and is a motivating factor for the current research in improved state estimation methods. This section outlines the current national SSA architecture, identifies current limitations, and outlines the specific needs identified by the SSA community both nationally and internationally.

SSA is primarily motivated by a recent proliferation of space objects, both operational payloads and debris from launches and collisions, leading to a national and global emphasis on protecting space assets and establishing a strong SSA program [21, 22]. This emphasis on SSA stems from several high profile collisions within the past decade. In 2007, the anti-satellite test on the Chinese weather satellite Fengyun-3 generated over 2,000 new pieces of

debris larger than 1cm in size [23]. In 2009, a defunct Russian Cosmos-2251 satellite collided with an operational Iridium-33 satellite, generating nearly 2,000 new debris objects larger than 10cm in size and several thousands of smaller debris objects [24]. It is estimated that high profile collisions such as these are likely to occur every 5 to 9 years [25]. Debris objects pose major threats as debris larger than 10cm can cause catastrophic failure and even objects between 1 and 10cm can cause major damage to core mission components [26, 27, 28]. An additional concern is raised due to the high area-to-mass ratio (HAMR) of some debris objects due to complex interactions with drag, in low Earth orbit (LEO), solar radiation pressure, as well as the dynamics of the Earth-Sun-Moon system [29]. These complex interactions make it difficult to consistently detect, track, and associate HAMR debris objects [30]. International measures are in place to attempt to mitigate the future proliferation of debris, but it is expected that with the number of payloads per launch growing, several planned large constellation missions^{1,2}, and the rapidly growing small satellite market³ there will continue to be demand for high quality SSA data to protect these assets.

The current operational system that detects and tracks these space objects is the US Space Surveillance Network (SSN), a combination of electro-optical (EO), conventional radar, and phase-array radar sensors [31, 32]. The electro-optical systems (EOS) are mainly used to augment the tracking capabilities of the SSN at geosynchronous Earth orbit (GEO) altitudes where it is difficult to do radar ranging [33, 34, 35]. The radar systems are most effective in LEO where systems like the Haystack radar and Millstone observatory are able to make accurate observations of objects down to around 10 centimeters in size [36, 37]. Phase-array systems enable enhanced performance and the new Space-Fence radar array is capable of making nearly 1.5 million observations of LEO objects daily [38]. Generally, the radar systems are capable of regularly detecting and tracking objects only down to 10cm

¹<http://www.spaceflightinsider.com/organizations/space-exploration-technologies/spacex-seeks-permission-4425-satellite-internet-constellation/>

²<https://www.itu.int/en/ITU-R/space/workshops/2015-prague-small-sat/Presentations/Planet-Labs-Safyan.pdf>

³<http://spacenews.com/smallsat-operators-have-yet-to-allay-concerns-about-space-junk/>

due to Raleigh scattering [39]. Objects much smaller than this cannot be regularly detected and thus cannot be regularly tracked, yet still pose significant threats. EOS are not limited by Raleigh scattering but by the amount of light that can be gathered from the objects, which decreases with the distance to and size of the object, as well as ambient weather conditions, such as day vs. night or clear vs. cloudy skies. This limitation of electro-optical sensors can be improved with space based optical sensors such as the one aboard the Space Based Space Surveillance (SBSS) satellite [40]. SBSS is one of the premier satellites for SSA as it is capable of surveying GEO objects and contributes significant detection and tracking data on GEO objects to the SSN. Due to the limitations of each type of sensor, while it is estimated that there are over 700,000 objects larger than 1cm in Earth orbit, only a small portion of these objects are actually regularly detected and tracked [41]. While a constantly evolving value, only about 18,000 objects of the estimated 700,000 are currently represented in the publicly available space object catalog (SOC) from Space-track.org⁴. Thus, in order to support the need for improved SSA, there is significant research to be done in the software, algorithms, and theory supporting the characterization of space objects as motivated by a recent review of the current Air Force Space Command Astrodynamics's Standards [42].

One of the areas noted for improvement in SSA by the National Academy is data association and orbit determination, the primary topics of this thesis. The observation of space objects with either radar or EO sensors is an example of an underdetermined system, as there is only partial information available to the observer over a short period of time. The measurement of the state of the object with an EOS consists of the angular position of the object relative to the observer. A time convolution of these individual angular measurements can yield the rate of change of the angular position of the object. An example of this is when a streak is captured as an object moves across the field of view of the sensor over a finite integration time. This information alone is not enough to fully determine the

⁴https://www.space-track.org/basicspacedata/query/class/tle_latest/ORDINAL/1/EPOCH/%3Enow-30/orderby/NORAD_CAT_ID/format/tle

state of the space object. The EO sensor is incapable of obtaining any range or range-rate information, the remaining two states needed to fully define the state of the object. Conversely, a radar measurement allows for computation of the range to the target, as well as the range-rate and the angular location of the target. However, radar measurements cannot determine the angle rate information which is necessary to determine the full state of the space object. For radar systems the angle rate states are undetermined and likewise the range and range-rate states are undetermined for EO sensors. Over long observation periods, i.e. when a component in the SSN is capable of long duration tracking of an object, then traditional initial orbit determination approaches such as Gauss' or Lambert's method will produce a full state estimate from these measurements [43]. Furthermore, if both types of sensors are used simultaneously to observe an object then a full state estimate is possible with high accuracy [44].

These traditional initial orbit determination methods rely conceptually on the curvature of the measurements to produce a state estimate. However, the measurements obtained from a short observation or a very short sequence of observations have essentially linear dynamics and the traditional methods begin to fail as the observation time decreases. This issue is the main motivation behind this thesis, particularly how to best initialize state estimation in underdetermined systems using uninformative priors from short measurement sequences and additionally how can disparate short measurement sequences be associated with one another. The next subsections describe the current state of the art research approaches for initial orbit determination in such systems and identify this problem as open area of research in the field which this thesis attempts to address.

Thesis Statement: State estimation in unobservable dynamical systems may be improved through the approximate, analytic consideration of systemic uncertainties, a rigorous application of probability and evidential reasoning theory to the resulting set membership functions, and a probabilistic methodology to associate uncorrelated

observations.

1.1.1 The Admissible Region Method

The failure of traditional orbit determination for short observations lead to the definition of this phenomenon as the Too Short Arc (TSA) problem in initial orbit determination since the measurement arc is not long enough to produce a state estimate through traditional means [45] [46]. As shown by Milani et. al. the TSA problem implies that for a particular TSA measurement, a continuum of potential states are possible for the range and range-rate (or equivalently angle rates in radar systems). This continuum of potential states is a direct result of the unobservability of these states in a TSA measurement. Milani et. al. introduced the admissible region method to solve this problem by imposing hypothesized constraints on this continuum of potential solutions based on the dynamics of the system [45]. The admissible region is then defined as a bounded set which contains all of the undetermined states which satisfy the hypothesized constraints. This is a major result for initial orbit determination from TSA measurements since Bayesian estimation requires an a priori distribution and the admissible region provides a straightforward method by which to define such a distribution. However, the admissible region as introduced gives a binary probability of set membership function which assigns a given undetermined state either 0 or 1 probability of satisfying the constraint hypothesis. Real systems are prone to observation error, timing uncertainty, and parameter uncertainties, each of which contributes to how well it can be known if a certain state satisfies the hypothesized constraint. As such, one open area of research is determining how to incorporate the measurement, parameter, and timing uncertainty into the admissible region method to yield an uncertain admissible region set. Additionally, it is desired then to determine how to find the intersection of a finite number of these uncertain sets, or equivalently how to account for several constraint hypotheses simultaneously.

Chapter 2 introduces a generalized notation for the admissible region and uses this notation to define an approximate analytic framework which accounts for the various uncertainties of the observation system in the degree to which a given state satisfies a given constraint

hypothesis. Then Bonferroni's inequality is applied to generalize the ability to account for an arbitrary number of constraint hypotheses while still incorporating the systemic uncertainties, providing a concise method to quantify the probability of set membership function for the resulting joint admissible region 'fuzzy' set.

Contribution 1: An analytic approximation framework to determine probability of set membership in admissible regions, accounting for systemic uncertainties (from the measurements, parameters, timing, etc.), yielding uncertain admissible regions which can incorporate any finite number of constraint hypotheses.

1.1.2 The Statistical Representation of the Admissible Region

After the introduction of the admissible region method, much work was done to improve its application to initial orbit determination. In particular, there has been work done to show different ways to represent the admissible region probabilistically for Bayesian estimation. Fujimoto et. al. show that the admissible region may be represented statistically as a uniform distribution over all states satisfying the hypothesized constraints [47]. This approach shows that since no particular state may be shown more likely to be the truth than another, each state must necessarily have equal probability. This approach also must assume that the admissible region has a discontinuity at the boundary where inside the boundary there is uniform probability but outside the boundary there is zero probability [48]. The work done by DeMars and Jah show that the bounded, compact set represented by the admissible region may be represented as a Gaussian mixture model (GMM) [49]. The GMM approach allows the admissible region to be represented in a computationally tractable way without discretizing individual points. It also gives 'fuzzy' boundaries which avoids the issue of a probability discontinuity at the constraint hypothesis boundary. Hussein et. al. show that by uniformly sampling from a set of alternative undetermined states and mapping the uniform distribution into the desired range and range-rate undetermined state space, a probabilistic representation of the admissible region could be generated [50]. This approach is

ultimately similar to the GMM approach except the resulting PDF over the admissible region is not uniform. Each of the approaches discussed above make the inherent assumption that a PDF is the correct statistical representation of the admissible region. There are certain mathematical and statistical requirements that must be met if this assumption is true, and if not then it is an open area of research to determine the correct statistical representation of the admissible region and how to initiate sequential Bayesian estimators from the admissible region.

Chapter 3 uses a rigorous application of probability theory to show that the admissible region is an uninformative, or diffuse, prior and demonstrates the inherent inconsistencies caused by treating the admissible region as a well defined PDF. Due to these inconsistencies, several theorems are presented which define necessary conditions for an admissible region to be treated as PDF to be transformed into an alternative state space for use in a Bayesian estimator. Furthermore, an observability condition is defined based on the length of the observation which indicates whether or not the system is observable, and equivalently whether or not the admissible region can be sufficiently treated as a PDF. When the condition is not met then the admissible region remains a diffuse prior for the purposes of state estimation and should be treated accordingly.

Contribution 2 : A presentation of the statistically correct method by which a sequential Bayesian estimator may be initiated with an admissible region, and more generally any uninformative prior as well as an observability condition indicating when a true PDF is defined.

1.1.3 Evidential Reasoning Theory Applied to the Admissible Region

What is desired is a methodology which does not require specific assumptions about the underlying probabilities, and thus does not permit inconsistent estimator behavior when different assumed probabilities are used. The problem arises again because the system is unobservable, which makes a traditional application of Bayesian probability difficult.

Bayesian probability requires that there is either support for or evidence against a given hypothesis. In general, when the problem is unobservable no such definitive support is available given a single measurement or an unobservable sequence of measurements. While states within the admissible region must support the hypothesized constraint being true, this in itself does not offer support for any one state being the true solution. Treating this region as a PDF is probabilistically incorrect because this region is just an uninformative prior which violates the the principles of probability theory [4].

An alternative branch of information fusion, Dempster-Shafer theory, deals with this problem by introducing plausibility as a third option which enables assignment of plausibility to states which neither directly support nor refute a hypothesis [10] [51]. The application of Dempster-Shafer theory to the admissible region problem can address the ambiguities that exist in the assignment of prior probability by recasting the admissible region as a region of plausibility. This falls directly in line with the Dempster-Shafer method since while the hypothesized constraints bound the set of potential solutions, they do not directly convey support for any one solution but define all plausible solutions to the problem.

A primary theoretical construct of Dempster-Shafer theory is the frame of discernment which contains the propositions which must be assigned belief mass [52, 53]. The proper construction of the frame of discernment for the admissible region problem enables the constraint hypothesis itself to be tested along with the individual states in the admissible region [54]. This provides the ability for a sequential estimator to potentially identify when a hypothesized constraint is incorrect and another hypothesis should be used. The aim of this work is to detail a generalized form of the frame of discernment for the admissible region problem and use it to define a Dempster-Shafer sequential estimation scheme. A unique feature of Dempster-Shafer theory is the concept of a probability bound provided by belief and plausibility [55]. While pignistic probabilities can be determined from belief, it can be shown also that as evidence is gathered belief and plausibility collapse to a sin-

gle value, the probability. The point when belief and plausibility become equal could, in general, indicate observability in dynamical systems and signify that a traditional Bayesian estimator could be initiated with the now fully defined probability distribution. This alternative approach offers an attractive method by which the a-priori probability assignment problem may be altogether avoided in favor of a more generalized estimation methodology.

Chapter 4 introduces Dempster-Shafer theory and defines its use for general underdetermined systems for which the belief assignment function is essentially vacuous. Through set membership, and more generally fuzzy set membership, functions, plausibility functions can be defined which serve as an upper bound to the underlying probability. Then, a combination rule is defined which takes plausibility functions from independent sources and combines them instead of the traditional Bayesian probability update. Furthermore, through a concise linear relationship specific to the construction of the admissible region problem under Dempster-Shafer theory, updated plausibility values can be mapped to updated belief values for a given state. These concepts are then implemented in a filter which tracks both belief and plausibility values over time and shows that when belief and plausibility are equal, the system is observable and a traditional Bayesian sequential estimation scheme could be used.

Contribution 3 : An evidential reasoning based sequential estimator which resolves the use of uninformative priors to initialize estimation schemes and enables the testing of the constraint hypotheses used to construct an admissible region.

1.1.4 Association of observations via the admissible region

Another area of research in the admissible region method, and SSA in general, is in observation association. Since obtaining TSA measurements from the SSN or other observation platforms do not give a full state estimate of the object, there is no direct way to associate two disparate observations. Furthermore, high area-to-mass ratio (HAMR) objects have varying dynamics due to the interaction of drag and solar radiation pressure which makes

it difficult to confidently association observations of the same HAMR objects [56, 30]. As such, there is much interest in using the admissible region as a foundation for observation correlation methods. Milani et. al. gives an overview of two object associate methods based on the admissible region in [57]. The first method uses an attribution penalty computed for each sampled point in the admissible region to identify potentially associated objects by setting a maximum allowed attribution penalty [58]. The second method uses integrals of motion by requiring that the energy and angular momentum at the time of the two disparate observations are equal. Siminski et. al. introduces an optimization based approach to identify a best fitting orbit solution from an admissible region [59]. Fujimoto et. al. and Maruskin et. al. show that by discretizing the admissible region and mapping to either Delunay or Poincaré elements, there is only one intersection, if it exists, of two disparate admissible regions and that intersection is the full state solution [60, 48, 61]. DeMars et. al. introduces a multiple hypothesis filter approach which does data association and state estimation with several TSA measurements [62]. A boundary value problem approach is applied to the admissible region by Fujimoto and Alfriend which uses the angle-rate information to eliminate potential state hypotheses [63]. Each of these methods have been implemented and demonstrated with some success but an open area of research is in the intersection theories introduced by Fujimoto et. al. and Maruskin et. al. [60, 48]. These intersection approaches require a single intersection between disparate admissible regions, but it is not shown that in general there is only one intersection possible. The inclusion of uncertainties will also affect the intersection approach since the intersection is no longer a single point but a distribution of points that could be the intersection. An open area of research in observation association is determining a computationally tractable approach which quantifies the probability of intersection, and thus the probability of correlation, of two or more disparate observations.

Chapter 5 introduces a generalization of the optimization problem defined to identify the point of intersection, if it exists, between two or more higher dimensional surfaces,

or manifolds. This optimization problem is then cast as a set of reduced order intersection problems which is shown to reduce computation time, thus improving computational tractability. The systemic uncertainties are then used to define a probability of association metric which attempts to quantify the extent to which a set of observations are associated based on the known or assumed uncertainties in the system. Furthermore, due to the inclusion of systemic uncertainties through the results presented in Chapter 2, the statistics about the solution(s) to the intersection problem can be defined analytically through the Hessian matrix. Examples are also included in Chapter 5 from real observation data to demonstrate the capabilities of this approach.

Contribution 4 : An optimization based methodology for uninformative priors that determines the probability that two underdetermined measurements are correlated, and simultaneously yields the resulting PDF about the estimated state if the observations are correlated.

1.2 Summary of contributions and relevant literature

To summarize, the proposed contributions of this work aim to improve the applicability of the admissible region approach to state estimation. Figure 1.1 illustrates the relationships between each of the contributions of this work. The first contribution involves incorporating uncertainties in the system into the construction of the admissible region. The second contribution involves using fundamental statistical properties to define the proper use of the admissible region prior for bayesian estimation. The third contribution expands further on the second by utilizing Dempster-Shafer theory to avoid the assignment of a prior until the system is observable through the concepts of belief and plausibility. The fourth contribution builds upon the first by utilizing uncertainties in the admissible region to determine a probability of association for two disparate observations. Table 1.1 summarizes the relevant literature for each of the contributions identified.

Table 1.1: Overview of contributions, existing literature and publications

	Define an approximate analytic method to account for systemic uncertainties in the admissible region method	Define the necessary conditions for an uninformative prior to be treated as a PDF based on observability	Define a method by which to avoid a priori assumptions on uninformative priors and test hypothesized constraints using Dempster-Shafer theory	Define a probabilistic methodology to associate disparate observations using the uncertain admissible region
Existing Literature	<p>Definition of the admissible region for optical and radar measurements</p> <p>Shows that the admissible region is a uniform PDF</p>	<p>Admissible region methods relying on a transformation of the admissible region</p> <p>Uses a transformation of a uniform PDF to obtain a probabilistic admissible region</p> <p>Defines the transformation of a multivariate PDF</p>	<p>Introduction to DS reasoning, defining the frame of discernment, and the belief mass assignment, belief and plausibility functions</p> <p>Brief discussion of why diffuse priors are incorrect for initialization in Bayesian state estimation</p> <p>Discussion of belief functions on real numbers and the assignment of belief for a finite subset of non-singleton propositions</p> <p>An introduction to Dempster-Shafer theory from a linear algebra point of view</p>	<p>Uses multiple hypothesis filtering to correlate TSA measurements</p> <p>Attribution penalty defined for correlating discretized admissible regions</p> <p>States two admissible regions intersect if they are correlated, the intersection is the state solution</p> <p>Gives optimization based approach for orbit fitting admissible regions for correlation</p>
	[64], [45], [65]			
	[66]			
	[67], [61], [68]			
	[50]			
	[69]			
	[10], [51], [52], [55], [53]			
	[4]			
	[4], [55], [54]			
	[70]			
[62]				
[57], [58]				
[67], [61]				
[68]				
Thesis Chapter	<p>Approximate analytical expression to account for uncertainty in admissible regions</p> <p>Correct approximate analytic expression to account for multiple constraint hypotheses</p>	<p>Statistically rigorous proof of admissible region prior, strict definition of proper filter instantiation</p>	<p>Employs DS theory to avoid the need to assign prior probability and instead update belief and plausibility until the system is observable</p>	<p>Defines the probability of association of observations using uncertain admissible regions</p>
Chapter 2 (J.1)				
Chapter 3 (J.2)				
Chapter 4 (J.4)				
Chapter 5 (J.3)				

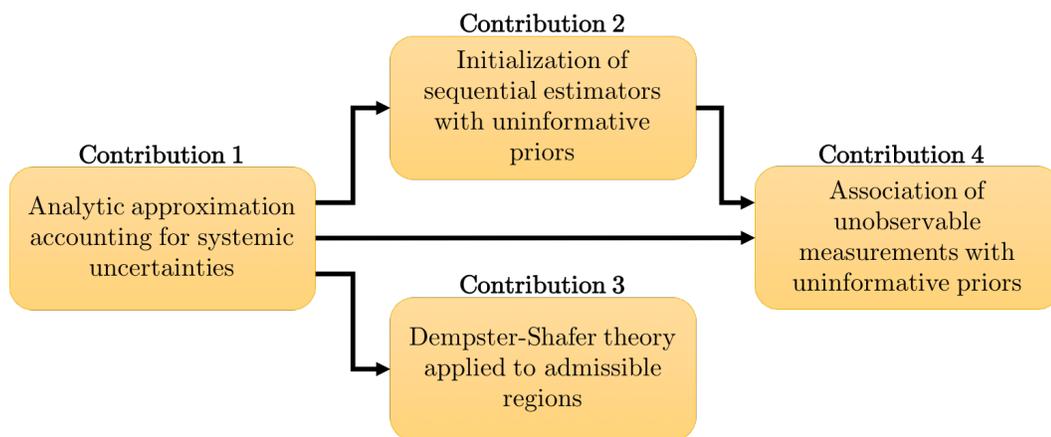


Figure 1.1: Summary of the contributions

1.3 List of Publications

The publications which contribute to the contents of this thesis are listed below in chronological order, separated by publication type.

Peer reviewed journal articles

J.1 J. L. Worthy and M. J. Holzinger, “Incorporating uncertainty in admissible regions for uncorrelated detections,” *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 9, pp. 1673-1689, 2015, doi:10.2514/1.G000890

J.2 J. L. Worthy and M. J. Holzinger, “Use of Uninformative Priors to Initialize State Estimation for Dynamical Systems,” *Advances in Space Research*, Submitted Nov 2016

J.3 J. L. Worthy, M. J. Holzinger, and D. J. Scheeres, “An Optimization Approach for Observation Association with Systemic Uncertainty Applied to Electro-Optical Systems,” *Advances in Space Research*, Submitted Dec 2016

J.4 J. L. Worthy, M. J. Holzinger, “Sequential Estimation from Uninformative Priors Using Dempster-Shafer Theory,” *IEEE Transactions in Aerospace and Electronic Systems*, To Be Submitted Apr 2017

Conference Presentations

C.1 J. L. Worthy, M. J. Holzinger, K. Fujimoto “Optical Sensor Constraints on Space Object Detection and Admissible Regions”, *AAS/AIAA Astrodynamics Specialist Conference*, Hilton Head, SC, August 2013

C.2 J. L. Worthy, M. J. Holzinger, “Incorporating Uncertainty in Admissible Regions for Uncorrelated Detections”, *AAS/AIAA Astrodynamics Specialist Conference - SPACE 2014*, San Diego, CA, August 2014

- C.3** J. L. Worthy, M. J. Holzinger, “Uncued Satellite Initial Orbit Determination Using Signals of Opportunity”, AAS/AIAA Astrodynamics Specialist Conference, AAS 15-733, Vail, CO, August 2015, doi:10.2514/6.2014-4307
- C.4** J. L. Worthy, M. J. Holzinger, “Application of Probability Transformation Mappings to the Admissible Region Method”, AAS/AIAA Astrodynamics Specialist Conference, AAS 15-733, Vail, CO, August 2015
- C.5** J. L. Worthy, M. J. Holzinger, D. J. Scheeres “An Optimization Based Approach to Correlation of Observations with Uncertainty”, AAS/AIAA Spaceflight Mechanics Meeting, Napa, CA, February 2016
- C.6** J. L. Worthy, M. J. Holzinger, “Dempster-Shafer Theory Applied to Admissible Regions”, AAS/AIAA Spaceflight Mechanics Meeting, San Antonio, TX, February 2017

CHAPTER 2

ANALYTIC CONSIDERATION OF SYSTEMIC UNCERTAINTIES IN THE ADMISSIBLE REGION APPROACH

This chapter introduces the admissible region approach as originally defined by Milani et. al. [45][65][64]. Since the introduction of the admissible region, the methodology has been applied to various phenomenologies. Milani et. al. originally defined the method for optical observations of near Earth objects [45]. Tommei et. al. discusses the extension of the method for radar observations [64]. Holzinger extended the method for application to magnetometer based detections of space object [71]. The notation defined to set up the admissible region problem in this chapter is generalized to permit the application of the admissible region approach to any underdetermined system. Further, the use of this generalized notation throughout the thesis permits the methodologies and concepts defined within to be applied not only to optical observations of space object, but also to any observation phenomenology of an unobservable system. Given the generalized notation, an approximate analytic inclusion of systemic uncertainties is defined and compared against numerical simulations. The resulting general approximate, analytic, uncertain admissible region formulation forms the basis for each of the following contributions in this thesis.

2.1 Admissible States

Admissible state approaches are useful in a variety of applications, and thus a general measurement model is used to derive the approach. A general nonlinear measurement model is assumed,

$$\mathbf{y} = \mathbf{h}(\mathbf{x}; \mathbf{k}, t) \tag{2.1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector, $\mathbf{x} \in \mathbb{R}^n$ is the state, $\mathbf{k} \in \mathbb{R}^l$ is a parameter vector, and $t \in \mathbb{R}$ is the time. In the case of an underdetermined system the number of states that can be observed, or measured, is less than the number of states in the system, giving $m < n$. This is significant because it implies there is no unique solution for \mathbf{x} given \mathbf{y} in Eqn. (2.1). This signifies that the undetermined states play no role in determining \mathbf{y} . This enables a partitioning of the state vector into the observable and undetermined states.

$$\mathbf{x}^T = \begin{bmatrix} \mathbf{x}_d^T & \mathbf{x}_u^T \end{bmatrix} \quad (2.2)$$

Eqn. (2.1) becomes

$$\mathbf{y} = \mathbf{h}(\mathbf{x}_d, \mathbf{x}_u; \mathbf{k}, t) \quad (2.3)$$

where $\mathbf{x}_d \in \mathbb{R}^m$ are the determined states, $\mathbf{x}_u \in \mathbb{R}^u$ are the undetermined states, and $u + m = n$. Since it has already been stated that the undetermined states have no impact on \mathbf{y} , it follows that Eqn. (2.3) can be written simply as

$$\mathbf{y} = \mathbf{h}(\mathbf{x}_d; \mathbf{k}, t) \quad (2.4)$$

This implies that there exists a unique relationship between \mathbf{x}_d and \mathbf{y} given the observation parameters \mathbf{k} and time t . Equivalently, there is a one-to-one and onto mapping from \mathbf{x}_d to \mathbf{y} , which means that the inverse mapping is guaranteed provide one-to-one and onto mappings from \mathbf{y} to \mathbf{x}_d . Thus,

$$\mathbf{x}_d = \mathbf{h}^{-1}(\mathbf{y}; \mathbf{k}, t) \quad (2.5)$$

To bound the undetermined state solution space, admissible region methods impose a set of state constraint hypotheses $\mathcal{H} = \{\mathcal{H}_1, \dots, \mathcal{H}_c\}$. Then an admissible region can be constructed from a given constraint $\mathcal{H}_i \in \mathcal{H}$ under the assumption that a given hypothesis \mathcal{H}_i is true. These constraints may result from the dynamics of the system or from physical

limitations and can be represented in the following form

$$g_i(\mathbf{x}_d, \mathbf{x}_u; \mathbf{k}, t) \leq 0 \quad (2.6)$$

Combining with Eqn. (2.5) yields

$$g_i(\mathbf{h}^{-1}(\mathbf{y}; \mathbf{k}, t), \mathbf{x}_u; \mathbf{k}, t) \leq 0 \quad (2.7)$$

This constraint is significant because a given observation will result in a continuum of possible solutions for \mathbf{x}_u , but Eqn. (2.7) defines an $n - m$ dimensional continuum of admissible solutions that all generate the observed measurement. This definition requires that the constraint reduces the set of solutions from an infinite continuum to a compact set with an integrable area. In order to formalize this set, $\mathcal{R}_i \in \mathbb{R}^u$ is defined.

$$\mathcal{R}_i := \{\mathbf{x}_u \in \mathbb{R}^u \mid g_i(\mathbf{h}^{-1}(\mathbf{y}; \mathbf{k}, t), \mathbf{x}_u; \mathbf{k}, t) \leq 0\} \quad (2.8)$$

The set of solutions in \mathcal{R}_i is also known as the admissible region for hypothesis \mathcal{H}_i in the literature [45][65][64]. The combined admissible region for all hypotheses \mathcal{H} is then given by

$$\mathcal{R} = \mathcal{R}_1 \cap \dots \cap \mathcal{R}_i \cap \dots \cap \mathcal{R}_c \quad (2.9)$$

The boundary of \mathcal{R}_i can be defined by setting the inequality in the set definition to an equality. This represents the constraint as a surface $\mathcal{B}_i \in \mathbb{R}^{u-1}$.

$$\mathcal{B}_i := \{\mathbf{x}_u \in \mathbb{R}^{u-1} \mid g_i(\mathbf{h}^{-1}(\mathbf{y}; \mathbf{k}, t), \mathbf{x}_u; \mathbf{k}, t) = 0\} \quad (2.10)$$

Notationally, the constraint of set \mathcal{B}_i can be simply defined as a constraint function

$$g_i(\mathbf{h}^{-1}(\mathbf{y}; \mathbf{k}, t), \mathbf{x}_u; \mathbf{k}, t) \equiv \kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) = 0 \quad (2.11)$$

which implies that given \mathbf{y} , \mathbf{k} and t , satisfying this equation gives the undetermined states \mathbf{x}_u in the set \mathcal{B}_i , which is the boundary of \mathcal{R}_i .

In the absence of uncertainty, the volume enclosed by \mathcal{B}_i , or equivalently the admissible region \mathcal{R}_i , is represented as a uniform distribution [45]. This admissible region probability of set membership in the notation proposed in this paper is formally stated as

$$\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] = \mathbb{P}[(\kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) \leq 0)] \quad (2.12)$$

which, without the effects of uncertainty, gives a uniform value for each state satisfying the constraint. Using Eqn. (2.12) and the following standard definition, a PDF typically used to instantiate a filter can be determined.

$$f_i[\mathbf{x}_u] = \frac{\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i]}{\int_{\mathcal{R}_i} \mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] d\mathbf{x}_u} \quad (2.13)$$

Because satisfaction of the constraint is binary, $\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i]$ takes on values of 0 or 1, yielding uniform distributions.

2.2 Uncertainties in the Admissible States

With an expression for the boundary of the admissible region method \mathcal{B}_i defined, the effects of uncertainty can be investigated. The uncertainty is generated by error in the measurement devices as well as uncertainties in the observer's parameters, such as the observer's position or velocity, and timing inaccuracies. A Taylor series expansion of Eqn. (2.11) allows for the analysis of the effects of uncertainty on the undetermined states by providing an expression containing the uncertainty contributions of the undetermined states, the measurements, the observations parameters and time. The Taylor series expansion of Eqn.

(2.11) is given by

$$\begin{aligned} \kappa_i(\mathbf{x}_u + \delta\mathbf{x}_u, \mathbf{y} + \delta\mathbf{y}, \mathbf{k} + \delta\mathbf{k}, t + \delta t) &= \kappa_i(x_u, \mathbf{y}, \mathbf{k}, t) + \frac{\partial\kappa_i}{\partial\mathbf{x}_u}\delta\mathbf{x}_u + \frac{\partial\kappa_i}{\partial\mathbf{y}}\delta\mathbf{y} \\ &+ \frac{\partial\kappa_i}{\partial\mathbf{k}}\delta\mathbf{k} + \frac{\partial\kappa_i}{\partial t}\delta t + \dots + \text{H.O.T} \end{aligned} \quad (2.14)$$

The full Taylor series expansion captures exactly the effects of uncertainty in the constraint equation, but the effects of the higher order terms can be assumed negligible for sufficiently small errors and small curvature of the constraint. As will be shown in later sections, the Taylor series approximation performs poorly when the constraint has high curvature or when the uncertainties are large due to the contributions of the higher order terms that are being neglected. Given any variation in the measurement or parameter vectors, the constraint defines the perturbed location of the admissible region boundary and must necessarily be satisfied, requiring

$$\kappa_i(\mathbf{x}_u + \delta\mathbf{x}_u, \mathbf{y} + \delta\mathbf{y}, \mathbf{k} + \delta\mathbf{k}, t + \delta t) = 0 \quad (2.15)$$

Using Eqn. (2.15) and assuming that the variations in $\delta\mathbf{y}$, $\delta\mathbf{k}$, and δt are sufficiently small, the higher order terms are negligible. The Taylor series simplifies to

$$-\frac{\partial\kappa_i}{\partial\mathbf{x}_u}\delta\mathbf{x}_u = \frac{\partial\kappa_i}{\partial\mathbf{y}}\delta\mathbf{y} + \frac{\partial\kappa_i}{\partial\mathbf{k}}\delta\mathbf{k} + \frac{\partial\kappa_i}{\partial t}\delta t \quad (2.16)$$

Eqn. (3.70) only relates a constant variation in \mathbf{y} , \mathbf{k} , and t to \mathbf{x}_u . It does not give information on how random uncertainties and errors in \mathbf{y} , \mathbf{k} , and t ($\delta\mathbf{y}$, $\delta\mathbf{k}$, δt) contribute to a resulting distribution of the undetermined states. The first order sensitivity of the undetermined states in Eqn. (3.70) can be rewritten as a similar equation in terms of random variables $\delta\mathbf{Y}$, $\delta\mathbf{K}$, and δT with each random variable belonging to a Gaussian distribution. Thus, combining the errors $\delta\mathbf{Y}$, $\delta\mathbf{K}$, and δT into a single variable $\delta\mathbf{Z} \in \mathbb{R}^{m+l+1}$, Eqn. (3.70) can be rewritten

as

$$-\frac{\partial \kappa_i}{\partial \mathbf{x}_u} \delta \mathbf{X}_u = \frac{\partial \kappa_i}{\partial \mathbf{z}} \delta \mathbf{Z} \quad (2.17)$$

where $\delta \mathbf{X}_u$ is now a random variable, and the uncertainty distributions

$$\delta \mathbf{Z} = \begin{bmatrix} \delta \mathbf{Y}^T & \delta \mathbf{K}^T & \delta T \end{bmatrix}^T$$

can be written as

$$\delta \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_z) \quad (2.18)$$

where \mathbf{P}_z is the known covariance matrix of $\delta \mathbf{Z}$. Eqn. (2.17) as written is a scalar equation, preventing a direct solution for $\delta \mathbf{X}_u$ without the addition of other constraints.

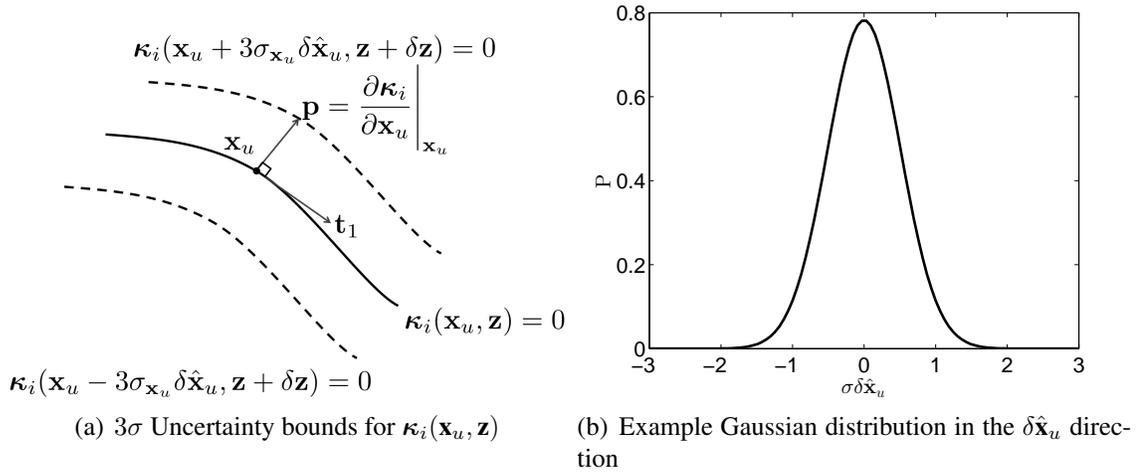


Figure 2.1: Determining the constraints for $\delta \mathbf{X}_u$

Considering the constraint $\kappa_i(\mathbf{x}_u, \mathbf{y}; \mathbf{k}, t)$, the derivative with respect to \mathbf{x}_u will be perpendicular to \mathcal{B}_i at the point \mathbf{x}_u . This perpendicular vector is defined as the \mathbf{p} vector

$$\mathbf{p} = \left. \frac{\partial \kappa_i}{\partial \mathbf{x}_u} \right|_{\mathbf{x}_u} \quad (2.19)$$

With this definition, Eqn. (2.17) can be rewritten as

$$\mathbf{p}^T \delta \mathbf{X}_u = -\frac{\partial \kappa_i}{\partial \mathbf{z}} \delta \mathbf{Z} \quad (2.20)$$

Consider the uncertainty $\delta \mathbf{x}_u$ at different locations \mathbf{x}_u along the constraint κ_i as shown in Figure 2.1. The variational location of the boundary can be described by a curve locally parallel to κ_i , and so at the point \mathbf{x}_u , can be fully described by the projection of $\delta \mathbf{x}_u$ in the direction of \mathbf{p} . Thus, the solution $\delta \mathbf{x}_u$ should have no component in any tangential directions, enabling the remaining additional constraints to be defined. For an m dimensional measurement, $n - m - 1 = u - 1$ tangential directions exist. The first tangential direction \mathbf{t}_1 is obtained by solving for any vector perpendicular to \mathbf{p} . Each subsequent tangential vector in \mathbb{R}^u can be obtained recursively by taking the cross product.

$$\mathbf{t}_j = \mathbf{p} \times \mathbf{t}_{j-1} \text{ for } j = 2, \dots, u - 1 \quad (2.21)$$

Thus the matrix of tangential directions is formed by

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_{u-1}^T \end{bmatrix} \quad (2.22)$$

where $\mathbf{T} \in \mathbb{R}^{u-1 \times u}$. To ensure a particular solution of $\delta \mathbf{x}_u$ has no tangential component, the following is defined.

$$\mathbf{t}_i^T \delta \mathbf{X}_u = 0 \quad (2.23)$$

Combined, the constraints on the solution for $\delta\mathbf{x}_u$ can be written as

$$\begin{bmatrix} \mathbf{p}^T \\ \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_{u-1}^T \end{bmatrix} \delta\mathbf{X}_u = \begin{bmatrix} \mathbf{p}^T \\ \mathbf{T} \end{bmatrix} \delta\mathbf{X}_u = \begin{bmatrix} -\frac{\partial\kappa_i}{\partial\mathbf{z}} \\ \mathbf{0} \end{bmatrix} \delta\mathbf{Z} \quad (2.24)$$

Because each row and column of the matrix on the left hand side are mutually orthogonal, it is by definition full rank and always invertible, yielding

$$\delta\mathbf{X}_u = \begin{bmatrix} \mathbf{p}^T \\ \mathbf{T} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{\partial\kappa_i}{\partial\mathbf{z}} \\ \mathbf{0} \end{bmatrix} \delta\mathbf{Z} \quad (2.25)$$

This equation expresses the relationship between uncertainties in \mathbf{z} and \mathbf{x}_u such that $\delta\mathbf{X}_u$ is orthogonal to the constraint boundary \mathcal{B}_i . Let

$$\mathbf{M}_i = \begin{bmatrix} \mathbf{p}^T \\ \mathbf{T} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{\partial\kappa_i}{\partial\mathbf{z}} \\ \mathbf{0} \end{bmatrix}$$

Then, to first order, the random variable $\delta\mathbf{X}_u$ orthogonal to the constraint surface is linearly related to the parameter random variable $\delta\mathbf{z}$.

$$\delta\mathbf{X}_u = \mathbf{M}\delta\mathbf{Z} \quad (2.26)$$

Now, the statistics of the uncertain boundary location $\delta\mathbf{X}_u$ must be determined. Since $\delta\mathbf{Z}$ belongs to a Gaussian distribution, the expected value of $\delta\mathbf{X}_u$ can be taken as

$$\mathbb{E}[\delta\mathbf{X}_u] = \mathbb{E}[\mathbf{M}\delta\mathbf{Z}] = \mathbf{0} \quad (2.27)$$

indicating $\delta\mathbf{X}_u$ has $\mathbf{0}$ mean, which is as expected since the uncertainties in $\delta\mathbf{Z}$ are assumed

to belong to Gaussian distributions with $\mathbf{0}$ mean. The covariance of $\delta\mathbf{X}_u$ is derived as follows

$$\begin{aligned}
\mathbf{P}_{\mathbf{x}_u} &= \mathbb{E} [(\mathbb{E}[\delta\mathbf{X}_u] - \delta\mathbf{X}_u)(\mathbb{E}[\delta\mathbf{X}_u] - \delta\mathbf{X}_u)]^T \\
&= \mathbb{E}[(\delta\mathbf{X}_u)(\delta\mathbf{X}_u)^T] \\
&= \mathbb{E}[\mathbf{M}\delta\mathbf{Z}\delta\mathbf{Z}^T\mathbf{M}^T] \\
&= \mathbf{M}\mathbb{E}[\delta\mathbf{Z}\delta\mathbf{Z}^T]\mathbf{M}^T
\end{aligned}$$

By definition $\mathbb{E}[\delta\mathbf{Z}\delta\mathbf{Z}^T] = \mathbf{P}_z$ is the variance of $\delta\mathbf{Z}$ which allows

$$\mathbf{P}_{\mathbf{x}_u} = \mathbf{M}\mathbf{P}_z^T\mathbf{M}^T \quad (2.28)$$

Where $\mathbf{P}_{\mathbf{x}_u} \in \mathbb{R}^{u \times u}$ is the covariance matrix for $\delta\mathbf{X}_u$ at a given point in \mathcal{B}_i . Because the relationship between $\delta\mathbf{x}_u$ is linear, if $\delta\mathbf{z}$ is assumed Gaussian then $\delta\mathbf{X}_u$ has no higher moments. Thus, with Equation (3.72), the distribution of the undetermined states in the direction of \mathbf{p} can be directly analytically approximated from knowledge of the distribution of the uncertainties in $\delta\mathbf{Z}$ at each point on the set \mathcal{B}_i of the admissible region \mathcal{R}_i . Furthermore, the Gaussian cumulative distribution function giving $\mathbb{P}[(x \leq X)]$ is fully defined by mean and variance, so the probability that a given \mathbf{x}_u is in the admissible region \mathcal{R}_i can be defined analytically as

$$\mathbb{P}[(\mathbf{x}_u \in \mathcal{R}_i)] = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\|\mathbf{x}_u - \mathbf{x}_{u,\mathcal{B}_i^\perp}\|}{\sqrt{2\operatorname{tr}\mathbf{P}_{\mathbf{x}_u,\mathcal{B}_i^\perp}}} \right) \right] \quad (2.29)$$

where $\mathbf{x}_{u,\mathcal{B}_i^\perp}$ is the point on the set \mathcal{B}_i that is perpendicular to \mathbf{x}_u , $\mathbf{P}_{\mathbf{x}_u,\mathcal{B}_i^\perp}$ is the covariance matrix computed for that point, and $\operatorname{erf}()$ is the Gauss error function given below.

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (2.30)$$

The approximate analytical probability of set membership can be fully obtained using Eqn.

(4.33). This particular formulation gives an analytical approach to determine the probability that a given solution \mathbf{x}_u is in the set \mathcal{R}_i . Also, note that \mathbf{x}_d is by construction Gaussian ($\mathbf{x}_d \sim \mathcal{N}(\mathbf{x}_d, \mathbf{P}_d)$), so a complete distribution for \mathbf{x} is found. This fundamental result allows for full state uncertainty initialization for an underdetermined detection.

Additionally, the contributions of the uncertainty can be quantified analytically. Since the distribution of \mathbf{x}_d is Gaussian, the trace of $\mathbf{P}_{\mathbf{x}_u}$ defines the distribution. The partial derivatives of $\mathbf{P}_{\mathbf{x}_u}$ with respect to \mathbf{P}_z gives the sensitivity of the distribution to the individual sources of error and uncertainty and is derived by the following.

$$\begin{aligned} \frac{\partial}{\partial \sigma_i^2} \text{tr}[\mathbf{P}_{\mathbf{x}_u}] &= \frac{\partial}{\partial \sigma_i^2} \text{tr}[\mathbf{M}\mathbf{P}_z^T\mathbf{M}^T] \\ &= \text{tr}[\mathbf{M}\mathbf{e}_i\mathbf{e}_i^T\mathbf{M}^T] \end{aligned} \quad (2.31)$$

where $\sigma_i^2 = \mathbf{P}_{z(i,i)}$ and \mathbf{e}_i is the i^{th} column of the $m + l + 1$ dimensional identity matrix.

2.3 Joint Admissible Region for Multiple Constraints

Given the set of constraints \mathcal{H} , it is likely that the probability of set memberships of each individual constraint hypothesis \mathcal{H}_i overlaps with those of the other constraints in \mathbb{R}^u . The interaction of these probabilities is of interest since the combined membership functions provides a full joint distribution under the assumption that all constraint hypotheses are true. To characterize the joint probability, a PDF for a single constraint can be written as

$$f(\mathbf{x}_u \in \mathcal{R}_i) = \frac{\mathbb{P}[\boldsymbol{\kappa}_i(\mathbf{x}_u) \leq 0]}{\int_{\mathcal{R}_i} \mathbb{P}[\boldsymbol{\kappa}_i(\mathbf{x}_u) \leq 0] d\mathbf{x}_u} \quad (2.32)$$

which can otherwise be written as

$$f(\mathbf{x}_u \in \mathcal{R}_i) = \frac{\mathbb{P}[\boldsymbol{\kappa}_i(\mathbf{x}_u) < 0]}{\int_{\mathcal{R}_i} \mathbb{P}[\boldsymbol{\kappa}_i(\mathbf{x}_u) < 0] d\mathbf{x}_u} \cup \frac{\mathbb{P}[\boldsymbol{\kappa}_i(\mathbf{x}_u) = 0]}{\int_{\mathcal{R}_i} \mathbb{P}[\boldsymbol{\kappa}_i(\mathbf{x}_u) = 0] d\mathbf{x}_u} \quad (2.33)$$

Furthermore, since $\mathbb{P}[\kappa_i(\mathbf{x}_u) < 0]$ and $\mathbb{P}[\kappa_i(\mathbf{x}_u) = 0]$ are mutually exclusive events

$$f(\mathbf{x}_u \in \mathcal{R}_i) = \frac{\mathbb{P}[\kappa_i(\mathbf{x}_u) < 0]}{\int_{\mathcal{R}_i} \mathbb{P}[\kappa_i(\mathbf{x}_u) < 0] d\mathbf{x}_u} + \frac{\mathbb{P}[\kappa_i(\mathbf{x}_u) = 0]}{\int_{\mathcal{R}_i} \mathbb{P}[\kappa_i(\mathbf{x}_u) = 0] d\mathbf{x}_u} \quad (2.34)$$

Define $\mathbb{P}[\mathbf{x}_u \in \mathcal{R}]$ as the joint probability that $\kappa_i(\mathbf{x}_u) \leq 0$ for $i = 1, \dots, c$ such that

$$\mathbb{P}[\mathbf{x}_u \in \mathcal{R}] = \mathbb{P}[\kappa_1(\mathbf{x}_u) \leq 0] \cap \dots \cap \mathbb{P}[\kappa_i(\mathbf{x}_u) \leq 0] \cap \dots \cap \mathbb{P}[\kappa_c(\mathbf{x}_u) \leq 0] \quad (2.35)$$

Substituting Eqn. (2.34)

$$\begin{aligned} \mathbb{P}[\mathbf{x}_u \in \mathcal{R}] &= \{\mathbb{P}[\kappa_1(\mathbf{x}_u) < 0] + \mathbb{P}[\kappa_1(\mathbf{x}_u) = 0]\} \cap \dots \\ &\quad \cap \{\mathbb{P}[\kappa_i(\mathbf{x}_u) < 0] + \mathbb{P}[\kappa_i(\mathbf{x}_u) = 0]\} \cap \dots \\ &\quad \cap \{\mathbb{P}[\kappa_c(\mathbf{x}_u) < 0] + \mathbb{P}[\kappa_c(\mathbf{x}_u) = 0]\} \end{aligned} \quad (2.36)$$

The distribution of $\delta\mathbf{X}_u$ along the \mathbf{p} direction is a continuous Gaussian distribution, thus $\mathbb{P}[\kappa_i(\mathbf{x}_u) = 0] = 0$ which after expansion of Eqn (2.36) yields

$$\mathbb{P}[\mathbf{x}_u \in \mathcal{R}] = \mathbb{P}[\kappa_1(\mathbf{x}_u) < 0] \cap \dots \cap \mathbb{P}[\kappa_i(\mathbf{x}_u) < 0] \cap \dots \cap \mathbb{P}[\kappa_c(\mathbf{x}_u) < 0] \quad (2.37)$$

For independent events A and B , Bayes' Theorem simplifies to $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Under the assumption that the constraints are independent, the probability that \mathbf{x}_u is in the set joint set \mathcal{R} is given by

$$\mathbb{P}[\mathbf{x}_u] = \mathbb{P}[\kappa_1(\mathbf{x}_u) < 0] \times \dots \times \mathbb{P}[\kappa_i(\mathbf{x}_u) < 0] \times \dots \times \mathbb{P}[\kappa_c(\mathbf{x}_u) < 0] \quad (2.38)$$

were the multiplication is the pointwise multiplication of the probabilities computed for each constraint. Thus, using Eqn. (2.13), Eqn. (2.38) and Eqn. (4.33) the joint probability of set membership for an arbitrary number of constraints can be defined analytically. Note,

in general the constraints are not independent and a full treatment of joint probability of set membership is given in §2.9 where the results are compared against those of the fully independent constraint assumption.

2.4 Track Correlation

The last section presents a method by which several constraints can be combined to generate a joint probability distribution. Similarly, this method can be used to joint probability distributions from different epochs. Fujimoto’s work shows that observations can be associated by applying Bayes’ rule to an admissible region generated from two epochs [66]. A non-zero result indicates the observations are associated. Specifically, if the objects are associated then there is exactly one solution with non-zero probability. An exception to this occurs when there are multiple closely spaced objects being tracked or if there are multi-revolutions [72].

However, due to observation errors and uncertainties, this particular solution does not fully capture the probability that the measurements are associated. It is possible to, through the inclusion of uncertainty, instead obtain a distribution around this point. Consider two admissible region PDFs $f(\mathbf{x}(t_2); \kappa_i(\mathbf{x}_u, \mathbf{z}_1, t_1))$ and $f(\mathbf{x}(t_2); \kappa_j(\mathbf{x}_u, \mathbf{z}_2, t_2))$, where $\mathbf{x}(t_2)$ denotes that the admissible region generated from $\kappa_i(\mathbf{x}_u, \mathbf{z}_k, t_k)$ is propagated to a common time t_2 . Assuming that each observation is statistically independent, then applying Bayes’ rule as before

$$f(\mathbf{x}(t_2); \kappa_i(\mathbf{x}_u, \mathbf{z}_1, t_1)) \cap f(\mathbf{x}(t_2); \kappa_j(\mathbf{x}_u, \mathbf{z}_2, t_2)) = f(\mathbf{x}(t_2); \kappa_i(\mathbf{x}_u, \mathbf{z}_1, t_1)) \times f(\mathbf{x}(t_2); \kappa_j(\mathbf{x}_u, \mathbf{z}_2, t_2)) \quad (2.39)$$

Eqn. (2.39) gives the joint probability that a particular state \mathbf{x}_u is in the admissible region at both times. Using Eqn. (2.13) the joint probability density can then be computed and applied to a discretized state space similar to the approach taken by Fujimoto. Importantly, it is not necessary that each measurement result from the same sensor phenomenology, nor

is it required that $t_1 \leq t_2$. A full treatment of track correlation with the uncertain admissible region is further defined in §5.

2.5 Summary

The presented approach is valid for any measurement sensor with undetermined states, such as optical, radar, or magnetometer sensors. Furthermore, any constraint hypothesis that is valid for the system observed may be used with this approach. For SSA, this may include minimum or maximum orbit energy constraints, illumination constraints, inclination constraints, etc. This approach provides a methodology to approximate uncertainty in the admissible states based on the observer and assumed parameter uncertainty. Then, assuming a known Gaussian distribution for the observer's uncertainty $\delta\mathbf{Z}$, the distribution of the admissible region is analytically approximated. This enables a probability distribution function in the admissible region \mathcal{R}_i to be analytically generated simply by knowing the uncertainties associated with the measurement sensors and observer. This analytical method can generate an approximate distribution several orders of magnitude faster than Monte Carlo numerical analyses, improving computational tractability for practical use. The analytical method is also extended to multiple constraints to generate a combined admissible region \mathcal{R} and the resulting combined joint probability of set membership associated with \mathcal{R} . The proposed method can also be used for associating admissible regions at different epochs by computing the joint probability distribution formed by each observation.

2.6 Application

This section explores application of the analytic consideration of systemic uncertainties to example observation scenarios to show agreement between the approximation and numerical simulations.

2.6.1 Angles-Only Initial Orbit Determination Initialization

The proposed methodology in the Theory section is applied in this section to the problem of initial orbit determination for space objects using optical measurements. This section begins with a review of the relevant results originally developed by Milani et. al. [45][65], then applies the theoretical contributions of this paper. Optical measurements generate angle and angle rates of the objects tracked using a streak or sequence of angle measurements in right ascension α and declination δ . The parameters associated with optical measurements include the observer position, \mathbf{o} and velocity, $\dot{\mathbf{o}}$, and the times at which the observations are made (or the start/stop times of a streak). Using this information, the position \mathbf{r} and velocity \mathbf{v} of the object are given by

$$\begin{aligned}\mathbf{r} &= \mathbf{o} + \rho \hat{\mathbf{l}} \\ \mathbf{v} &= \dot{\mathbf{o}} + \dot{\rho} \hat{\mathbf{l}} + \rho \dot{\alpha} \hat{\mathbf{l}}_{\alpha} + \rho \dot{\delta} \hat{\mathbf{l}}_{\delta}\end{aligned}$$

where α is the right ascension, δ is the declination, ρ is the range to the target, $\dot{\rho}$ is the range-rate, and $\hat{\mathbf{l}}$, $\hat{\mathbf{l}}_{\alpha}$, and $\hat{\mathbf{l}}_{\delta}$ are given by

$$\hat{\mathbf{l}} = \begin{bmatrix} \cos \alpha \cos \delta \\ \sin \alpha \cos \delta \\ \sin \delta \end{bmatrix}$$

$$\hat{\mathbf{l}}_{\alpha} = \begin{bmatrix} -\sin \alpha \cos \delta \\ \cos \alpha \cos \delta \\ 0 \end{bmatrix}$$

$$\hat{\mathbf{1}}_\delta = \begin{bmatrix} \cos \alpha \sin \delta \\ -\sin \alpha \sin \delta \\ \cos \delta \end{bmatrix}$$

For this system the states \mathbf{x} , observations \mathbf{y} , and parameters \mathbf{k} are defined as follows.

$$\begin{aligned} \mathbf{x}^T &= \begin{bmatrix} \alpha & \dot{\alpha} & \delta & \dot{\delta} & \rho & \dot{\rho} \end{bmatrix} \\ \mathbf{y}^T &= \begin{bmatrix} \alpha_1 & \cdots & \alpha_q & \delta_1 & \cdots & \delta_q \end{bmatrix} \\ \mathbf{k}^T &= \begin{bmatrix} \mathbf{o}^T & \dot{\mathbf{o}}^T \end{bmatrix} \end{aligned}$$

where q is the number of observations made and $\dot{\alpha}$ and $\dot{\delta}$ are the angle rates.

For an observation with two measurements, the combined measurement and parameter vector $\mathbf{z} \in \mathbb{R}^{12}$ is simply given by

$$\mathbf{z}^T = \begin{bmatrix} \alpha_1 & \alpha_2 & \delta_1 & \delta_2 & t_1 & t_2 & \mathbf{o}^T & \dot{\mathbf{o}}^T \end{bmatrix} \quad (2.40)$$

To generate the angle rate data from angle pairs in \mathbf{z} , a Lagrange Interpolation shown in Eqn. (2.41) is used. Using this interpolation, the effects of time uncertainty can be accounted for analytically in the dynamics by substituting directly for $\dot{\alpha}$ and $\dot{\delta}$.

$$\begin{aligned} \dot{\alpha} &= \alpha(t_1) \frac{(t-t_2) + (t-t_3) + \cdots + (t-t_n)}{(t_1-t_2)(t_1-t_3) \cdots (t_1-t_q)} + \alpha(t_2) \frac{(t-t_2) + (t-t_3) + \cdots + (t-t_q)}{(t_2-t_1)(t_2-t_3) \cdots (t_2-t_q)} \\ &+ \cdots + \alpha(t_l) \frac{(t-t_1) + (t-t_2) + \cdots + (t-t_{q-1})}{(t_q-t_1)(t_q-t_2) \cdots (t_l-t_{q-1})} \end{aligned} \quad (2.41)$$

From Eqn. (2.41) and \mathbf{y} , 4 of the 6 states of \mathbf{x} can be observed or determined and the

undetermined states for this system are given by

$$\mathbf{x}_u = \begin{bmatrix} \rho \\ \dot{\rho} \end{bmatrix}_{2 \times 1} \quad (2.42)$$

2.6.2 Example Constraints for Initial Orbit Determination

With the undetermined states \mathbf{x}_u defined, constraint hypotheses on those states must be imposed. Many constraints exist for orbiting bodies based on the assumptions made. A primary assumption for simple analyses is that of 2-body motion, allowing for the specific orbital energy equation to be used. The constraint κ_1 is generated by enforcing the energy equation such that the space object orbits about the Earth. To constrain the solutions to only objects in orbit about Earth, the admissible region set \mathcal{R} can be defined as $\{\mathbf{x}_u \in \mathbb{R}^2 | \varepsilon(\mathbf{r}, \dot{\mathbf{r}}) \leq 0\}$. The set \mathcal{B}_1 is then defined as $\{\mathbf{x}_u \in \mathbb{R}^2 | \varepsilon(\mathbf{r}, \dot{\mathbf{r}}) = 0\}$, which is given as the solution to the following equation.

$$\kappa_1(\mathbf{x}_u, \mathbf{z}) = 2\varepsilon(\mathbf{r}, \dot{\mathbf{r}}) = \dot{\rho}^2 + w_1\dot{\rho} + w_2\rho^2 + w_3\rho + w_4 - \frac{2\mu}{\sqrt{\rho^2 + w_5\rho + w_0}} = 0 \quad (2.43)$$

with w_0 through w_5 defined as in [64].

$$\begin{aligned} w_0 &= \|\mathbf{o}\|^2 & w_3 &= 2\dot{\alpha} \langle \dot{\mathbf{o}} \cdot \hat{\mathbf{i}}_\alpha \rangle + 2\dot{\delta} \langle \dot{\mathbf{o}} \cdot \hat{\mathbf{i}}_\delta \rangle \\ w_1 &= 2 \langle \dot{\mathbf{o}} \cdot \hat{\mathbf{i}} \rangle & w_4 &= \|\dot{\mathbf{o}}\|^2 \\ w_2 &= \dot{\alpha}^2 \cos^2 \delta + \dot{\delta}^2 & w_5 &= 2 \langle \mathbf{o} \cdot \hat{\mathbf{i}} \rangle \end{aligned}$$

From Eqn. (2.43) the solution continuum \mathbf{x}_u satisfying the constraint can be determined given \mathbf{z} . Also from Eqn. (2.43) \mathbf{t} and \mathbf{p} directions are defined at each point on \mathcal{B}_1 simply as

$$\begin{aligned}\mathbf{p}(\mathbf{x}_u)^T &= \left[\begin{array}{cc} \frac{\partial \kappa_1}{\partial \rho} & \frac{\partial \kappa_1}{\partial \dot{\rho}} \end{array} \right]_{\mathbf{x}_u} \\ \mathbf{t}(\mathbf{x}_u)^T &= \left[\begin{array}{cc} -\frac{\partial \kappa_1}{\partial \dot{\rho}} & \frac{\partial \kappa_1}{\partial \rho} \end{array} \right]_{\mathbf{x}_u}\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \kappa_1}{\partial \rho} &= 2\rho w_2 + w_3 + \frac{\mu(2\rho + w_5)}{(\rho^2 + w_5\rho + w_0)^{3/2}} \\ \frac{\partial \kappa_1}{\partial \dot{\rho}} &= 2\dot{\rho} + w_1\end{aligned}$$

and the terms of $\frac{\partial \kappa}{\partial \mathbf{z}}$ are given in the Appendix. Note that the Lagrange interpolation (Eqn (2.41)) is used for the angular rates, enabling computation of the partial derivatives of κ_1 with respect to time. With κ_1 defined, example deterministic admissible regions are computed based on the measurements given in Tables 2.1 and 2.2 and the observer parameters given in Table 2.3. The cases considered include an object at LEO and an object at GEO with two observation intervals to demonstrate the effects including uncertainty can have. Optical measurements can be generated from streaks for which a single detection yields two angles or from multiple point detections over time. Table 2.3 shows measurements of the latter form with 3 detections forming the observation. The admissible regions computed from the Case 1 observations are shown in Figure 2.2 and Figure 2.3. As can be seen in Figures 2.2a and 2.3a, without inclusion of measurement uncertainty, the approach generates an uniform distribution with a discontinuous probability density at the boundary.

2.6.3 Accounting for Uncertainty

Recall the standard admissible region (shown in Figures 2.2 and 2.3) has a uniform probability distribution within the constraint. To approximate the true probability of set member-

Table 2.1: GEO Optical Observation Measurements

5 minute interval			1 minute interval		
Time (UTC)	α (rad)	δ (rad)	Time (UTC)	α (rad)	δ (rad)
02:01:36	-1.5696	-0.09473	02:01:36	-1.5696	-0.09473
02:06:36	-1.5501	-0.09380	02:02:36	-1.5657	-0.09454
02:11:36	-1.5307	-0.09287	02:03:36	-1.5618	-0.09436

Table 2.2: LEO Optical Observation Measurements

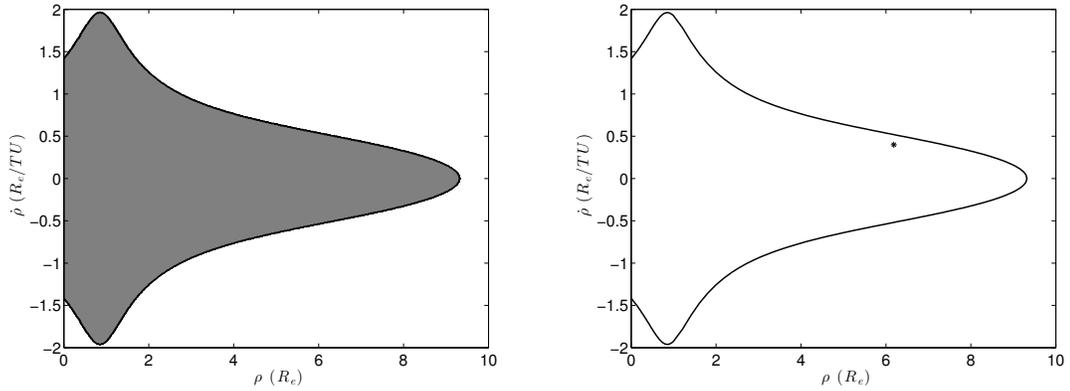
10 second interval			1 second interval		
Time (UTC)	α (rad)	δ (rad)	Time (UTC)	α (rad)	δ (rad)
02:01:36	-0.88785	0.30175	02:01:36	-0.88785	0.30175
02:01:46	-0.89577	0.30866	02:01:37	-0.88853	0.30233
02:01:56	-0.90370	0.31558	02:01:38	-0.88919	0.30292

Table 2.3: Observer Parameters (Atlanta, GA) and True Object States

Parameter	\hat{i}	\hat{j}	\hat{k}
Location (Lat, Lon, Alt)	33.77°N	84.39°W	340m
Cartesian Position (km)	-1359.0	5128.8	3527.9
LEO Object Position (km)	-6012.2	-316.8	3995.1
LEO Object Velocity (km/s)	-4.0881	1.0216	-6.0568
GEO Object Position (km)	-38845	16361	1417
GEO Object Velocity (km/s)	-1.1907	-2.2837	-0.1151

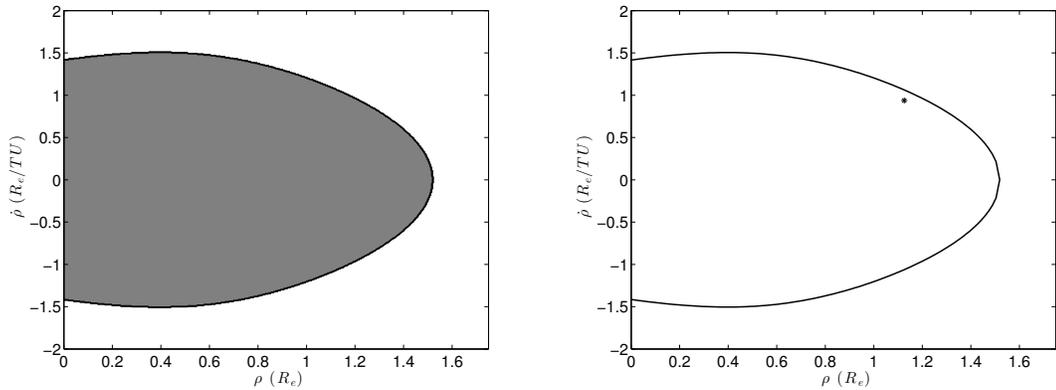
ship, the uncertainties must be taken into account while computing the admissible region. The measurements and parameters are assumed to have normal distributions with $\mu = 0$ and σ defined by the error associated with each quantity. The errors are given below and are consistent with or better than the performance of Raven-class telescopes used for SSA.

As stated, the analytical approach does not require discretization of the state space. However, in order to both visualize and compare the results with a traditional Monte Carlo approach a discretization of the state space is utilized. Since the approach is analytical, the accuracy of the PDF generated is only dependent upon the approximation of the linearization and the resolution of the discretization. For each observation case, the state space is discretized to a 500 by 500 grid for a total of 250,000 points. Equation (4.33) is then eval-



(a) Probability over the undetermined state space for (b) The true state (denoted with the marker) shown the energy constraint for the Case 1 GEO observation along with the boundary \mathcal{B}_1 of the admissible region. with no uncertainty effects

Figure 2.2: Admissible region for the Case 1 GEO observation



(a) Probability over the undetermined state space for (b) The true state (denoted with the marker) shown the energy constraint for the Case 1 LEO observation along with the boundary \mathcal{B}_1 of the admissible region. with no uncertainty effects

Figure 2.3: Admissible region for the Case 1 LEO observation

Table 2.4: Measurement Error and Parameter Uncertainty

Right Ascension uncertainty, σ_α	10 arcseconds
Declination uncertainty, σ_δ	10 arcseconds
Timing error, σ_t	0.0001 s
Position error (each axis), $\sigma_{\mathbf{o}}$	1 m
Velocity error (each axis), $\sigma_{\dot{\mathbf{o}}}$	1 m/s

uated at each point \mathbf{x}_u in the grid to generate $\mathbb{P}[\mathbf{x}_u \in \mathcal{A}R_1]$ for each GEO and LEO case, which can then be normalized to compare the approximate probability of set membership. The constraint curve \mathcal{B} is sampled by 100 points equally spaced from $\rho = 0$ to ρ_{\max} with $\pm\rho$ evaluated at each value of ρ . The quantity $\|\mathbf{x}_u - \mathbf{x}_{u,\mathcal{B}_\perp}\|$ is calculated at each point as the minimum distance from \mathbf{x}_u to the points in the sample set of points defining the constraint curve \mathcal{B} . The covariance $\mathbf{P}_{\mathbf{x}_u,\mathcal{B}_\perp}$ is calculated at each point \mathbf{x}_u by evaluating Eqn (3.72). This enables a direct comparison between the numerical and analytical approaches as the discretization grid used by each is identical. Figure 2.4 shows the resulting analytically generated probability of set membership. To visualize the results, Figure 2.5 shows the $\pm 3\sigma$ bounds of the approximate probability of set membership as well as the nominal constraint curve. These plots show how significant an effect the measurement uncertainties have on the admissible region. Parts of the state space are unaccounted for without the inclusion of uncertainty, thus, to appropriately use admissible regions, observation uncertainty must be accounted for. As an example application, the results shown in Figure 2.4 could be used to improve Multiple Hypothesis Tracking (MHT) techniques by providing more accurate initial probability distributions. Additionally, the $+3\sigma$ curve can be computed analytically as an upper bound to admissible region methods such as the GMM approach.

The GEO 1 minute interval and the LEO 1 second interval cases display large effects of uncertainty. The uncertainty at the location of maximum possible ρ is larger than the curvature of the constraint in both of these cases. Thus, there is a region in both cases where the expected symmetry of the $\pm n\text{-}\sigma$ curves is no longer true. Due to the large curvature of the constraint curve, these cases show that with large enough uncertainty effects or large enough curvature the assumptions made in the derivation of the approach can be violated. In the observation scenarios presented, as the observation interval approaches 1 minute for the GEO object and 1 second for the LEO object, the asymmetry of the $\pm n\text{-}\sigma$ curves shows that the assumptions are starting to be violated. However even in these regions where the curvature and sufficiently small uncertainty assumptions are violated, the % error in the

approach is less than 6% in the GEO 1 minute observation and less than 10% in the LEO 1 second observation.

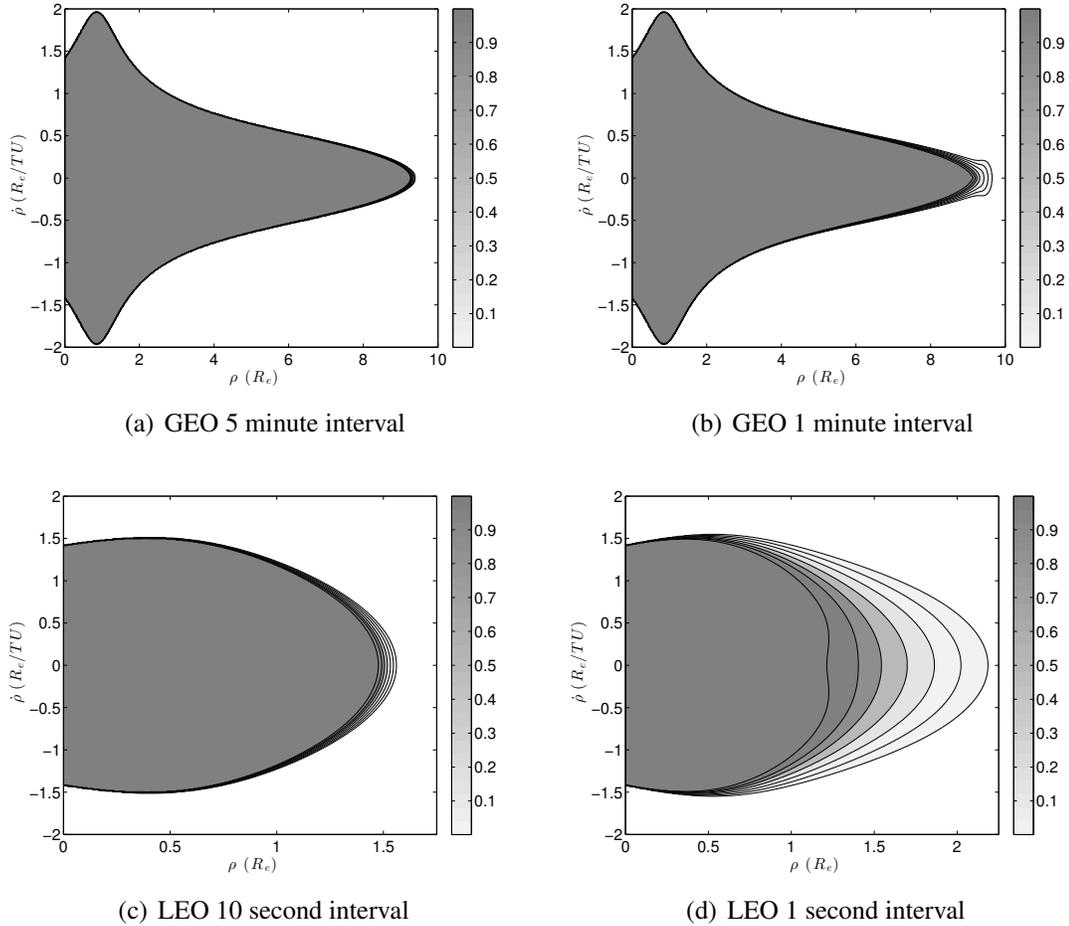


Figure 2.4: Analytical probabilities including uncertainty over the undetermined state space

2.7 Validation of the Approximation

The analytical approach presented is based on a first order approximation using a Taylor series expansion. Since it is an approximation, error between analytically and numerically obtained results are expected. The purpose of this section is to quantify this error.

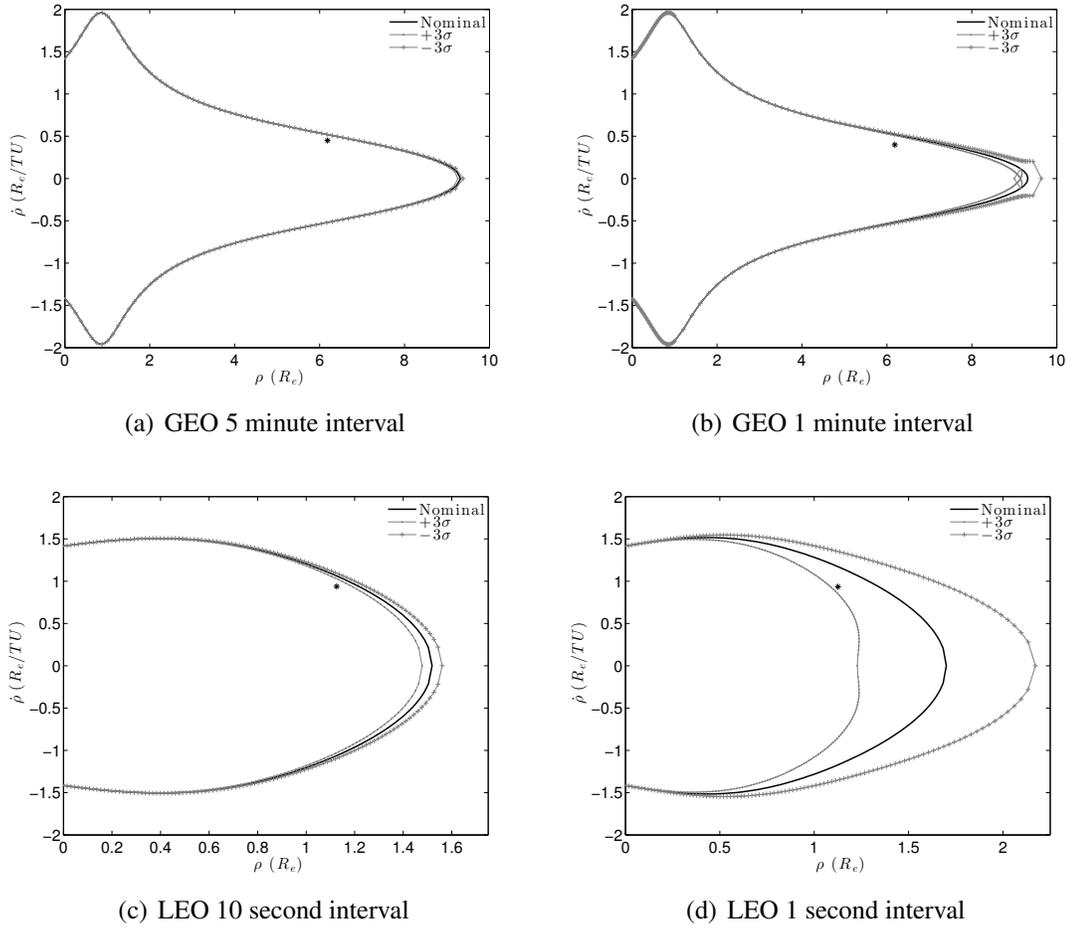


Figure 2.5: The $\pm 3\sigma$ bounds for the approximate analytical probability of set membership shown in addition to the nominal boundary. The true state is denoted with the marker

2.7.1 True PDF

Figure 2.4 shows the full probability of set membership with uncertainty calculated from the analytical approximation. Traditional approaches also use a discretization of the undetermined state space to generate an admissible region probability through Monte Carlo methods. The discretized grid generates a sampling of the state space. Since $\delta \mathbf{Z}$ is given by Eqn. (2.18) and the measurement and parameter errors are given in Table 5.1, a Monte Carlo simulation can provide the true distribution in $\delta \mathbf{x}_u$. For every point in the grid, randomly generated uncertainties with a Gaussian distribution are calculated and the constraint function evaluated. Each point in the grid acts as a bin recording the number of times the

constraint is met. The probabilities are then obtained by dividing the resulting bin values by the total number of trials. Since $\delta\mathbf{Z} \in \mathbb{R}^{12}$ for this example application, the Monte Carlo simulation requires a large sample size to sufficiently capture the distribution of the data. Thus, a Monte Carlo approach is limited by both the dimensionality of the problem and the resolution of the discretization in addition to long computation times. A Monte Carlo simulation with one million samples is generated for an identical discretized state space grid as the analytical approximation and each point in the grid is evaluated one million times. To expedite this process, the algorithm was programmed to utilize a Tesla C2050 GPU through the GPU programming functionality in MATLAB. The results from the analysis are shown in Figure 2.6. Even with the calculations being performed on the GPU, the run time for the Monte Carlo algorithm is approximately 2 hours. In contrast, the analytical probability of set membership is generated in 2 minutes on the same machine (specs listed in Table 2.5) without GPU processing. The analytical approach presented is not limited by dimensionality because it does not require a discretization to generate the PDF, the discretization is being shown to compare the two results. Furthermore, the computational requirements for the analytical approach presented are much less than those of Monte Carlo methods.

Table 2.5: Computer Specs

CPU	Intel Xeon E5520 @ 2.26GHz
GPU	Tesla C2050
OS	RHEL 5

A piecewise % Error calculation is shown in Figure 2.7. In the LEO cases, the % Error is below 10% in general. Note that the uncertainty effects in the GEO cases decreases as ρ decreases. In general, the GEO cases appear to have good agreement for large ρ but as ρ decreases so does the effect of uncertainty, reducing the comparability of the numerical and analytical results. Thus, the % Error calculations in the GEO cases is truncated to the region where the discretization is sufficient to compare the two. Figure 2.8 shows features of the % Error plots not easily seen in Figure 2.7. As seen in Figure 2.7a, it appears the GEO 5

minute observation is not in well agreement, but after discretizing the region for $\rho \geq 8$ it can be seen in Figure 2.8 that the analytical and numerical cases are in very well agreement with less than 2% error between them. Conversely, in the GEO 1 minute interval case the % Error is not symmetric due to the large effect of the uncertainty. The major effect seen in the GEO 1 minute interval case is that the high curvature allows the uncertainty orthogonal to the constraint to cross the plane of symmetry which contributes to high error near the location of ρ_{max} . In contrast, the GEO 5 minute interval case is expected to be in much better agreement since the errors are lower in this case. However, what is seen here is a limitation of the spatial resolution of the discretization. The entire distribution is contained in 3 pixels of the chosen discretization which is insufficient to capture the true distribution, contributing to the area near the mean being near 0% error but just outside being close to 20% error. In order to better assess how well the analytical probabilities agrees with the Monte Carlo probabilities without considering the limitations of the discretization chosen, an additional metric is defined in the following section.

2.7.2 Kullback-Leibler divergence

The comparison of the analytical and numerical results is fundamentally a comparison of two distributions. Thus, a measure of the similarity of the distributions can be used as a metric of the quality of the analytical approximation. Eqn. (4.33) defines the probability distribution along the lines normal to the constraint curve, in the expected value case, and both the numerical and analytical solutions should have identical curves for $\sigma = 0$. Equivalently, the analytical and numerical curves (i.e. $\sigma = 0$) without considering uncertainty should be identical. This implies that for this specific case the distribution in the \mathbf{p} direction at identical points on the should also be identical, enabling a comparison of the accuracy of the analytical solution. To define these points a line integral is defined for the $\sigma = 0$ curve such that for each point on the line, the distribution in the \mathbf{p} direction is obtained for both approaches. The line integral along a given constraint curve κ_i parameterized by $s \in [0, 1]$

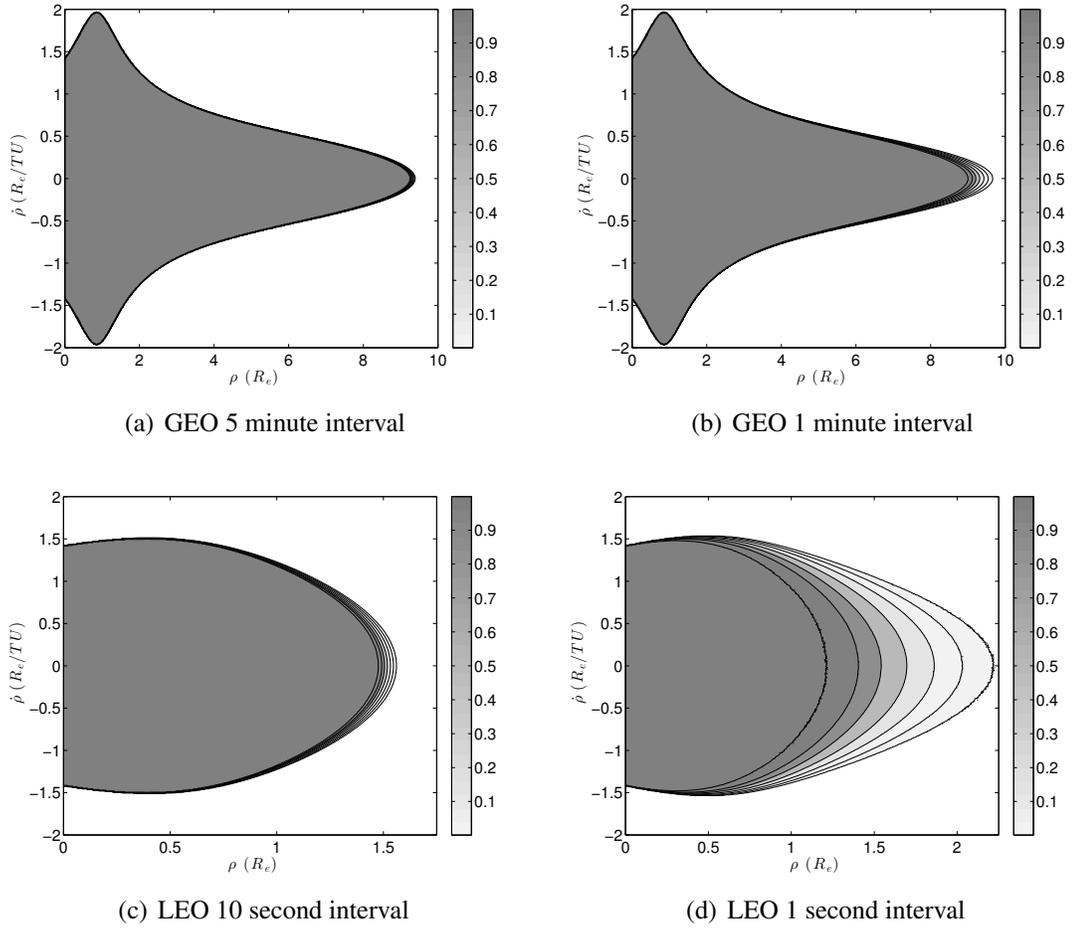


Figure 2.6: ‘True’ Monte Carlo numerical probabilities over the undetermined state space

is defined as follows. Defining the j^{th} point in the set of n states comprising \mathcal{B}_i as $\mathbf{x}_{j,\mathcal{B}_i}$, the total length, L , of κ_i for $\mathbf{x}_u \in \mathbb{R}^2$ can be defined by

$$L = \sum_{j=1}^{n-1} \|\mathbf{x}_{j+1,\mathcal{B}_i} - \mathbf{x}_{j,\mathcal{B}_i}\|_2 \quad (2.44)$$

where the set of points $\mathbf{x}_{j,\mathcal{B}_i}$ depend on the chosen discretization. Figure 2.9 illustrates this for an example admissible region. As shown, $\sum \Delta L$ gives the total length of the admissible region curve enabling the line integral s to be defined such that Δs , along the line s_i can be

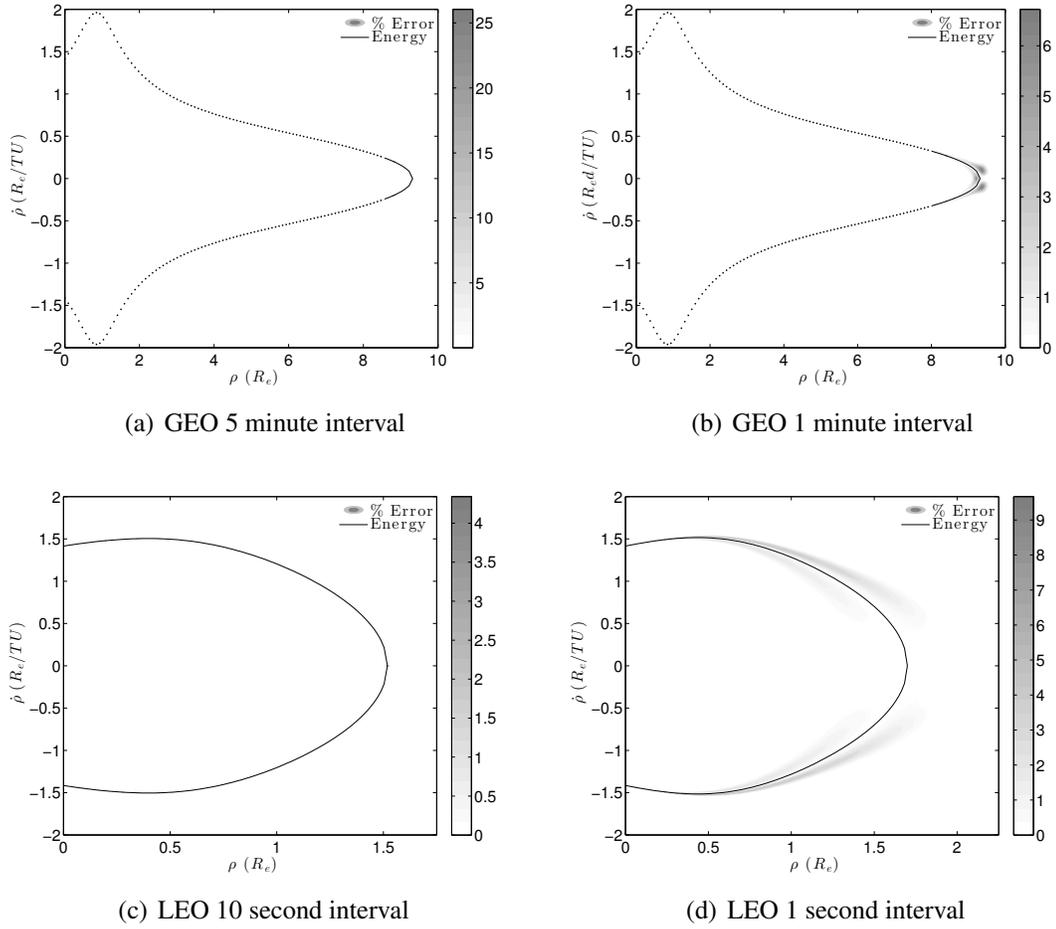


Figure 2.7: % Error between the analytical and numerical probabilities over the undetermined state space. The dotted lines in the GEO cases denote a region where the discretization is too small to produce meaningful % error results.

computed as

$$\begin{aligned} \Delta s_j &= \frac{\Delta L_j}{L} \\ &= \frac{\|\mathbf{x}_{j+1, B_i} - \mathbf{x}_{j, B_i}\|_2}{L} \end{aligned}$$

This procedure normalizes the constraint curve to the domain $s \in [0, 1]$ allowing for the analytically and numerically computed uncertainties to be conveniently compared. For the numerically generated probability of set membership the -3σ through 3σ contours are extracted and by generating a line integral for each contour. Then an interpolation through

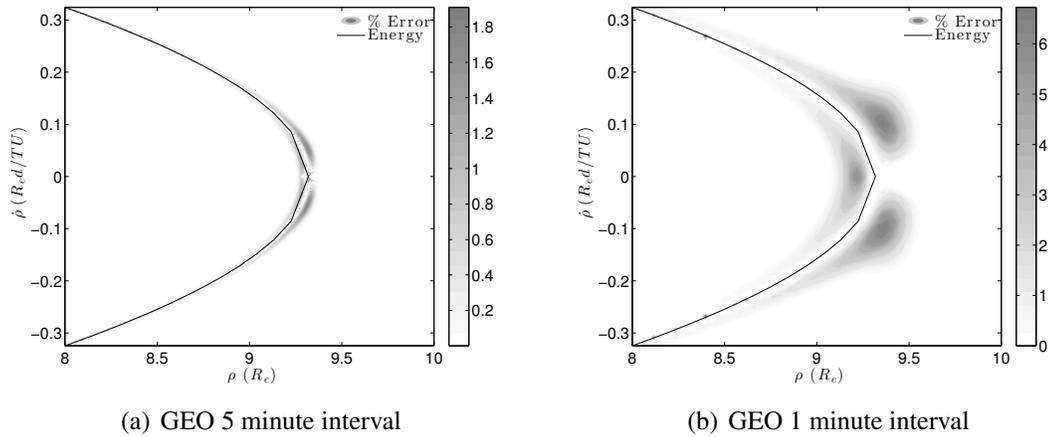


Figure 2.8: % Error focused on regions of interesting behavior.

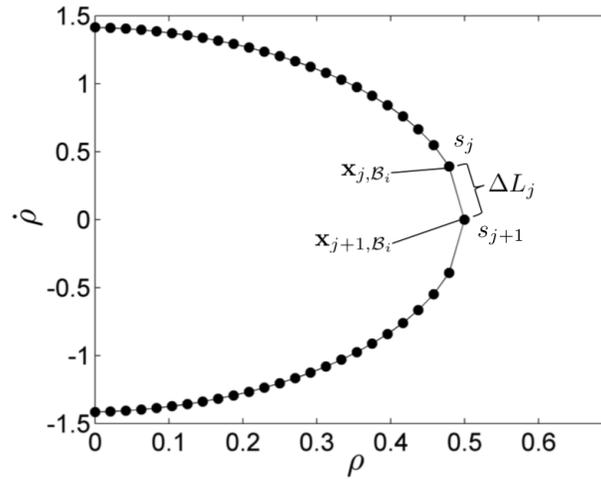


Figure 2.9: Two points along L used to calculate ΔL_j

the extracted points can be used to generate the numerical distribution corresponding to each point along s . The analytical distributions are fully defined at each point along s with zero mean and variance given by Eqn (3.72). With both the analytically and numerically computed distributions in the \mathbf{p} direction along s defined, the Kullback-Leibler divergence is utilized to numerically compare the distributions.

The Kullback-Leibler (KL) divergence metric, D_{KL} , also known as the information divergence, is a measure of the difference between two probability distributions P and Q [73]. It is a measure of information lost when Q is used to approximate P . It should be

noted that the KL divergence is not symmetric, implying D_{KL} from P to Q is not the same as from Q to P .

$$D_{kl}(P||Q) = \sum_i \ln \left(\frac{P(i)}{Q(i)} \right) P(i) \quad (2.45)$$

The equation given above is the form of the KL divergence for discrete probability distributions, where i is an index of the discrete probabilities being compared. Though the probability distributions approximated for the uncertainties are continuous, only discrete points from that distribution are known and thus the approximation of the KL divergence for the numerically and analytically computed distributions as discrete is appropriate.

Because both distributions necessarily have zero mean, quantifying error in the analytical approximation can be accomplished by examining error in the standard deviation. To assess how the values determined for D_{KL} relate to the quality of the approximation, two standard Normal distributions are generated as follows

$$P \sim \mathcal{N}(0, \sigma_P)$$

$$Q \sim \mathcal{N}(0, \sigma_P(1 + \%err))$$

where $\sigma_{\%err}$ is the percent error difference from σ_P . By varying the % error up to 100%, Figure 2.10 shows how D_{KL} changes for standard normal distributions. Using this plot, an interpolation can be defined to approximate the % error that corresponds to a given D_{KL} . The resulting interpolation is then used along with Eqn. (2.45) such that both % error and D_{KL} can be evaluated. The numerically computed distributions are defined as Q and the analytically computed distributions are defined as P . Thus at each point along s_i both P and Q are used to compute D_{KL} and, using the inverse of the relationship shown in Figure 2.10, the corresponding percent error between the numerical and analytical cases is computed.

Since the LEO 1 second interval showed the highest effect of the uncertainty, this case is chosen for further analysis with the Kullback-Leibler approach. It is expected that as the

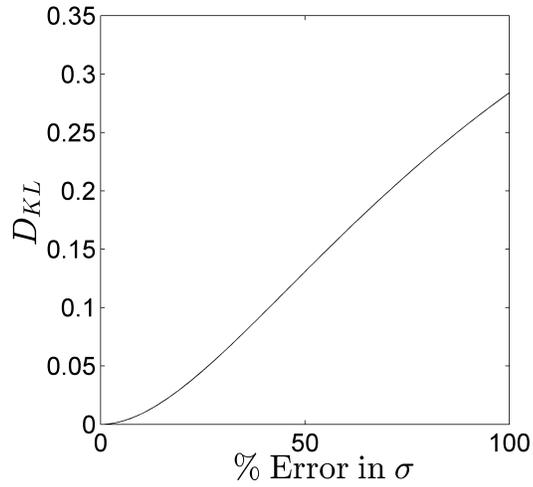
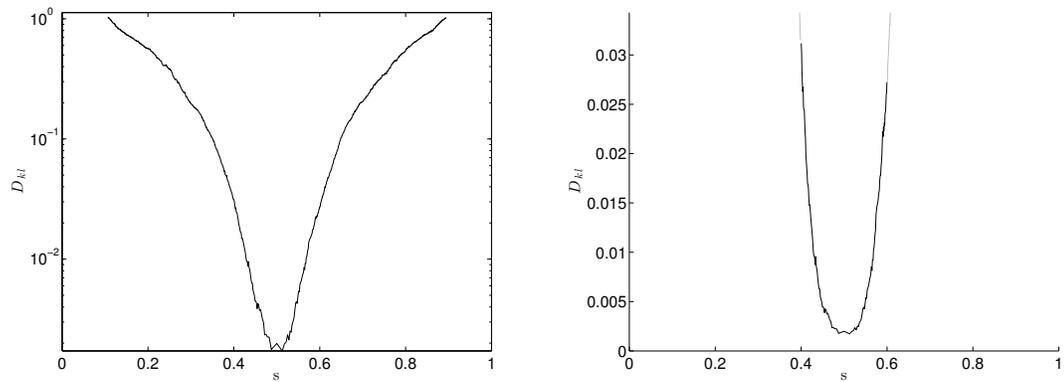


Figure 2.10: D_{KL} as a function of % error for a standard Normal distribution



(a) Near the ends of s , D_{KL} behaves poorly due to the effects of discretization of the state space. (b) Utilizing $s \in [0.4, 0.6]$ avoids some of the effects of discretization

Figure 2.11: Plots of D_{KL} comparing the analytical and Monte Carlo simulation

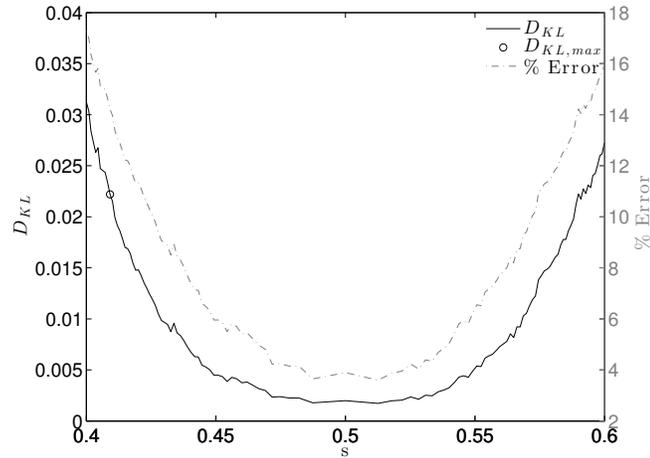


Figure 2.12: D_{KL} over the well behaved region of the line s plotted along with % Error

uncertainty decreases the accuracy of the approximation increases, thus if the accuracy of the approach is validated for the highest uncertainty case, it is expected to be validated for the other cases as well. Figure 2.11a shows D_{KL} for the full length of s . However, at the ends of the distribution D_{KL} increases substantially towards infinity. This increase is due to the effects of the discretization as well as the effects of uncertainty. Recalling Figure 2.4, as the constraint approaches $\rho = 0$, the distribution narrows. This can be explained probabilistically as the variances computed by Eqn (3.72) decrease as $\rho \rightarrow 0$. As the distribution narrows with a constant, predefined discretization, the width of the distribution becomes too small for the discretization to adequately sample from the distribution. This can be avoided by selecting a discretization size based on the smallest distribution width, however such an approach would be computationally infeasible. Alternatively, a discretization could be selected such that a significant portion of s provides adequate sampling of the probability distribution in that region. Thus, 2.11b shows D_{KL} for the region of $s \in [0.4, 0.6]$. Over this range, the values of D_{KL} are better behaved implying that over this range of s the discretization is sufficient for analysis.

Analyzing the results of Figure 2.12, the greatest Kullback-Leibler divergence is expected to be found to be at $s = 0.5$. Since this problem is derived using a Taylor series

approach, thus the largest errors are expected in the high curvature regions of the approximation. However, as the analytical results are continuous and the numerical results are discrete, $s = 0.5$ has lower error than its neighboring region since the normal direction is solely in the ρ axis, thus effects of the discretization are minimized. For this reason, the error increases as the slope of the normal direction increases and Figure 2.12 confirms this. In order to minimize this the selected discretization was 3500 points in each axis which required 11 hours to generate. It is expected that a finer discretization will show the expected behavior however such a discretization is computationally infeasible. However it is found that outside of $0.40 \leq s \leq .60$ is where the discretization starts to fail so $s = 0.41$ is chosen for further analysis. The analytically and numerically obtained distribution functions are shown in Figure 2.13 for $s = 0.41$. Note that the numerical distributions are fit to the points from the Monte Carlo simulation as a method of comparison. The Monte Carlo results are indeed non-Gaussian but can be approximated as Gaussian for the Kullback-Leibler approach. It is evident in Figure 2.13a that the means of the two approaches are well in agreement. This is as expected as the uncertainties belong to a distribution with zero mean. Further, it is shown in Eqn. (2.27) that the mean of the distribution in the undetermined states should be zero. At this point, the location of the highest error between the approaches, the main difference between the numerical and analytical approaches is the apparent overestimation of the standard deviation from the analytical approach. This is evident due to the narrower peak of the numerical distribution, indicating the true distribution has a lower standard deviation than what is approximated analytically. The contributions of each source of uncertainty or error can also be quantified at $s = 0.41$ using Eqn (2.31). Table 2.6 shows the sensitivity of the overall uncertainty to each contributing factor and it is apparent that the angular accuracy and timing accuracy are the most influential sources of error.

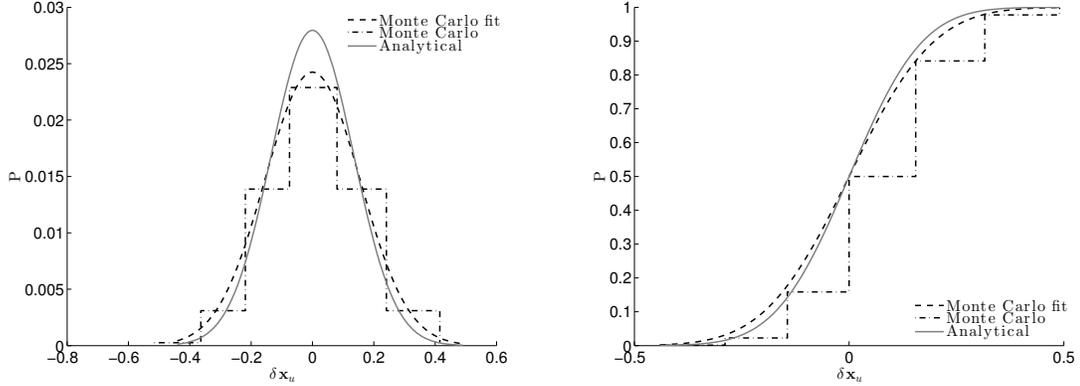
Using the inverse of the relationship between D_{KL} and % error shown in Figure 2.10, Figure 2.12 plots both D_{KL} and % error on a single axis. The location of the expected

Table 2.6: Sensitivity of Undetermined State Uncertainty to Measurement Error and Parameter Uncertainty

Contribution	Parameter
$\sigma_{\alpha_1} = 17.20\%$	Right Ascension Measurement 1
$\sigma_{\alpha_2} = 17.19\%$	Right Ascension Measurement 2
$\sigma_{\delta_1} = 15.77\%$	Declination Measurement 1
$\sigma_{\delta_2} = 15.78\%$	Declination Measurement 2
$\sigma_{t_1} = 17.02\%$	Time of Measurement 1
$\sigma_{t_2} = 17.02\%$	Time of Measurement 2
$\sigma_{\mathbf{o}_x} = 1.58 \times 10^{-6}\%$	X Component of Observer Position
$\sigma_{\mathbf{o}_y} = 4.21 \times 10^{-7}\%$	Y Component of Observer Position
$\sigma_{\mathbf{o}_z} = 2.92 \times 10^{-6}\%$	Z Component of Observer Position
$\sigma_{\dot{\mathbf{o}}_x} = 4.22 \times 10^{-6}\%$	X Component of Observer Velocity
$\sigma_{\dot{\mathbf{o}}_y} = 7.60 \times 10^{-6}\%$	Y Component of Observer Velocity
$\sigma_{\dot{\mathbf{o}}_z} = 1.19 \times 10^{-5}\%$	Z Component of Observer Velocity
100%	

maximum D_{KL} is marked with a circle on the plot and at this point the % error of the approach is 3.638%. Overall, the approach appears to be in good agreement with the numerical results. It should be noted that several effects contribute to the accuracy of the analytical approach. The first order approximation used to derive the approach contributes to the errors presented. A numerical method is capable of capturing the higher order effects whereas in the presented analytical approach they are assumed negligible, so a certain amount of error is to be expected. However, it is shown that this error is relatively small and additional factors can contribute to lowering the error. Lower uncertainties will improve the analytical approximation, which can be seen to some extent in Figure 2.7. Additionally, having larger time steps between observations will improve the agreement of the analytical and numerical PDFs reducing error. Uncertainty increases as the time between observations decreases. Larger uncertainties contribute more to the higher order effects neglected in the

approach. With longer time steps, uncertainty decreases overall and the error between the numerical and analytical results will decrease as well.



(a) Comparison of the analytical probability distribution with a fit of the Monte Carlo (b) Comparison of the cumulative distribution with a fit of the Monte Carlo

Figure 2.13: Validation of the probability and cumulative distribution functions for the numerical and analytical cases at $s = 0.41$ (peak D_{KL} error)

2.8 Joint probability of set membership with Two Constraints

Since many underdetermined systems can be subject to several constraint hypotheses, the joint probability of set membership is of much interest. The approach presented in Section II.C for generating a joint probability of set membership is applied to the same observation information \mathbf{z} by considering both an energy constraint as well as a constraint on the periapse radius of the object. For ground observers the periapse radius of the object can be constrained using the following approach [45]. The apparent angular rate of the object with respect to the observer is given by the proper motion η where

$$\eta = \sqrt{w_2} \quad (2.46)$$

Thus the range-rate can be equivalently written as $\dot{\rho} = \rho\eta$ and thus it is required that

$$|\dot{\rho}\tau| \geq R_e$$

$$|\rho\eta\tau| \geq R_e$$

where R_e is the radius of the Earth. Additionally, $\tau = \rho/|\dot{\rho}|$ which, after simplification, leaves

$$\frac{\rho^2\eta}{|\dot{\rho}|} \geq R_e$$

which can be written as a quadratic constraint in ρ .

$$\kappa_2(\mathbf{x}_u, \mathbf{z}) = \frac{R_e|\dot{\rho}|}{\eta} - \rho^2 \leq 0 \quad (2.47)$$

Eqn. (2.47) represents the second constraint $\kappa_2(\mathbf{x}_u)$. Following the same procedure as with the energy constraint, $\frac{\partial \kappa_2}{\partial \mathbf{x}_u}$ is given by

$$\begin{aligned} \frac{\partial \kappa_2}{\partial \rho} &= -2\rho \\ \frac{\partial \kappa_2}{\partial \dot{\rho}} &= \frac{\text{sgn}(\dot{\rho})R_e}{\eta} \end{aligned}$$

where $\text{sgn}()$ is the sign function and $\frac{\partial \kappa_2}{\partial \mathbf{z}}$ are given in the Appendix.

The approximate and true Monte Carlo probability of set membership for $\kappa_1(\mathbf{x}_u)$ and $\kappa_2(\mathbf{x}_u)$ are generated and the joint PDF is computed from the element-wise multiplication of the two distributions. Figure 2.14 shows the probability of set membership generated for the energy constraint alone. The joint approximate analytic probability of set membership is computed using Eqn. (4.33) and shown in Figure 2.15 from the multiplication of the PDF shown in Figure 2.14 with the probability of set membership from Figure 2.4. A numerical Monte Carlo joint probability of set membership is generated as well and the results are

shown in Figure 2.16. To generate a metric of comparison for the joint PDF, a % error calculation is performed assuming the numerical distribution as the true distribution. Due to the very tight distribution for the periapse constraints in both GEO observations as well as the LEO 10 second interval observation, the % error calculation is only presented for the LEO 1 second interval observation. Since the LEO 1 second interval give the largest effect of uncertainty, it is expected that it will have the highest error overall and it can be expected that the overall % error in the rest of the observations should be lower. Figure 2.17 shows the computed % error between the analytical probability of set membership shown in Figure 2.15 and that of Figure 2.16 for the LEO 1 second interval observation. Overall, the joint probability of set membership approximated from the analytical approach agrees well with the numerical case. For solutions not in the vicinity of the intersection, the analytical approach almost exactly matches, with less than 10% error in these regions. At the intersections, however, the approximation has up 25% error. For reference, the two admissible region curves are plotted as well in Figure 2.17 highlighting the location of the intersection. It is stated previously that one of the contributions to the error in the analytical approach is the first order approximation. However, another assumption made is that the constraints are independent probabilistically which is inherently not true for the selected constraints. The question then, is how to address dependent constraints in a computationally tractable way.

2.9 Joint Probability for Dependent Constraints

The previous section highlights the error induced by assuming independence of all hypothesized constraints. The error in the boundary can exceed 20% under this assumption which indicates there must exist a more correct way to find the joint probability of set membership for a joint admissible region. The full consideration of the joint admissible region

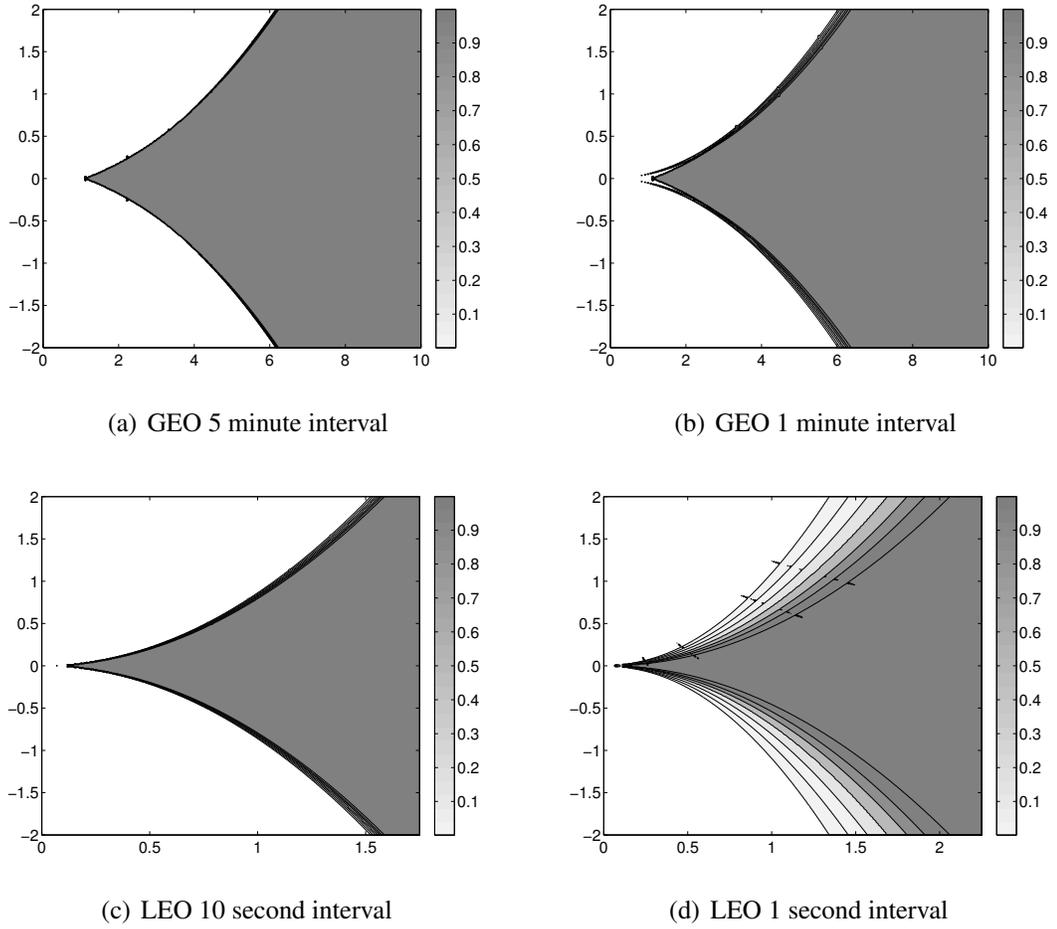


Figure 2.14: Analytical probabilities including uncertainty over the undetermined state space for each case

probabilities of set membership is based on the inclusion-exclusion principle

$$|\cup_{i=1}^c A_i| = \sum_{k=1}^c (-1)^{k+1} \left(\sum_{1 \leq i_1 < \dots < i_k \leq c} |A_{i_1} \cap \dots \cap A_{i_k}| \right) \quad (2.48)$$

where A_i are the sets of interest and $|\cdot|$ denotes the cardinality of the set [74]. Eqn. (2.48) is a combinatorial methodology which is useful for finding partitions and cuts of sets [75],

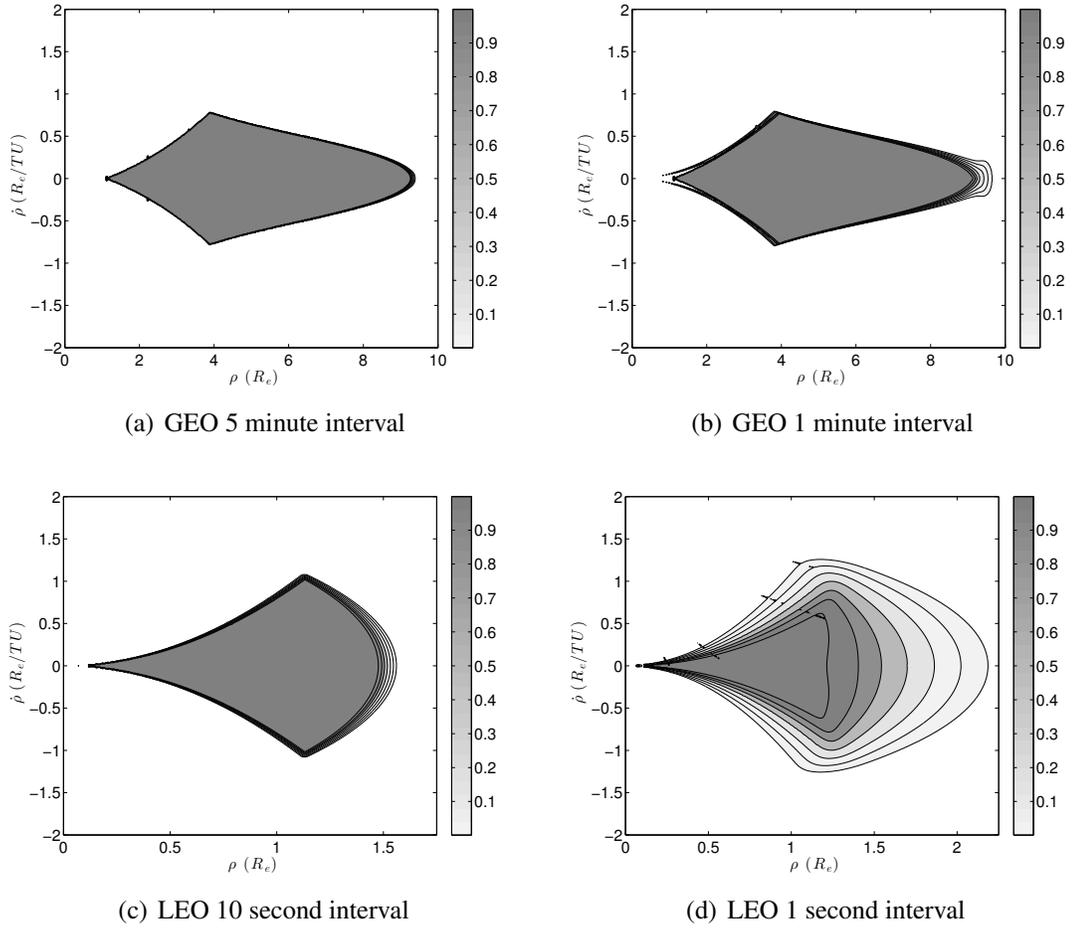


Figure 2.15: The joint probability as obtained analytically for the energy and periape constraints

which when applied to probability theory yields, in general form,

$$\begin{aligned}
 P(\cup_{i=1}^c A_i) &= \sum_{i=1}^c P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) \\
 &\quad - \dots + (-1)^{c-1} P(\cap_{i=1}^c A_i)
 \end{aligned}
 \tag{2.49}$$

To apply Eqn. (2.49) to the admissible region problem, let each set A_i represent an admissible region \mathcal{R}_i . Then the probabilities shown represent the probability that a given

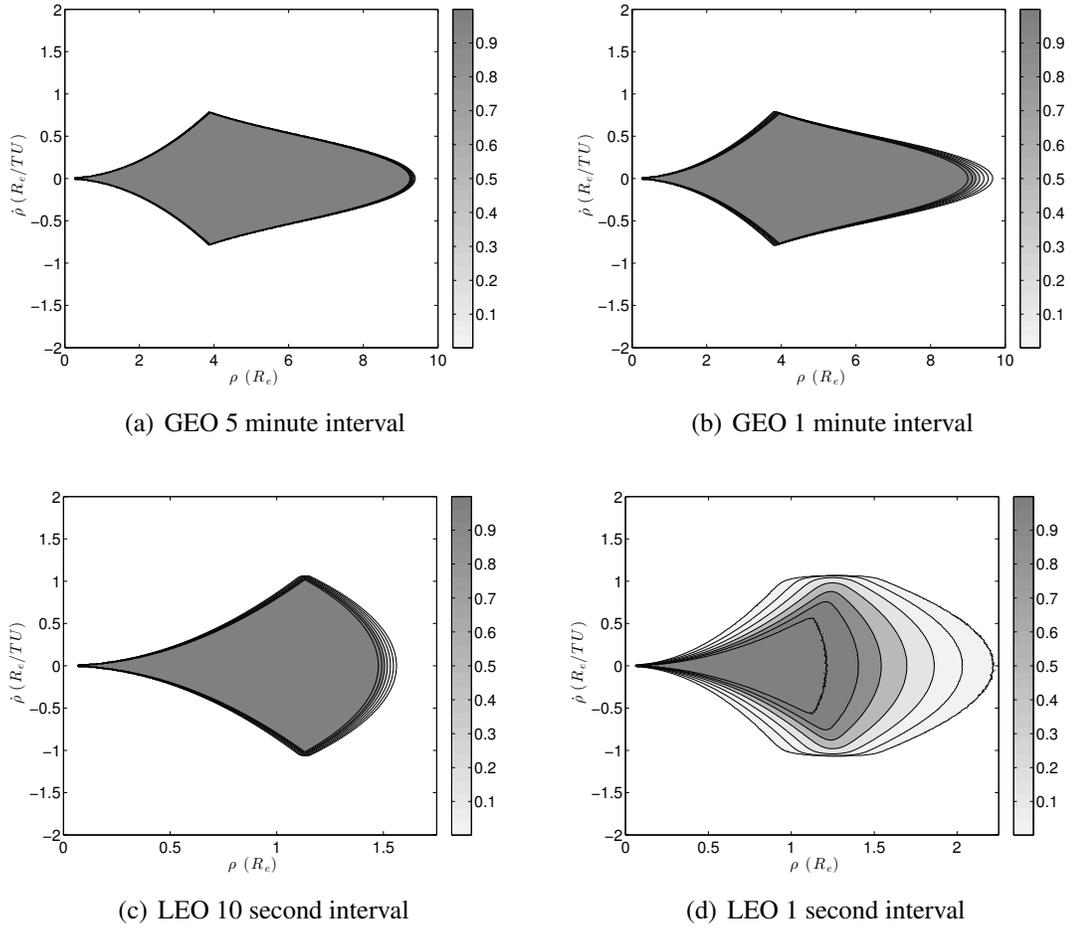


Figure 2.16: The joint probability as obtained from the Monte Carlo simulation for the energy and periaapse constraints

undetermined state \mathbf{x}_u lies in the admissible region and Eqn. (2.49) can be written as

$$\begin{aligned}
 P(\mathbf{x}_u \in \cup_{i=1}^c A_i) &= \sum_{i=1}^c P(\mathbf{x}_u \in A_i) - \sum_{i < j} P(\mathbf{x}_u \in (A_i \cap A_j)) + \sum_{i < j < k} P(\mathbf{x}_u \in (A_i \cap A_j \cap A_k)) \\
 &\quad - \dots + (-1)^{c-1} P(\mathbf{x}_u \in \cap_{i=1}^c A_i)
 \end{aligned} \tag{2.50}$$

Note that the term of interest appears on the right hand side of Eqn. (2.50). That is, the probability quantity desired is $P(\mathbf{x}_u \in \cap_{i=1}^c \mathcal{R}_i)$. Consider the case when $c = 2$, then Eqn.

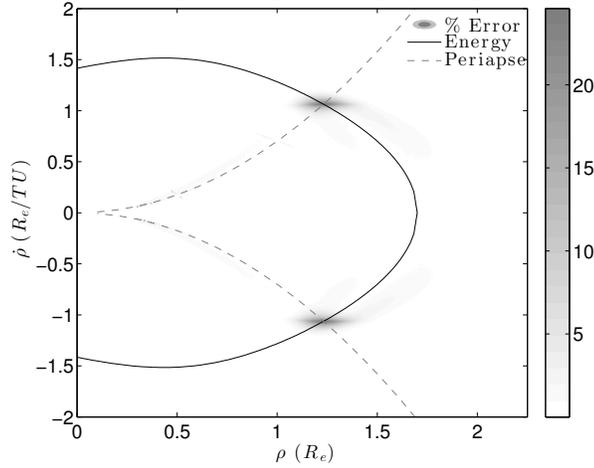


Figure 2.17: The percent error between the numerically computed joint probability of set membership and the analytically computed joint probability of set membership for the LEO 1 second interval case.

(2.50) becomes

$$P(\mathbf{x}_u \in \mathcal{R}_1 \cap \mathcal{R}_2) = P(\mathbf{x}_u \in \mathcal{R}_1) + P(\mathbf{x}_u \in \mathcal{R}_2) - P(\mathbf{x}_u \in \mathcal{R}_1 \cup \mathcal{R}_2) \quad (2.51)$$

The term $P(\mathbf{x}_u \in \mathcal{R}_1 \cup \mathcal{R}_2)$ can be described without loss of generality as a piecewise function

$$P(\mathbf{x}_u \in \mathcal{R}_1 \cup \mathcal{R}_2) = \begin{cases} 1 & (\kappa_1 \leq 0) \wedge (\kappa_2 \leq 0) \\ 1 & (\kappa_1 \leq 0) \wedge (\kappa_2 > 0) \\ 1 & (\kappa_1 > 0) \wedge (\kappa_2 \leq 0) \\ z & (\kappa_1 > 0) \wedge (\kappa_2 > 0) \end{cases} \quad (2.52)$$

where \wedge is the logical "and" operator and $0 \leq z \leq 1$. It is shown that $z > 0$ only when systemic uncertainties are taken into consideration where it is possible to have a non-zero probability of set membership outside of the nominal constraint boundary, but only in close vicinity to the boundary [76]. By application of the Bonferroni inequality it is then possible

to bound Eqn. (2.53) as follows

$$P(\mathbf{x}_u \in \mathcal{R}_1 \cap \mathcal{R}_2) \geq P(\mathbf{x}_u \in \mathcal{R}_1) + P(\mathbf{x}_u \in \mathcal{R}_2) - 1 \quad (2.53)$$

without needing to determine an analytical approximation of $P(\mathbf{x}_u \in \mathcal{R}_1 \cup \mathcal{R}_2)$ [77]. Without loss of generality, this simplification through Bonferroni's inequality may then be applied to any number of constraint hypotheses.

$$P(\mathbf{x}_u \in \cap_{i=1}^c \mathcal{R}_i) \geq \left[\sum_i^c P(\mathbf{x}_u \in \mathcal{R}_i) \right] - (c - 1) \quad (2.54)$$

This provides a concise way to handle two or more hypothesized constraints without the increased approximation error induced by assuming the constraint hypotheses are independent.

Recall the errors shown in Figure 2.17 under the incorrect assumption that the constraint hypotheses are independent. Using instead Eqn. (2.54) to define the joint probability of set membership for the energy and periapse radius constraints these errors are reduced substantially. Figure 2.18 shows the resulting probability of set membership function over the joint admissible region using the Bonferroni approach and as can be seen in Figure 2.19 the percent error between the analytical and numerical approach is significantly improved. There are still regions of large percent error near the location of the intersection, however this is not solely due to the application of Bonferroni's inequality. Examining Figure 2.7d shows that there is increased error in the approximation of the energy constraint probability of set membership even without considering additional constraints, noting again that some of this error is induced by the limitations of computational complexity and the discretization used for the Monte Carlo simulation. This suggests that the errors in the joint probability of set membership functions are largely bounded by the errors of the independent membership functions for each constraint hypothesis.

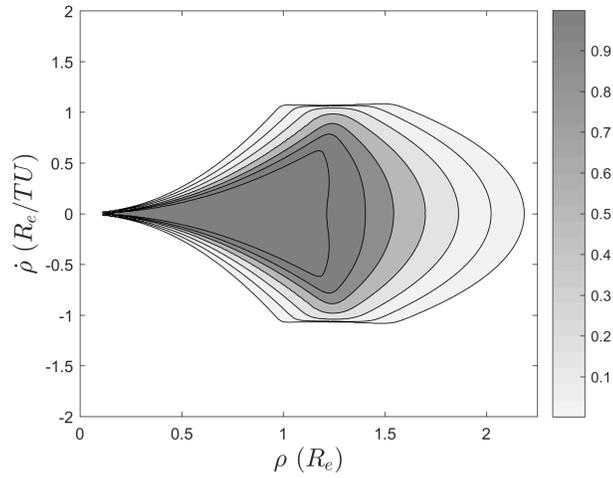


Figure 2.18: The approximate joint probability of set membership for the LEO 1 second interval case.

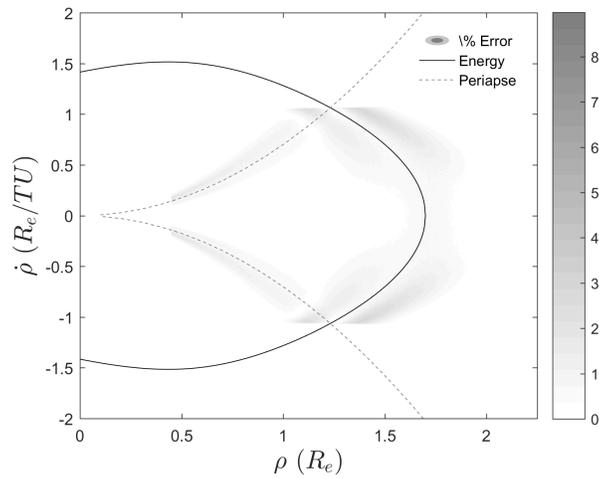


Figure 2.19: The percent error between the numerically and analytically computed joint probability of set membership for the LEO 1 second interval case.

2.10 Conclusions

This work generalizes the admissible region method for initial orbit determination and presents a method by which measurement and observer uncertainties can be rigorously included in the admissible region. From a general measurement model, the admissible region method is generalized to define the probability of set membership function associated with a given admissible region. Using a Taylor series expansion, an analytical expression is derived which enables the generation of an approximate analytical model of the true probability of set membership. This approximation is shown to be in good agreement with a numerically computed ‘true’ probability of set membership. The methodology is then extended to include multiple constraint hypotheses enabling a closed-form approximation of a joint admissible region probability of set membership. The method presents an improved initial orbit determination initialization with a single observation sequence by generating an initial probability of set membership, which may then be used as a PDF in a filter, that correctly incorporates observation and assumed parameter uncertainties.

CHAPTER 3

PROBABILISTIC INTERPRETATION OF THE ADMISSIBLE REGION

The previous chapter introduces the inclusion of systemic uncertainties in the admissible region by defining a membership function which is essentially a Gaussian CDF defined by how the errors in the system map to the constraint surface. However, it does not address the problem of how to assign prior probability density to the resulting ‘fuzzy’ set for use in bootstrapping estimation methods. The purpose of this chapter is to define guidelines for how prior probability densities are assigned for the uncertain admissible region based on rigorous application of probability theory and the probability transformation theorem. The general theory of probability transformations is an exhaustively studied topic in statistics and probability with a wide range of applications [69] [78] [79]. The fundamental results contained in this chapter are demonstrated through examples which highlight the ambiguities present in assigning prior density in underdetermined systems.

3.1 General Probability Transformations

Given the PDF $f_{\mathbf{X}} : \mathbb{R}^n \rightarrow \mathbb{R}_+$ of a random variable $\mathbf{X} \in \mathbb{R}^n$, $\mathbf{x} \sim f_{\mathbf{X}}(\mathbf{x})$, the cumulative distribution function (CDF) can be written as

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}[\mathbf{X} \leq \mathbf{x}] = \int_A f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \quad (3.1)$$

where the volume of integration is given by $A = (-\infty, x_1] \times \cdots \times (-\infty, x_n]$. Define a transformation $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ where $n \geq m$. Applying the transformation $\tilde{\mathbf{x}} = \mathbf{g}(\mathbf{x})$, the CDF for the transformed variable is obtained using integration by substitution and is given

by

$$F_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = \mathbb{P}[\tilde{\mathbf{X}} \leq \tilde{\mathbf{x}}] = \int_{\tilde{A}} f_{\mathbf{X}}(\mathbf{g}^{-1}(\tilde{\mathbf{x}})) \cdot \text{abs} \left(\left| \frac{\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}} \right| \right) \quad (3.2)$$

where $\tilde{A} = (-\infty, \tilde{x}_1] \times \cdots \times (-\infty, \tilde{x}_n]$ and $\text{abs}(|\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})/\partial \tilde{\mathbf{x}}|)$ is the determinant of the Jacobian matrix and the absolute value ensures $f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}})$ is non-negative for all values of $\tilde{\mathbf{x}}$ [69]. The integrand of Eqn. (3.2) is by definition the PDF of $\tilde{\mathbf{X}} = \mathbf{g}(\mathbf{X})$ and the following foundational theorem in multivariate statistics gives the PDF of the transformed variable.

Theorem 1 (Transformation theorem for continuous random variables [69]). *Given a PDF $f_{\mathbf{X}}(\mathbf{x})$ and a left-invertible transformation $\tilde{\mathbf{x}} = \mathbf{g}(\mathbf{x})$ the PDF of the transformed variable $f_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}}_u)$ is given by*

$$f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = \begin{cases} f_{\mathbf{X}}(\mathbf{g}^{-1}(\tilde{\mathbf{x}})) \cdot \text{abs}(|\frac{\partial}{\partial \tilde{\mathbf{x}}} \mathbf{g}^{-1}(\tilde{\mathbf{x}})|) & \text{for } \tilde{\mathbf{x}} \in \mathcal{R}(\mathbf{g}(\tilde{\mathbf{x}})) \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

Proof. The proof of Theorem 1 is given in [69]. □

This transformation of \mathbf{X} into $\tilde{\mathbf{X}}$ must also satisfy

$$F_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = F_{\mathbf{X}}(\mathbf{g}^{-1}(\tilde{\mathbf{x}})) \quad (3.4)$$

where $F_{(\cdot)}$ denotes the CDF over $\tilde{\mathbf{x}}$ or \mathbf{x} [80]. This implies that for a given transformation $\tilde{\mathbf{x}} = \mathbf{g}(\mathbf{x})$, the CDF must not be changed. In other words, if the CDF is known for \mathbf{X} then the CDF is known for $\tilde{\mathbf{X}}$.

Corollary 1 (Equivalence of CDFs). *Given a known CDF $F_{\mathbf{X}}(\mathbf{x})$ for \mathbf{x} and a once differentiable and right-invertible transformation $\tilde{\mathbf{x}} = \mathbf{g}(\mathbf{x})$, the CDF $F_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}})$ for $\tilde{\mathbf{x}}$ must satisfy $F_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = F_{\mathbf{X}}(\tilde{\mathbf{x}})$.*

Proof. The proof of Corollary 1 follows directly from the derivation and analysis of Eqn. (3.1). By definition Eqn. (3.1) is equal to Eqn. (3.2) and thus Eqn. (3.4) must hold. □

3.2 Admissible Region Transformations

The purpose of the following subsections is to prove why, in general, an admissible region prior cannot be transformed since a general transformation of the admissible region yields the assignment of different prior probabilities, violating the Principle of Transformation Groups. The first subsection shows the application of the derivation of Eqn. (3.3) to the admissible region problem. Then the necessary conditions for an admissible region prior to be transformed based on the definition of an admissible region are defined, followed by a discussion of the limitation of practical transformations satisfying these necessary conditions. The second subsection considers the case when an admissible region prior is not considered to be uniform. The third subsection discusses the observability condition in the admissible region problem and discusses when Eqn. (3.3) may be applied to an a posteriori PDF based on an admissible region without any additional conditions.

3.2.1 Observability of Admissible States

The observations relevant for the admissible region approach are typically short enough relative to the dynamics that a continuum of states could have generated the measurements observed. In optical observations this is readily realized as a short streak from which not enough information is available to obtain a full state estimate. The admissible region approach allows the continuum of possible solutions for an underdetermined system to be bounded based on hypothesized constraints as described above. The continuum of possible solutions for an underdetermined system indicates that the system is unobservable. The undetermined states may be considered the unobservable states, and the admissible region must be a subset of this unobservable subspace. It is then desired to determine the observability of the dynamical system being observed since, for an admissible region to exist, the system must be unobservable. Conversely, if the dynamical system can be shown to be observable then the admissible region is not defined.

Consider the general nonlinear dynamical system and measurement model

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \quad (3.5)$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}; \mathbf{k}, t) \quad (3.6)$$

where the measurement function is defined as $\mathbf{h} : \mathbb{R}^n \times \mathbb{R}^z \times \mathbb{R} \rightarrow \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector, $\mathbf{x} \in \mathbb{R}^n$ is the state, $\mathbf{k} \in \mathbb{R}^z$ is the parameter vector that may include the observer state and any other necessary parameters, and t is the time. Several approaches exist to show observability of the general system given by Eqns. (3.5) and (3.6). For linear systems, the conditions for observability of this system, over a time interval $t \in [t_0, t_f]$, can be assessed by the observability gramian $\mathbf{P} \in \mathbb{S}_+^{n \times n}$ [81] which is given in most general form as

$$\mathbf{P}(t_f, t_0, \mathbf{x}(t)) = \int_{t_0}^t \Phi^T(\tau, t_0) \frac{\partial \mathbf{h}(\mathbf{x}(\tau); \mathbf{k}, \tau)^T}{\partial \mathbf{x}(\tau)} \frac{\partial \mathbf{h}(\mathbf{x}(\tau); \mathbf{k}, \tau)}{\partial \mathbf{x}(\tau)} \Phi(\tau, t_0) d\tau \quad (3.7)$$

where $\Phi : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ is the state transition matrix (STM). The observability gramian as defined above is also valid for linearized system in a region near the point of linearization, however it does not provide information of observability of other states. The rank of the above observability gramian gives the dimension of the observable subspace of the system along $\mathbf{x}(t), t \in [t_0, t_f]$. A point in state space $\mathbf{x}(t)$ is observable if and only if $\text{rank}[\mathbf{P}(t_f, t_0, \mathbf{x}(t))] = n$. If $\text{rank}[\mathbf{P}(t_f, t_0, \mathbf{x}(t))] < n$ then there is an unobservable subspace which is realized as $\mathcal{N}(\mathbf{P}(t_f, t_0, \mathbf{x}(t)))$, the nullspace of the observability gramian about $\mathbf{x}(t)$ over the time interval $t \in [t_0, t_f]$, and a state estimate admits a continuum of solutions that generate the same measurement sequence.

Hermann and Krener show that for a nonlinear system, the necessary condition for local observability is that a one-to-one mapping exist between the output (and derivatives of the output) and the input or initial conditions [82]. It is shown in [83] that this condition may be sufficiently satisfied by linearizing Eqns. (3.5) and (3.6) about a reference trajectory and

showing that for any reference trajectory in the domain $(\mathbf{x}_r(t), \mathbf{k}_r(t)) \in \mathbf{W}$, the linearized system is observable. For the admissible region approach \mathbf{W} is defined as

$$\mathbf{W} = \{(\mathbf{x}_r(t), \mathbf{k}_r(t)) : \mathbf{x}_{r0,u} \in \mathcal{R}, \mathbf{k}_{r0,d} = \mathbf{h}^{-1}(\mathbf{y}, \mathbf{k}, t_0)\} \quad (3.8)$$

Linearization of Eqns. (3.5) and (3.6) about trajectories in \mathbf{W} yields

$$\delta \dot{\mathbf{x}}(t) = \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_r(t), \mathbf{k}_r(t)} \delta \mathbf{x}(t) \quad (3.9)$$

$$\delta \mathbf{y}(t) = \left. \frac{\partial \mathbf{h}(\mathbf{x}, \mathbf{k}, t)}{\partial \mathbf{x}} \right|_{\mathbf{x}_r(t), \mathbf{k}_r(t)} \delta \mathbf{x}(t) \quad (3.10)$$

Observability of the linearized system may be determined directly by application of Eqn. (3.7) and checking the rank of the local linearized observability gramian. By Theorem 3.2 in [83], if $\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t))) > 0$ for all $(\mathbf{x}_r(t), \mathbf{k}_r(t)) \in \mathbf{W}$ then since no two trajectories in \mathbf{W} can yield an identical observation, the system must be observable at t_0 over the entire domain \mathbf{W} . Satisfying Theorem 3.2 in [83] is equivalent to showing that, for every point $\mathbf{x}_u \in \mathcal{R}$, the mapping between the output and input is indeed one-to-one. For autonomous systems, this observability mapping is defined as

$$\mathcal{O}(\mathbf{x}) = \begin{bmatrix} \mathbf{y} \\ \dot{\mathbf{y}} \\ \ddot{\mathbf{y}} \\ \vdots \end{bmatrix} \quad (3.11)$$

where the order of the derivatives of the output depend on the system [82]. The rank of the Jacobian of this mapping, $d\mathcal{O}(\mathbf{x}) = \partial \mathcal{O}(\mathbf{x}) / \partial \mathbf{x}$, is of particular interest. A rank-deficient Jacobian of this mapping implies there exists an equivalent unobservable space for nonlinear systems which is the null space of this Jacobian. Thus, if Theorem 3.2 in [83] is not satisfied then an unobservable space $\mathcal{N}(d\mathcal{O}(\mathbf{x}))$ exists and a state estimate admits

a continuum of solutions generating the same measurements over the time interval $t \in [t_0, t_f]$.

For observation of general dynamical systems, it is clear that if a continuum of solutions is yielded from a measurement, the continuum of solutions form nullspace of either \mathbf{P} or $d\mathcal{O}(\mathbf{x})$. The admissible region is a bounded subset of this nullspace, which implies that \mathcal{R} can only be formed if the system is unobservable. Alternatively, this implies that if a system can be shown to be observable, then \mathcal{R} must be an empty set. Statistically this implies that if $\mathcal{R} = \emptyset$ then a state estimate and corresponding distribution around that state estimate exists and the admissible region approach is not necessary. Applying this theory to observation of space objects leads to the following Lemma.

Lemma 1 (Admissible Regions and System Observability). *Observation of an object following Newtonian dynamics over a time period $t \in [t_0, t_f]$ such that $\Delta t = t_f - t_0 \ll \lambda_{\max}^{-1}$ yields a local linearized observability gramian with rank $\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t))) = d < n \forall (\mathbf{x}_r(t), \mathbf{k}_r(t)) \in \mathbf{W}$. Every point $\mathbf{x}_u \in \mathcal{R}$ is therefore unobservable for this short observation sequence.*

Proof. Utilizing the linearized system described in Eqns. (3.9) and (3.10), the observability gramian is given by Eqn. (3.7). Consider the Taylor series approximation of the state transition matrix.

$$\Phi(\tau, t_0) = \mathbb{I}_6 + \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\tau - t_0) + \text{H.O.T} \quad (3.12)$$

where $\mathbf{f}(\mathbf{x}, \mathbf{p}, t)$ represents the dynamics of a Newtonian system with the state given by

$$\mathbf{x} = \begin{bmatrix} \mathbf{r} \\ \dot{\mathbf{r}} \end{bmatrix} \quad (3.13)$$

where \mathbf{r} and $\dot{\mathbf{r}}$ are the position and velocity respectively and the system parameters given

by \mathbf{p} . The dynamics may be written in general as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{p}, t) = \begin{bmatrix} \dot{\mathbf{r}} \\ \mathbf{a}(\mathbf{x}, \mathbf{p}, t) \end{bmatrix} \quad (3.14)$$

The linearization of the dynamics is then given as follows

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0} & \mathbb{I}_3 \\ \frac{\partial \mathbf{a}(\mathbf{x}, \mathbf{p}, t)}{\partial \mathbf{r}} & \frac{\partial \mathbf{a}(\mathbf{x}, \mathbf{p}, t)}{\partial \dot{\mathbf{r}}} \end{bmatrix} \quad (3.15)$$

where \mathbb{I}_v is the $v \times v$ identity matrix. Let $\mathbf{F}_r = \partial \mathbf{a}(\mathbf{x}, \mathbf{p}, t) / \partial \mathbf{r}$ and $\mathbf{F}_{\dot{r}} = \partial \mathbf{a}(\mathbf{x}, \mathbf{p}, t) / \partial \dot{\mathbf{r}}$. Taking the first order terms of the Taylor series approximation of the state transition matrix,

$$\Phi(\tau, t_0) \approx \mathbb{I}_6 + \begin{bmatrix} \mathbf{0} & (\tau - t_0)\mathbb{I}_3 \\ (\tau - t_0)\mathbf{F}_r & (\tau - t_0)\mathbf{F}_{\dot{r}} \end{bmatrix} \quad (3.16)$$

where the acceleration of the linearized dynamics depends on the time span and the magnitudes of the elements of \mathbf{F}_r and $\mathbf{F}_{\dot{r}}$. Given that \mathbf{F}_r and $\mathbf{F}_{\dot{r}}$ are square matrices, the eigenvalue decomposition provides

$$\mathbf{F}_r = \mathbf{U}_1 \Lambda_r \mathbf{U}_1^{-1} \quad (3.17)$$

$$\mathbf{F}_{\dot{r}} = \mathbf{U}_2 \Lambda_{\dot{r}} \mathbf{U}_2^{-1} \quad (3.18)$$

where Λ_r and $\Lambda_{\dot{r}}$ are diagonal matrices containing the eigenvalues of \mathbf{F}_r and $\mathbf{F}_{\dot{r}}$. Let λ_{\max} denote the maximal eigenvalue in Λ_r and $\Lambda_{\dot{r}}$.

In general, there exists a constant C such that if C is small enough then $C\mathbf{F}_r$ and $C\mathbf{F}_{\dot{r}}$ are negligible. In particular, let the time span of interest satisfy $C = (\tau - t_0) \ll 1/\lambda_{\max}$ then it is clear that $(\tau - t_0)\lambda_{\max} \ll 1$ which further implies that $(\tau - t_0)\mathbf{F}_r$ and $(\tau - t_0)\mathbf{F}_{\dot{r}}$ are negligible. Thus, if the time period is small enough to satisfy the condition $(\tau - t_0) \ll \lambda_{\max}^{-1}$

then the contributions of \mathbf{F}_r and $\mathbf{F}_\dot{r}$ are negligible and Eqn. (3.16) is approximated as

$$\Phi(\tau, t_0) \approx \mathbb{I}_6 + \begin{bmatrix} \mathbf{0} & (\tau - t_0)\mathbb{I}_3 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.19)$$

and the dynamics essentially follow straight line motion.

Define

$$\mathbf{H}_r(\tau) = \frac{\partial \mathbf{h}(\mathbf{x}; \mathbf{k}, \tau)}{\partial \mathbf{r}(\tau)} \quad (3.20)$$

$$\mathbf{H}_{\dot{r}}(\tau) = \frac{\partial \mathbf{h}(\mathbf{x}; \mathbf{k}, \tau)}{\partial \dot{\mathbf{r}}(\tau)} \quad (3.21)$$

$$\mathbf{H}^T \mathbf{H} = \frac{\partial \mathbf{h}(\mathbf{x}(\tau); \mathbf{k}, \tau)^T}{\partial \mathbf{x}(\tau)} \frac{\partial \mathbf{h}(\mathbf{x}(\tau); \mathbf{k}, \tau)}{\partial \mathbf{x}(\tau)} \quad (3.22)$$

$$= \begin{bmatrix} \mathbf{H}_r^T \mathbf{H}_r & \mathbf{H}_r^T \mathbf{H}_{\dot{r}} \\ \mathbf{H}_{\dot{r}}^T \mathbf{H}_r & \mathbf{H}_{\dot{r}}^T \mathbf{H}_{\dot{r}} \end{bmatrix} \quad (3.23)$$

By definition, the rank of $\mathbf{H}^T \mathbf{H}$ depends on the dimension of $\partial \mathbf{h} / \partial \mathbf{x}$. Introduce a bijective transformation $\zeta : \mathbb{R}^n \rightarrow \mathbb{R}^n$ which maps the state vector into a partitioned state vector containing the observable and unobservable states as follows

$$\mathbf{z} = \zeta(\mathbf{x}) = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \quad (3.24)$$

where $\mathbf{z}_1 \in \mathbb{R}^q$ are the observable states of the system and $\mathbf{z}_2 \in \mathbb{R}^{n-q}$ are the unobservable states of the system. Each of the partial derivatives may now be partitioned as well yielding

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \frac{\partial \mathbf{h}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{h}}{\partial \mathbf{z}_1} & \frac{\partial \mathbf{h}}{\partial \mathbf{z}_2} \end{bmatrix} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \quad (3.25)$$

where $\text{rank}[\frac{\partial \mathbf{z}}{\partial \mathbf{x}}] = n$ since ζ is a bijective transformation. The unobservable states play no

role in the measurements, $\partial \mathbf{h} / \partial \mathbf{z}_2 = \mathbf{0}$ leaving

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{h}}{\partial \mathbf{x}_d} & \mathbf{0} \end{bmatrix} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \quad (3.26)$$

which implies that $\text{rank}[\mathbf{H}^T \mathbf{H}] = q$, the dimension of the observable states. With the approximation for $\Phi(\tau, t_0)$ and the definition of $\mathbf{H}^T \mathbf{H}$, the local linearized observability gramian may be analytically integrated. Keeping the matrices expressed in block form, using Eqn. (3.19), and introducing a change of variables $s = \tau - t_0$ then

$$\mathbf{H}_{(1,1)} = \mathbf{H}_r^T \mathbf{H}_r \quad (3.27)$$

$$\mathbf{H}_{(1,2)} = \mathbf{H}_r^T \mathbf{H}_r s \mathbb{I}_3 + \mathbf{H}_r^T \mathbf{H}_{\dot{r}} \quad (3.28)$$

$$\mathbf{H}_{(2,1)} = s \mathbb{I}_3 \mathbf{H}_r^T \mathbf{H}_r + (\mathbf{H}_r^T \mathbf{H}_{\dot{r}})^T \quad (3.29)$$

$$\mathbf{H}_{(2,2)} = s^2 \mathbb{I}_3 \mathbf{H}_r^T \mathbf{H}_r + s \mathbb{I}_3 \mathbf{H}_r^T \mathbf{H}_{\dot{r}} \quad (3.30)$$

$$+ (\mathbf{H}_r^T \mathbf{H}_{\dot{r}})^T s \mathbb{I}_3 + \mathbf{H}_{\dot{r}}^T \mathbf{H}_{\dot{r}} \quad (3.31)$$

$$\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t))) = \int_0^{\Delta t} \frac{\partial \mathbf{z}^T}{\partial \mathbf{x}} \begin{bmatrix} \mathbf{H}_{(1,1)} & \mathbf{H}_{(1,2)} \\ \mathbf{H}_{(2,1)} & \mathbf{H}_{(2,2)} \end{bmatrix} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} ds \quad (3.32)$$

The three specific cases for evaluating $\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t)))$ that are possible are given as follows

$$\begin{cases} \text{Case 1 :} & \mathbf{h}(\mathbf{x}; \mathbf{k}, t_1) \\ \text{Case 2 :} & \mathbf{h}(\mathbf{x}; \mathbf{k}, \{t_1, \dots, t_v\}), \quad (t_v - t_1) \ll \lambda_{\max}^{-1} \\ \text{Case 3 :} & \mathbf{h}(\mathbf{x}; \mathbf{k}, \{t_1, \dots, t_v\}), \quad (t_v - t_1) \gg \lambda_{\max}^{-1} \end{cases} \quad (3.33)$$

where $v \in \mathbb{Z}^+, v > 1$. Case 1 details a measurement that is dependent upon a single instance in time t_1 . The integration of Eqn. (3.32) is then only dependent upon instantaneous evaluations of \mathbf{h} at a given time. Case 2 details a measurement function dependent upon a time interval t_1 to t_v where the total time interval satisfies $(t_v - t_1) \ll \lambda_{\max}^{-1}$. In this

case the measurement is essentially a convolution over time which must be evaluated when determining the rank of the observability gramian. This is particularly the case for optical measurements where \mathbf{y} is often obtained from a streak which is obtained over a short time interval. Case 3 details a measurement function dependent upon a sufficiently long time period where the assumption of Eqn. (3.19) is no longer valid.

For both Case 1 and Case 2, since $\Delta t \ll \lambda_{\max}^{-1}$ is small, it is reasonable to assume that any higher order powers of Δt in $\Phi(t_v, t_0)$ may be considered negligible. With this in mind, the integration of Eqn. (3.32) gives a simple result after neglecting higher order terms of Δt .

$$\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t))) \approx \Delta t \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^T & \mathbf{H}_3 \end{bmatrix} \quad (3.34)$$

The matrix $\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t)))$ has $\text{rank}[\mathbf{H}^T \mathbf{H}] = d$ where $d = q$ for Case 1. For Case 2, the value of d depends on the time convolution of the measurement function over the time of observation. For both Case 1 and Case 2 then, $\text{rank}[\mathbf{P}(t_f, t_0, \mathbf{x}(t))] = d$ as long as $\Delta t \ll \lambda_{\max}^{-1}$. Because this is true for any point \mathbf{x}_u , if $d < n$, all such points are unobservable. For an observation falling under Case 3, it is possible that the system is observable and $\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t)))$ may have full rank. \square

Example 1 (Observability in Keplerian Dynamics).

To demonstrate Lemma 1 for Keplerian dynamics, it is first necessary to determine C . Specifically, the dynamics are given by

$$\mathbf{f}(\mathbf{x}, t) = \begin{bmatrix} \dot{\mathbf{r}} \\ -\frac{\mu}{\|\mathbf{r}\|^3} \mathbf{r} \end{bmatrix} \quad (3.35)$$

where $\mathbf{x} = \begin{bmatrix} \mathbf{r}^T & \dot{\mathbf{r}}^T \end{bmatrix}^T$. Thus,

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0} & \mathbb{I}_3 \\ \mathbf{M} & \mathbf{0} \end{bmatrix} \quad (3.36)$$

$$\Phi(\tau, t_0) \approx \mathbb{I}_6 + \begin{bmatrix} \mathbf{0} & (\tau - t_0)\mathbb{I}_3 \\ (\tau - t_0)\mathbf{M} & \mathbf{0} \end{bmatrix} \quad (3.37)$$

The matrix, \mathbf{M} , is given by

$$\mathbf{M} = \begin{bmatrix} \frac{3\mu r_x^2}{\|\mathbf{r}\|^5} - \frac{\mu}{\|\mathbf{r}\|^3} & \frac{3\mu r_x r_y}{\|\mathbf{r}\|^5} & \frac{3\mu r_x r_z}{\|\mathbf{r}\|^5} \\ \frac{3\mu r_y r_x}{\|\mathbf{r}\|^5} & \frac{3\mu r_y^2}{\|\mathbf{r}\|^5} - \frac{\mu}{\|\mathbf{r}\|^3} & \frac{3\mu r_y r_z}{\|\mathbf{r}\|^5} \\ \frac{3\mu r_z r_x}{\|\mathbf{r}\|^5} & \frac{3\mu r_z r_y}{\|\mathbf{r}\|^5} & \frac{3\mu r_z^2}{\|\mathbf{r}\|^5} - \frac{\mu}{\|\mathbf{r}\|^3} \end{bmatrix} \quad (3.38)$$

which can be written as

$$\mathbf{M} = \frac{3\mu}{\|\mathbf{r}\|^3} \begin{bmatrix} \frac{r_x^2}{\|\mathbf{r}\|^2} - \frac{1}{3} & \frac{r_x r_y}{\|\mathbf{r}\|^2} & \frac{r_x r_z}{\|\mathbf{r}\|^2} \\ \frac{r_y r_x}{\|\mathbf{r}\|^2} & \frac{r_y^2}{\|\mathbf{r}\|^2} - \frac{1}{3} & \frac{r_y r_z}{\|\mathbf{r}\|^2} \\ \frac{r_z r_x}{\|\mathbf{r}\|^2} & \frac{r_z r_y}{\|\mathbf{r}\|^2} & \frac{r_z^2}{\|\mathbf{r}\|^2} - \frac{1}{3} \end{bmatrix} \quad (3.39)$$

It can be shown that the eigenvalues of the factored matrix in Eqn. (3.39) are simply

$$\Lambda = \frac{3\mu}{\|\mathbf{r}\|^3} \begin{bmatrix} -1/3 & 0 & 0 \\ 0 & -1/3 & 0 \\ 0 & 0 & 2/3 \end{bmatrix}$$

It follows that $\lambda_{max} = 2\mu/\|\mathbf{r}\|^3$. It is then possible to define a time interval sufficiently small enough that $\mathbf{M}(\tau - t_0)$ can be approximated to have a negligible contribution by letting $(\tau - t_0) \ll \sqrt{\|\mathbf{r}\|^3/2\mu}$ where the square root is necessary for consistent units of time.

The rest of Lemma 1 follows directly by continuing the proof with $(\tau - t_0) \ll \sqrt{\|\mathbf{r}\|^3/2\mu}$.

Lemma 1 gives the conditions under which an admissible region exists for a dynamical system. For optical observations, the instantaneous measurement consists of two angles giving $d = q = 2$, a Case 1 situation. However, since a truly instantaneous measurement is often not realizable, optical observations have a finite integration time. The convolution of these instantaneous measurements over the integration time provides angle rate information for the measurement and thus while $q = 2$, this Case 2 situation yields $d = 2q$. It is important to understand the type of measurement to properly determine the observability of the system. The next section builds upon the existence of the admissible region and shows its definition and properties.

3.2.2 Defining the Admissible Region

Defining the admissible region requires knowledge of a measurement model for the system being observed. Consider a general nonlinear measurement model given by

$$\mathbf{y} = \mathbf{h}(\mathbf{x}; \mathbf{k}, t) \quad (3.40)$$

As done in all admissible region approaches, the state vector is partitioned in to determined states $\mathbf{x}_d \in \mathbb{R}^d$ and undetermined states $\mathbf{x}_u \in \mathbb{R}^u$ where $u + d = n$ [76]. This means that

$$\mathbf{y} = \mathbf{h}(\mathbf{x}_d; \mathbf{k}, t) \quad (3.41)$$

Admissible region approaches constrain this continuum of solutions using hypothesized constraints in the form $\kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) \leq 0$ where $\kappa_i : \mathbb{R}^u \times \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R} \rightarrow \mathbb{R}$. The admissible region for the i^{th} hypothesized constraint $\kappa_i(\cdot)$ is then defined as

$$\mathcal{R}_i := \{\mathbf{x}_u \in \mathbb{R}^u \mid \kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) \leq 0\} \quad (3.42)$$

where $\mathcal{R}_i \subseteq \mathbb{R}^u$. Furthermore, if there are c such hypotheses then the total combined admissible region is given by

$$\mathcal{R} = \bigcap_{i=1}^c \mathcal{R}_i \quad (3.43)$$

where \mathcal{R} must be a bounded set [64]. The requirement that \mathcal{R} be compact ensures the assumed uniform distribution has non-zero probability. Thus, each state $\mathbf{x} \in \mathcal{R}$ can be assigned a non-zero uniform probability.

3.2.3 The Admissible Region Prior

The probability that a given state $\mathbf{x}_u \in \mathbb{R}^u$ satisfies the i^{th} admissible region constraint is then given by

$$\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] = \mathbb{P}[\kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) \leq 0] \quad (3.44)$$

Without any additional information, the inequality defining \mathcal{R}_i in Equation (3.42) is a binary constraint and $\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] \in \{0, 1\}$ since each \mathbf{x}_u has either 100% or 0% probability of satisfying the constraint. Thus the probability that \mathbf{x}_u satisfies a given constraint κ_i can be exactly expressed as a piecewise membership function defined as

$$m_i(\mathbf{x}_u) = \begin{cases} 1, & \kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) \leq 0 \\ 0, & \kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) > 0 \end{cases} \quad (3.45)$$

Thus $\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] = m_i(\mathbf{x}_u)$, and the prior distribution for a particular constraint hypothesis can then be defined as [78]

$$f_{i, \mathbf{x}_u}(\mathbf{x}_u) = \frac{m_i(\mathbf{x}_u)}{\int_{\mathcal{R}_i} d\mathbf{x}_u} \quad (3.46)$$

Eqn. (3.46) results in a uniform distribution, which is demonstrated in [47]. Applying the chain rule of probabilities, the general joint probability function over all k constraints can

be written as

$$\begin{aligned}
 f_{\mathbf{x}_u}(\mathbf{x}_u) &= \frac{\mathbb{P}[\mathbf{x}_u \in \mathcal{R}]}{\int_{\mathcal{R}} d\mathbf{x}_u} \\
 &= \frac{1}{\int_{\mathcal{R}} d\mathbf{x}_u} \prod_{k=1}^c \mathbb{P} \left[\mathbf{x}_u \in \mathcal{R}_k \mid \mathbf{x}_u \in \bigcap_{j=1}^{k-1} \mathcal{R}_j \right]
 \end{aligned} \tag{3.47}$$

where the bracketed term gives the probability that k^{th} constraint is satisfied given that each of the $k - 1$ previous constraints are satisfied [79]. If the constraints κ_i are assumed to be independent, then by Bayes' rules the conditional probability terms evaluate to 1 and Eqn. (3.47) simplifies to

$$f_{\mathbf{x}_u}(\mathbf{x}_u) = \frac{\prod_{k=1}^c \mathbb{P}[\mathbf{x}_u \in \mathcal{R}_k]}{\int_{\mathcal{R}} d\mathbf{x}_u} \tag{3.48}$$

$$= \frac{\prod_{k=1}^c m_k(\mathbf{x}_u)}{\int_{\mathcal{R}} d\mathbf{x}_u} \tag{3.49}$$

By this formulation, every $\mathbf{x}_u \in \mathcal{R}$ is a candidate solution that satisfies the c constraints and without additional information; no one state can be considered more likely than another. Thus $f_{\mathbf{x}_u}(\mathbf{x}_u)$ is a constant over \mathcal{R} and as such the admissible region must be considered a uniform distribution. This fact is consistent with the work presented by Fujimoto and Scheeres stating that without any a priori information regarding the observation, an admissible region is expressed as a uniform PDF [72]. However, it should be noted that the notation used in this paper will refer to the statistical representation of the admissible region as an admissible region prior, which is also consistent with a general uninformative prior. The reason for this notation will become clear in the next section.

3.2.4 Transformation of the Admissible Region Prior

Suppose a user wishes to use the admissible region method to initiate an estimation procedure in a state space different from the state space in which the admissible region constraints are formed. Following the general probability transformation approach, a transformation

$\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ can be defined. This transformation must also be able to be partitioned into $\mathbf{g}_u : \mathbb{R}^u \rightarrow \mathbb{R}^u$ and $\mathbf{g}_d : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that

$$\tilde{\mathbf{x}}_u = \mathbf{g}_u(\mathbf{x}_u; \mathbf{y}, \mathbf{k}, t) \quad (3.50)$$

$$\tilde{\mathbf{x}}_d = \mathbf{g}_d(\mathbf{x}_d; \mathbf{y}, \mathbf{k}, t) \quad (3.51)$$

For simplicity, this transformation will be expressed as $\mathbf{g}_u(\mathbf{x}_u; \cdot)$ for the remainder of this paper. In general, the transformation $\mathbf{g}_u(\mathbf{x}_u; \cdot)$ must be left invertible in order to preserve probability. Additionally, the transformation must satisfy the condition that the underdetermined and determined states in the transformed space are still capable of being partitioned, leading to the following Lemma.

Lemma 2 (Partitioned Transformed State). *An admissible region prior expressed in state space \mathbf{x}_u may be transformed to state space $\tilde{\mathbf{x}}_u = \mathbf{g}_u(\mathbf{x}_u; \cdot)$ only if there exist some $\tilde{\mathbf{x}}_d = \bar{\mathbf{g}}_d(\mathbf{y}; \cdot)$, $\bar{\mathbf{g}}_d : \mathbb{R}^m \rightarrow \mathbb{R}^d$ such that $\mathbf{y} = \mathbf{h}(\mathbf{x}_d; \mathbf{k}, t) = \tilde{\mathbf{h}}(\tilde{\mathbf{x}}_d; \mathbf{k}, t)$, $\tilde{\mathbf{h}} : \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R} \rightarrow \mathbb{R}^m$.*

Proof. The undetermined states \mathbf{x}_u are independent of the determined states \mathbf{x}_d as defined in [76]. This enables the partitioning of the state space such that the measurement \mathbf{y} is only a function of the determined states, the parameters \mathbf{k} , and time and can be expressed by

$$\mathbf{y} = \mathbf{h}(\mathbf{x}_u, \mathbf{x}_d; \mathbf{k}, t) = \mathbf{h}(\mathbf{x}_d; \mathbf{k}, t)$$

which by definition means $\mathbf{x}_d = \mathbf{h}^{-1}(\mathbf{y}, \mathbf{k}, t)$. If there is a transformation of \mathbf{x}_u , then the transformation can be given by

$$\begin{aligned} \tilde{\mathbf{x}}_d &= \mathbf{g}_d(\mathbf{x}_d) \\ &= \mathbf{g}_d(\mathbf{h}^{-1}(\mathbf{y}, \mathbf{k}, t)) \end{aligned}$$

which can be defined as $\bar{\mathbf{g}}_d = \mathbf{g}_d \circ \mathbf{h}^{-1} : \mathbb{R}^m \rightarrow \mathbb{R}^d$ giving,

$$\tilde{\mathbf{x}}_d = \bar{\mathbf{g}}_d(\mathbf{y}; \cdot)$$

Thus, the measurement function is now expressed by

$$\mathbf{y} = \tilde{\mathbf{h}}(\mathbf{g}_u(\mathbf{x}_u; \cdot), \bar{\mathbf{g}}_d(\mathbf{y}; \cdot), \mathbf{k}, t)$$

For the admissible region problem, it is required that $\tilde{\mathbf{x}}$ can be partitioned into $\tilde{\mathbf{x}}_u$ and $\tilde{\mathbf{x}}_d$ such that \mathbf{y} is independent of $\tilde{\mathbf{x}}_u$. In general $\tilde{\mathbf{h}}(\mathbf{g}_u(\mathbf{x}_u; \cdot), \bar{\mathbf{g}}_d(\mathbf{y}; \cdot), \mathbf{k}, t) \neq \tilde{\mathbf{h}}(\bar{\mathbf{g}}_d(\mathbf{y}; \cdot), \mathbf{k}, t)$ since the transformation is not necessarily a function solely of \mathbf{x}_u . Thus, the function $\bar{\mathbf{g}}_d(\mathbf{y}; \cdot)$ must be defined to ensure that the determined variables are transformed such that the transformed undetermined states remain independent of the measurements. If a transformation $\bar{\mathbf{g}}_u(\mathbf{y}; \cdot)$ cannot be defined such that this is true then

$$\mathbf{y} = \tilde{\mathbf{h}}(\mathbf{g}_u(\mathbf{x}_u; \cdot), \bar{\mathbf{g}}_d(\mathbf{y}; \cdot), \mathbf{k}, t) \neq \tilde{\mathbf{h}}(\bar{\mathbf{g}}_d(\mathbf{y}; \cdot), \mathbf{k}, t)$$

and the admissible region formulation is invalid. □

The result of Lemma 2 essentially requires that if the undetermined states can be transformed then they must remain unobservable with respect to the observations. Because this is a requirement for the formation of an admissible region, any transformation that does not satisfy Lemma 2 necessarily generates a region that can no longer be defined as an admissible region.

Assuming a transformation satisfying Lemma 2 exists, the admissible region in the transformed space can be defined. For the admissible region problem, since the constraint

hypothesis is a function of a unique state \mathbf{x}_u ,

$$\kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) = \tilde{\kappa}_i(\mathbf{g}_u(\mathbf{x}_u; \cdot), \mathbf{y}, \mathbf{k}, t) \quad (3.52)$$

$$\tilde{\kappa}_i(\tilde{\mathbf{x}}_u, \mathbf{y}, \mathbf{k}, t) = \kappa_i(\mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u; \cdot), \mathbf{y}, \mathbf{k}, t) \quad (3.53)$$

Eqns. (3.52) and (3.53) then imply that $\mathbb{P}[\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{R}}_i] = \mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i]$ and $m_i(\mathbf{x}_u) = m_i(\tilde{\mathbf{x}}_u)$ where,

$$\widetilde{\mathcal{R}}_i := \{\tilde{\mathbf{x}}_u \in \mathbb{R}^u \mid \tilde{\kappa}_i(\tilde{\mathbf{x}}_u, \mathbf{y}, \mathbf{k}, t) \leq 0\} \quad (3.54)$$

and

$$\tilde{m}_i(\tilde{\mathbf{x}}_u) = \begin{cases} 1, & \tilde{\kappa}_i(\tilde{\mathbf{x}}_u, \mathbf{y}, \mathbf{k}, t) \leq 0 \\ 0, & \tilde{\kappa}_i(\tilde{\mathbf{x}}_u, \mathbf{y}, \mathbf{k}, t) > 0 \end{cases} \quad (3.55)$$

The general admissible region prior in the transformed space is given by

$$f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u) = \frac{1}{\int_{\widetilde{\mathcal{R}}} d\tilde{\mathbf{x}}_u} \prod_{k=1}^c \mathbb{P} \left[\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{R}}_k \mid \tilde{\mathbf{x}}_u \in \bigcap_{j=1}^{k-1} \widetilde{\mathcal{R}}_j \right] \quad (3.56)$$

Assuming again that the constraint hypotheses are independent, the admissible region prior expressed in $\tilde{\mathbf{x}}_u$ is given by,

$$f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u) = \frac{\prod_{k=1}^c \tilde{m}_k(\tilde{\mathbf{x}}_u)}{\int_{\widetilde{\mathcal{R}}} d\tilde{\mathbf{x}}_u} \quad (3.57)$$

A general nonlinear transformation of a uniform PDF must yield a non-uniform PDF according to Eqn. (3.3). For an unobservable system, the PDF is defined by a marginalization over only the observable states [84]. The result of a nonlinear transformation of the n dimensional state space then has no probabilistic impact on this marginal PDF, and as such under this approach any initial distribution for the unobservable states may be used.

This is the justification for the use of probability transformations in the admissible region approach outlined in the introduction.

However, this application of probability theory implies that there is inherent benefit to be gained in selecting a different state space in which to express the problem which gives different initial distributions. While probabilistically correct, this approach can yield inconsistent results due to the implication that given a single, unobservable measurement the initial PDF can be better represented based solely on the state space in which the problem is expressed. This inconsistency is noted in alternative formulations of probability theory which aim to preserve objective information about the unobservable system, which does not follow from the straightforward application of probability theory [3].

Theorem 2 will prove by contradiction this inconsistency for the admissible region problem by showing how the choice of state space can impact the validity of the hypothesized constraint assumptions and derive a condition which preserves objective information about the unobservable system, namely the probability of set membership. The uniform PDF of an admissible region is a statistical representation of the fact that each state $\mathbf{x}_u \in \mathcal{R}$ is consistent with the measurement \mathbf{y} . Without any additional information, each state necessarily has equal probability which must also be true if \mathbf{x}_u is expressed in any other state space. Given this fact, there must exist a necessary relationship between $f_{\mathbf{x}_u}(\mathbf{x}_u)$ and $f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u)$ to preserve the set membership information.

Theorem 2 (Equivalence of Admissible Regions). *Given $\mathbf{x}_u \in \mathcal{R}$ and an invertible transformation $\tilde{\mathbf{x}}_u = \mathbf{g}_u(\mathbf{x}_u; \cdot)$, a reparameterization of the admissible region prior by \mathbf{g} is only valid if the transformation satisfies $|\partial\mathbf{x}_u/\partial\tilde{\mathbf{x}}_u| = \zeta \forall \mathbf{x}_u \in \mathcal{R}$ where ζ is the ratio of the volume of the admissible region as expressed in both state spaces and $f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u) = \zeta f_{\mathbf{x}_u}(\mathbf{x}_u)$.*

Proof. The proof of Theorem 2 is given by way of contradiction. Assume first that the statistical representation of the admissible region is given by a PDF. Then assume there exists an invertible transformation $\mathbf{g}_u(\mathbf{x}_u; \cdot)$ for which $|\partial\mathbf{x}_u/\partial\tilde{\mathbf{x}}_u| \neq \zeta$ for some $\mathbf{x} \in \mathcal{R}$. The relationship between $f_{\mathbf{x}_u}(\mathbf{x}_u)$ and $f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u)$ may be determined by applying Eqn. (3.3)

as follows

$$f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u) = \frac{\prod_{k=1}^c m_k(\mathbf{g}^{-1}(\tilde{\mathbf{x}}_u))}{\int_{\mathcal{R}} d\mathbf{x}_u} \text{abs} \left(\left| \frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u} \right| \right) \quad (3.58)$$

Each of the terms in Eqn. (3.58) have been defined thus far except for the Jacobian term $|\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}}_u)/\partial \tilde{\mathbf{x}}_u|$. Rearranging Eqn. (3.58), by substituting the $\tilde{\mathbf{x}}_u$ PDF on the left hand side and multiplying by the denominator of the right hand side,

$$\frac{\prod_{k=1}^c \tilde{m}_k(\tilde{\mathbf{x}}_u)}{\int_{\tilde{\mathcal{R}}} d\tilde{\mathbf{x}}_u} \int_{\mathcal{R}} d\mathbf{x}_u = \prod_{k=1}^c m_k(\mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)) \text{abs} \left(\left| \frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u} \right| \right) \quad (3.59)$$

Note that for the admissible region approach $m_k(\mathbf{x}_u) = \tilde{m}_k(\tilde{\mathbf{x}}_u)$ since it is necessary that $\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] = \mathbb{P}[\tilde{\mathbf{x}}_u \in \tilde{\mathcal{R}}_i]$. Thus, dividing each side by $\prod_{k=1}^c \tilde{m}_k(\tilde{\mathbf{x}}_u)$ results in,

$$\frac{\int_{\mathcal{R}} d\mathbf{x}_u}{\int_{\tilde{\mathcal{R}}} d\tilde{\mathbf{x}}_u} = \zeta = \text{abs} \left(\left| \frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u} \right| \right) \quad (3.60)$$

If $|\partial \mathbf{x}_u/\partial \tilde{\mathbf{x}}_u| \neq \zeta$ then,

$$\frac{\int_{\mathcal{R}} d\mathbf{x}_u}{\int_{\tilde{\mathcal{R}}} d\tilde{\mathbf{x}}_u} \neq \zeta \quad (3.61)$$

which then implies $\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] \neq \mathbb{P}[\tilde{\mathbf{x}}_u \in \tilde{\mathcal{R}}_i]$ for Eqn. (3.59) to hold. But this is a contradiction since, by definition, the admissible region approach gives that $\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] = \mathbb{P}[\tilde{\mathbf{x}}_u \in \tilde{\mathcal{R}}_i]$ regardless of the transformation. \square

Theorem 2 imposes a geometric constraint on the transformation \mathbf{g} through the determinant of the Jacobian. The constraint requires the determinant to be constant which implies the distortion of the $\tilde{\mathbf{x}}_u$ state space relative to the \mathbf{x}_u state space is the same at every point. This is necessary to ensure that any one point inside the admissible region in \mathbf{x}_u remains inside the equivalent admissible region expressed in $\tilde{\mathbf{x}}_u$. The constant Jacobian constraint limits the practical applicability of probability transformations to admissible region because

useful state space transformation are often complex, nonlinear functions. More importantly, Theorem 2 shows an admissible region must be uniform, or uninformative, regardless of the state space it is expressed in. Any general PDF must satisfy the probability transformation given by Eqn. (3.3), but as shown by Theorem 2, under most practical transformations, an admissible region fails to satisfy Eqn. (3.3). Moreover, the implications of Theorem 2 follow directly the main ideal behind the Principle of Transformation Groups through the preservation of the prior probability across a transformation.

The result of Theorem 2 is directly related to Jeffreys' prior [3]. A Jeffreys' prior is an uninformative prior which satisfies

$$f(\mathbf{x}) \propto \sqrt{\det \mathcal{I}(\mathbf{x})} \quad (3.62)$$

where $f(\cdot)$ denotes the prior and $\mathcal{I}(\mathbf{x})$ is the Fisher information matrix [85]. If $\mathcal{I}(\mathbf{x})$ is singular then Jeffreys' prior does not exist [86]. For the application of Jeffreys' prior in this paper, since \mathbf{x} essentially belongs to a uniform distribution, the Fisher information matrix will be non-singular. The proportionality of Eqn. (3.62) gives that a Jeffreys prior is invariant to a reparameterization of \mathbf{x} . Applying the previously derived probability transformation and defining a reparameterization or transformation of \mathbf{x} given by $\tilde{\mathbf{x}} = \mathbf{g}(\mathbf{x}; \cdot)$ then

$$f(\tilde{\mathbf{x}}) = f(\mathbf{x}) \left| \frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}} \right| \quad (3.63)$$

It can likewise be shown that

$$\sqrt{\det \mathcal{I}(\tilde{\mathbf{x}})} = \sqrt{\det \mathcal{I}(\mathbf{x})} \left| \frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}} \right| \quad (3.64)$$

thus for Jeffrey's prior to hold, Eqn. (3.62) may be rewritten as

$$f(\tilde{\mathbf{x}}) \left| \frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}} \right|^{-1} \propto \sqrt{\det \mathcal{I}(\tilde{\mathbf{x}})} \left| \frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}} \right|^{-1} \quad (3.65)$$

The proportionality of Eqn. (3.62) requires that $|\partial \mathbf{x} / \partial \tilde{\mathbf{x}}|^{-1} = |\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u) / \partial \tilde{\mathbf{x}}_u|$ to be a constant. This result is directly equivalent to Theorem 2 for admissible regions. Jeffreys' prior is based on Jeffreys' Rule which states that given an equation for $f(\mathbf{x})$, applying the equation to determine $f(\tilde{\mathbf{x}})$ directly should yield an identical result as computing $f(\mathbf{x})|\partial \mathbf{x} / \partial \tilde{\mathbf{x}}|$. Applying this to an admissible region system, Jeffreys' Rule states that if $f_{\mathbf{x}_u}(\mathbf{x}_u)$ is the prior, then a reparameterization of \mathbf{x} must satisfy

$$f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u) \left| \frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u} \right| \propto \sqrt{\det \mathcal{I}(\tilde{\mathbf{x}}_u)} \left| \frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u} \right| \quad (3.66)$$

which confirms Theorem 2 and shows it is consistent with the statistical representation of an admissible region being identified in this paper as an admissible region prior as opposed to a PDF.

Given that a transformation \mathbf{g} exists which satisfies Theorem 2, it is possible to define the transformed admissible region prior. The final expression for the transformed admissible region prior is then given by

$$f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u) = \zeta \frac{\prod_{k=1}^c m_k(\mathbf{x}_u)}{\int_{\mathcal{R}} d\mathbf{x}_u} \quad (3.67)$$

Eqn. (3.67) signifies that for the admissible region problem with no additional information, the admissible region of \mathbf{x}_u expressed in any transformed state space $\tilde{\mathbf{x}}_u$ such that $\mathbf{g}^{-1}(\tilde{\mathbf{x}}_u)$ exists is necessarily uniform and simply scaled by a factor ζ . Given that the transformation satisfies Theorem 2, the admissible region prior may be expressed in any state space which agrees with the work shown in [47]. It should be noted that useful transformations are often highly non-linear and as such will not typically satisfy the conditions presented

by Theorem 2. It is likely that, in general, an admissible region admissible region prior cannot be transformed since no practical transformations exists satisfying Theorem 2. If an admissible region prior is transformed by a transformation not satisfying Theorem 2, then the prior in the transformed space is no longer a uniform representation of the state space, and this non-uniform representation is not based on statistical information but based only on the transformation function. Because of this, any transformation not satisfying Theorem 2 generates a set of prior probabilities misrepresenting the true distribution by directly violating the Principle of Transformation Groups.

3.2.5 Priors Using Uncertain Admissible Region Constraints

While Eqn. (3.67) applies for transformations of uniform priors, it may also be applied to non-uniform admissible region priors. An approach for generating the non-uniform probability that \mathbf{x}_u is in \mathcal{R} is shown in [76]. The approximate analytical probability for a given admissible region is given by

$$\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] = m_i(\mathbf{x}_u) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\|\mathbf{x}_u - \mathbf{x}_{u,\mathcal{B}_\perp,i}\|}{\sqrt{2\operatorname{tr}\mathbf{P}_{\mathbf{x}_{u,\mathcal{B}_\perp,i}}}} \right) \right] \quad (3.68)$$

which updates the piecewise membership function given by Eqn. (3.45) to a continuous membership function by including uncertainty effects. These uncertainties are quantified as the covariance matrix \mathbf{P}_z where \mathbf{z} is the combined matrix of the measurements, parameters, and time. The quantity $\mathbf{x}_{u,\mathcal{B}_\perp,i}$ is the point on the boundary of \mathcal{R}_i orthogonal to \mathbf{x}_u and $\mathbf{P}_{\mathbf{x}_{u,\mathcal{B}_\perp,i}}$ is the covariance calculated at this boundary point. Substituting Eqn. (3.68) into Eqn. (3.49) then gives the non-uniform prior.

Corollary 2 (Systematic Uncertainty in Admissible Regions). *If the combined measurements, parameters, and time covariance matrix \mathbf{P}_z are known then transformation of the*

non-uniform admissible region probability is given by

$$\mathbb{P}[(\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{R}}_i)] = \tilde{m}_i(\tilde{\mathbf{x}}_u) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\|\mathbf{g}_u(\mathbf{x}_u; \cdot) - \mathbf{g}_u(\mathbf{x}_{u, \mathcal{B}_{\perp, i}}; \cdot)\|}{\sqrt{2\operatorname{tr}\mathbf{P}_{\tilde{\mathbf{x}}_u, \mathcal{B}_{\perp, i}}}} \right) \right] \quad (3.69)$$

where $\mathbf{P}_{\tilde{\mathbf{x}}_u, \mathcal{B}_{\perp, i}}$ is the modified covariance matrix.

Proof. Given the previous transformation $\tilde{\mathbf{x}}_u = \mathbf{g}_u(\mathbf{x}_u; \cdot)$, Eqn. (3.68) can be derived for $\tilde{\mathbf{x}}_u$. The simplified Taylor series expansion from Eqn. (17) in [76] now becomes

$$-\frac{\partial \kappa_i}{\partial \tilde{\mathbf{x}}_u} \frac{\partial \tilde{\mathbf{x}}_u}{\partial \mathbf{x}_u} \delta \mathbf{X}_u = \frac{\partial \kappa_i}{\partial \mathbf{z}} \delta \mathbf{Z} \quad (3.70)$$

Carrying the notation defined in [76], a new perpendicular vector $\tilde{\mathbf{p}}$ is defined as

$$\tilde{\mathbf{p}} = \left. \frac{\partial \kappa_i}{\partial \tilde{\mathbf{x}}_u} \frac{\partial \mathbf{g}_u^{-1}(\mathbf{x}_u; \cdot)}{\partial \mathbf{x}_u} \right|_{\mathbf{x}_u} \quad (3.71)$$

The rest of the derivation can be carried out as specified in [76] by replacing \mathbf{p} with $\tilde{\mathbf{p}}$ resulting in

$$\tilde{\mathbf{M}} = \begin{bmatrix} \tilde{\mathbf{p}}^T \\ \mathbf{T} \end{bmatrix}^{-1} \begin{bmatrix} -\frac{\partial \kappa_i}{\partial \mathbf{z}} \\ \mathbf{0} \end{bmatrix}$$

where $\mathbf{T} \in \mathbb{R}^{u-1 \times u}$ is a matrix of tangential unit vectors which gives

$$\mathbf{P}_{\tilde{\mathbf{x}}_u} = \tilde{\mathbf{M}} \mathbf{P}_z \tilde{\mathbf{M}}^T \quad (3.72)$$

Eqn. (3.69) is obtained by substituting $\mathbf{P}_{\tilde{\mathbf{x}}_u}$ and $\mathbf{g}(\mathbf{x}_u)$ into Eqn. (3.68). \square

Eqn. (3.69) defines the approximate analytical probability distribution function for an

admissible region \mathcal{R}_i in the $\tilde{\mathbf{x}}_u$ space. Alternatively, from Eqns. (3.52) and (3.53)

$$\mathbb{P}[\mathbf{x}_u \in \mathcal{R}_i] = \mathbb{P}[\tilde{\mathbf{x}}_u \in \tilde{\mathcal{R}}_i] \quad (3.73)$$

$$= \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\|\mathbf{g}_u(\mathbf{x}_u; \cdot) - \mathbf{g}_u(\mathbf{x}_u; \cdot)_{u, \mathcal{B}_\perp}\|}{\sqrt{2\operatorname{tr}\mathbf{P}_{\tilde{\mathbf{x}}_u, \mathcal{B}_\perp}}} \right) \right] \quad (3.74)$$

$$\approx \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\|\mathbf{x}_u - \mathbf{x}_{u, \mathcal{B}_\perp}\|}{\sqrt{2\operatorname{tr}\mathbf{P}_{\mathbf{x}_u, \mathcal{B}_\perp}}} \right) \right] \quad (3.75)$$

Because of this, it is equivalent to directly map each \mathbf{x}_u to $\tilde{\mathbf{x}}_u$ and assign each $\tilde{\mathbf{x}}_u = \mathbf{g}(\mathbf{x}_u)$ the probability of set membership $\mathbb{P}[\mathbf{x} \in \mathcal{R}_i]$ or vice-versa.

3.2.6 The Observability Condition

Lemma 1 shows that the existence of the admissible region implies that there is a non-trivial unobservable subspace of the system given a short enough observation. However, it is possible for the system to become fully observable given enough observations or a long enough observation of the system. Thus, it is of interest to understand how the observability of a system affects the transformation of the admissible region prior. If an initial observation is made such that the admissible region is non-empty then the admissible region prior is the statistical representation of the \mathcal{R} . However, if an additional measurement can be taken at a time t such that each state $\mathbf{x}_u \in \mathcal{R}$ is locally observable, then an a posteriori PDF can be constructed. This a posteriori PDF represents a true PDF over the state space and can be used directly with Eqn. (3.3) to transform probabilities between state spaces. As such, it is of interest to determine when the states $\mathbf{x}_u \in \mathcal{R}$ become locally observable.

Corollary 3 (Observability in Admissible Region Problems). *If the observability gramian for the admissible region system satisfies $\operatorname{rank}[\mathbf{P}(t_f, t_0, \mathbf{x}(t))] = n$ where $\mathbf{x}(t) = [\mathbf{x}_d(t) \ \mathbf{x}_u(t)] \forall \mathbf{x}_u \in \mathcal{R}$ then the PDF associated with the admissible region estimate may be transformed without the condition $|\partial \mathbf{x}_u / \partial \tilde{\mathbf{x}}_u| = \zeta \forall \mathbf{x}_u \in \mathcal{R}$.*

Proof. The admissible region \mathcal{R} is, as defined, a subset of the unobservable state space

where each state $\mathbf{x}_u \in \mathcal{R}$ has no effect on the measurements. Since the mapping \mathbf{h} from \mathbf{x} to \mathbf{y} cannot be a one-to-one and onto, each $\mathbf{x}_u \in \mathcal{R}$ must necessarily have a uniform probability. Because this is also true in any transformed state space $\tilde{\mathbf{x}}$, the admissible region must necessarily be uniform in any state space. If a system is locally observable at $\mathbf{x}_k \in \mathbb{R}^n$, where k is an arbitrary index, then there exists a measurement function $\mathbf{h}_o : \mathbb{R}^n \rightarrow \mathbb{R}^m$ where \mathbf{h}_o is a one-to-one and onto function. Thus, $\mathbf{x}_j \neq \mathbf{x}_k \implies \mathbf{h}_o(\mathbf{x}_j) \neq \mathbf{h}_o(\mathbf{x}_k)$ and each unique observation corresponds to a unique state \mathbf{x} . If the transformation $\mathbf{g}(\mathbf{x})$ is also one-to-one and onto then there must also exist a measurement function $\tilde{\mathbf{h}}_o : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that $\tilde{\mathbf{x}}_j \neq \tilde{\mathbf{x}}_k \implies \tilde{\mathbf{h}}_o(\tilde{\mathbf{x}}_j) \neq \tilde{\mathbf{h}}_o(\tilde{\mathbf{x}}_k)$ and $\mathbf{h}_o(\mathbf{x}_j) = \tilde{\mathbf{h}}_o(\tilde{\mathbf{x}}_j) = \mathbf{y}$. A unique solution exists for a given observation, or set of observations, and a PDF can then be defined about that solution. Because this unique PDF cannot be identical in both state spaces, the condition given by $|\partial \mathbf{x}_u / \partial \tilde{\mathbf{x}}_u| = \zeta$ can no longer hold, and for an observable system the PDF can simply be transformed by Eqn. (3.3). \square

The main result of Corollary 3 is that the PDF associated with a given \mathcal{R} may not generally be transformed until it is observable. This equivalently means that until the problem is observable, the prior defined over \mathcal{R} is uninformative. Since there are likely no practical transformations that satisfy Theorem 2, the significance of Corollary 3 is in the fact that general admissible region PDF transformations are possible, but only once each state in \mathcal{R} becomes locally observable. Furthermore, by Lemma 1, if every $\mathbf{x}_u \in \mathcal{R}$ is locally observable, then the region is necessarily not an admissible region.

3.3 Additional Transformations

This section discusses additional transformations that apply to the probability transformation theorems, corollaries, and lemmas presented in this work.

3.3.1 Linear Transformations

The only set of functions that will always satisfy Theorem 2 are linear transformations leading to Remark 1.

Remark 1: Any linear transformation $\tilde{\mathbf{x}}_u = \mathbf{g}_u(\mathbf{x}_u) = \mathbf{T}_u \mathbf{x}_u$ such that $\mathbf{T}_u \in \mathbb{R}^{n \times n}$, $\text{rank} \mathbf{T}_u = n$ that can be defined $\forall \mathbf{x}_u \in \mathcal{R}$ will satisfy the requirements given by Lemma 1 and Theorem 2. Thus, for any linear transformation of an admissible region, ζ can be defined such that $f_{\tilde{\mathbf{x}}_u}(\tilde{\mathbf{x}}_u) = \zeta f_{\mathbf{x}_u}(\mathbf{x}_u)$.

Any invertible linear transformation of covariance in extended Kalman filters satisfies Theorem 2 as long as the covariance is sufficiently small. While any linear transformation of the admissible region prior satisfies Theorem 2, it is unlikely that these transformations are practical or useful for the admissible region problem.

3.3.2 Sigma Point Transformations

An additional application of the general probability transformation comes from sigma point transformations and filters [87]. Sigma point filters use transformations of the sigma points of a Gaussian PDF to map the PDF over nonlinear transformations, used largely in the Unscented Kalman Filter. The sigma point transformation as originally defined relies on the fact that the transformation preserves the mean and covariance [88]. Alternatively, the sigma point transformation must preserve the PDF. Assume a PDF $f_{\mathbf{x}}(\mathbf{x})$ is known for a given \mathbf{x} , then the first order Taylor Series expansion of the inverse of the transformation $\tilde{\mathbf{x}} = \mathbf{g}(\mathbf{x})$ is given by

$$\mathbf{x} + \delta \mathbf{x} = \mathbf{g}^{-1}(\tilde{\mathbf{x}}) + \frac{\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}} \delta \tilde{\mathbf{x}} \quad (3.76)$$

$$\mathbf{x} + \delta \mathbf{x} = \mathbf{x} + \frac{\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}} \delta \tilde{\mathbf{x}} \quad (3.77)$$

$$\delta \mathbf{x} = \frac{\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}} \delta \tilde{\mathbf{x}} \quad (3.78)$$

Since a sigma point transformation aims to preserve the mean and covariance a transformation given by $|\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})/\partial \tilde{\mathbf{x}}| = 1$ is a valid sigma point transformation since the PDFs of \mathbf{x} and $\tilde{\mathbf{x}}$ are the same. However, if $|\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})/\partial \tilde{\mathbf{x}}| = c$ where c is a constant for all \mathbf{x} in the vicinity of the Gaussian PDF parameterized by the sigma points, then the PDF is also preserved by the scaling factor c . The PDFs can then be written as $f_{\mathbf{x}}(\mathbf{x}) = f_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}})/c$. This result is analogous to Theorem 2 since the admissible region prior must be preserved and the PDF must be preserved for sigma point transformations, the scaling factor c is equivalent to ζ for admissible regions.

3.3.3 Transformations Over Time

General probability transformations also apply to transformations through time as shown by Park and Scheeres [89]. Here it is shown that the framework presented in this paper is consistent with these existing methods. Given an initial PDF for a system, it is often useful to know how that PDF changes over time. Consider the following system dynamics

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \quad (3.79)$$

where $\mathbf{x} \in \mathbb{R}^n$ and $t \in \mathbb{R}$. The solution is expressed as

$$\mathbf{x}(t) = \phi(t; \mathbf{x}_0, t_0) \quad (3.80)$$

where the subscript ‘0’ denotes the initial state, $\mathbf{x}(t_0) = \mathbf{x}_0$ and ϕ is the flow function satisfying

$$\frac{d\phi}{dt} = \mathbf{f}(\phi(t; \mathbf{x}_0, t_0), t) \quad (3.81)$$

$$\phi(t_0; \mathbf{x}_0, t_0) = \mathbf{x}_0 \quad (3.82)$$

In the case of time transformations, the function ϕ is the transformation function $\mathbf{g}(\cdot)$. The PDF transformation of a dynamical system over time comes from analysis of the Fokker-Planck equation. If the system introduced above satisfies the Itô stochastic differential equation, then the time evolution of the PDF stochastic variable \mathbf{X} over time is given by the Fokker-Planck equation [90]

$$\frac{\partial f_{\mathbf{x}}(\mathbf{x}, t)}{\partial t} = - \sum_{i=1}^n \frac{\partial}{\partial \mathbf{x}_i} (f_{\mathbf{x}}(\mathbf{x}, t) \mathbf{f}_i(\mathbf{x}, t)) \quad (3.83)$$

assuming no diffusion terms. Park and Scheeres show the integral invariance of a PDF through the solution to this simplified Fokker-Planck equation for a system with no diffusion resulting in [89] [91].

$$f(\phi(t; \mathbf{x}_0, t_0), t) = f(x_0, t_0) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{x}_0} \right|^{-1} \quad (3.84)$$

which is the exact form given in Eqn. (3.3). Under Hamiltonian dynamics, Liouville's theorem proves that $|\partial \mathbf{x} / \partial \mathbf{x}_0| = 1$ for all time t since the transformation over time is a Canonical transformation [89]. For a Hamiltonian system Eqn. (3.84) simplifies further since the Jacobian term evaluates to unity. Thus, if the PDF is known at any time, it is known for all time for Hamiltonian systems. This exactly matches with Theorem 2 since $|\partial \mathbf{x} / \partial \mathbf{x}_0| = \zeta = 1$ and at any time t the PDF is given by $\zeta f(\mathbf{x}_0, t_0) = f(\mathbf{x}_0, t_0)$.

3.4 Discussion

The results presented in this paper show that, following the Principle of Transformation Groups, transformations of admissible region probabilities are only possible under strict conditions outlined by Theorem 2. This restriction ensures that the prior probabilities assigned originally are equivalent to the prior probabilities assigned after a transformation of the state space. In other words, this restriction ensures that the prior probabilities are equivalent regardless of how the problem is posed, which is consistent with the Principle

of Transformation Groups. Notably acceptable transformations include linear transformations and transformations with constant Jacobians over the admissible region. If a nonlinear transformation is applied to an admissible region prior that does not satisfy Theorem 2, then the resulting prior is necessarily a mis-representation of the statistical representation of the admissible region under the Principle of Transformation Groups. Furthermore, if a filter is instantiated from this improperly transformed prior then it may cause unnecessary inefficiency in filter convergence. However, once every state in the admissible region becomes observable then Theorem 1 can be applied to transform the true a posteriori PDF with appropriate $\tilde{\mathbf{x}} = \mathbf{g}(\mathbf{x})$ as desired. As such, for any filter to be properly instantiated under the Principle of Transformation Groups, the admissible region should remain expressed in the state space of the original admissible region prior formulation unless either Theorem 2 or Corollary 3 is satisfied.

As mentioned, this approach which preserves prior probability across transformations is only an alternative formulation of the problem. It is up to the user to decide which approach is appropriate for the specific application. The next section shows a comparison of this alternative Principle of Transformation Groups approach with a traditional application of the probability transformations.

3.5 Simulation and Results

To demonstrate probability transformations as applied to admissible regions, consider the observation of an object in LEO from an observer in Socorro, NM. Following the approach described in [76], the measurement vector is given by,

$$\mathbf{y} = \begin{bmatrix} \alpha & \delta & \dot{\alpha} & \dot{\delta} \end{bmatrix}^T \quad (3.85)$$

with the object state vector,

$$\mathbf{x} = \begin{bmatrix} \mathbf{r} & \mathbf{v} \end{bmatrix} \quad (3.86)$$

where \mathbf{r} and \mathbf{v} are position and velocity of the space object. The state matrix may also be represented by the topocentric spherical coordinates,

$$\tilde{\mathbf{x}} = \begin{bmatrix} \alpha & \delta & \dot{\alpha} & \dot{\delta} & \rho & \dot{\rho} \end{bmatrix}^T \quad (3.87)$$

For this observation model the undetermined states are given by $\tilde{\mathbf{x}}_u = [\rho \dot{\rho}]$, where ρ is the range and $\dot{\rho}$ is the range-rate. The true state of the object at time t_0 is given in canonical units as

$$\mathbf{r} = \begin{bmatrix} -0.9281 \\ -0.0489 \\ 0.6167 \end{bmatrix} \text{DU} \quad \mathbf{v} = \begin{bmatrix} -0.5171 \\ 0.1292 \\ -0.7662 \end{bmatrix} \text{DU/TU} \quad (3.88)$$

where $1 \text{ DU} = 6378 \text{ km}$ and $1 \text{ DU/TU} = 7.90538 \text{ km/s}$. An initial series of 2 measurements of the inertial bearings are gathered at 20 second intervals producing the following determined states, or observation, vector

$$\mathbf{x}_d = \begin{bmatrix} -3.0337 \text{ rad} & -0.0538 \text{ rad} & -0.1003 \text{ rad/TU} & -0.4482 \text{ rad/TU} \end{bmatrix} \quad (3.89)$$

From this information an admissible region can be constructed. The admissible region is then constructed such that the constraint hypotheses give a region where $10000 \text{ km} \leq a \leq 50000 \text{ km}$ and $e < 0.4$. A set of 5000 points are uniformly sampled from the admissible region to demonstrate the requirements on admissible region transformations and are shown in Figure 3.1. The upper bound on semi major axis is given by the solid line and the upper bound on eccentricity is given by the dotted line in Figure 3.1.

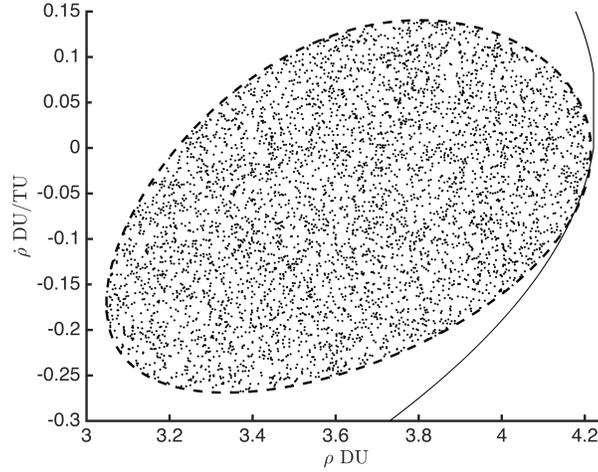


Figure 3.1: A set of 5000 points sampled uniformly from the admissible region.

Initial orbit determination methods can then use these sampled points to initiate particle filters or multiple hypothesis filters to process new observations. For these particle filter methods, the state vector can be converted to cartesian coordinates for propagation. However, this involves a transformation of the state space which implies either Theorem 2 or Eqn. (3.3) must be applied. The transformation from $\tilde{\mathbf{x}}$ to \mathbf{x} is given by,

$$\mathbf{r} = \mathbf{o} + \rho \hat{\mathbf{l}} \quad (3.90)$$

$$\mathbf{v} = \dot{\mathbf{o}} + \dot{\rho} \hat{\mathbf{l}} + \rho \dot{\alpha} \hat{\mathbf{l}}_{\alpha} + \rho \dot{\delta} \hat{\mathbf{l}}_{\delta} \quad (3.91)$$

where,

$$\hat{\mathbf{l}}^T = \begin{bmatrix} \cos \alpha \cos \delta & \sin \alpha \cos \delta & \sin \delta \end{bmatrix}$$

$$\hat{\mathbf{l}}_{\alpha}^T = \begin{bmatrix} -\sin \alpha \cos \delta & \cos \alpha \cos \delta & 0 \end{bmatrix}$$

$$\hat{\mathbf{l}}_{\delta}^T = \begin{bmatrix} \cos \alpha \sin \delta & -\sin \alpha \sin \delta & \cos \delta \end{bmatrix}$$

and $\mathbf{o} \in \mathbb{R}^3$ is the observer position and $\dot{\mathbf{o}} \in \mathbb{R}^3$ is the observer velocity. This transformation is both one-to-one and onto as there is only one cartesian state corresponding to a given ρ ,

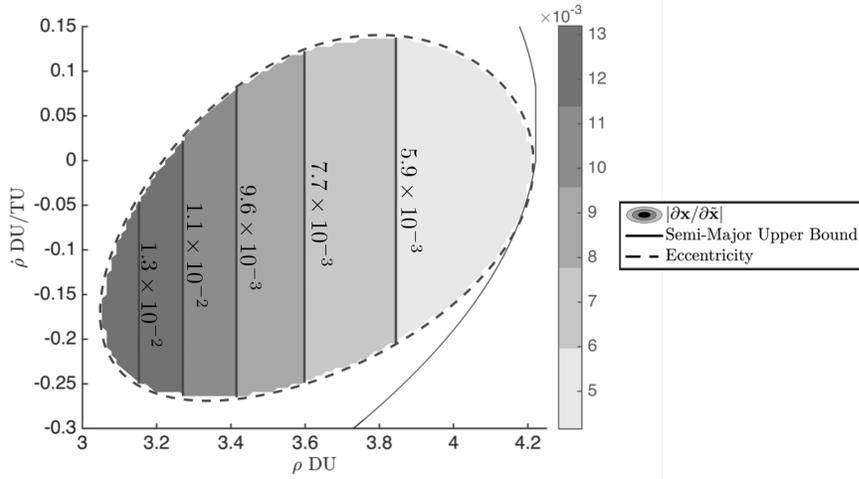


Figure 3.2: Values of $|\partial \mathbf{x} / \partial \tilde{\mathbf{x}}|$ evaluated for each particle $\mathbf{x}(t)$

$\dot{\rho}$, and observation vector. The Jacobian of this transformation is clearly a function of ρ and $\dot{\rho}$ and thus cannot be constant over the admissible region. After a single observation, the admissible region must still be expressed as a uniform distribution and transforming the sampled points into cartesian coordinates and expressing the admissible region prior in cartesian coordinates violates Theorem 2. To demonstrate this, Figure 3.2 shows the values of the determinant of the Jacobian over the admissible region. Since the probability transformation of an admissible region requires this value to be constant, it is clear that the transformation to cartesian coordinates violates Theorem 2.

With a single observation and no consideration of uncertainty, each of the points sampled from the admissible region necessarily has a uniform spatial distribution. New measurements should allow the admissible region to become observable by taking into account the new information provided by the measurements. Once the system is observable, by Corollary 3, the admissible region prior becomes a true PDF and the transformation is given directly by Eqn. (3.3). To test for observability, the condition number, $K(\mathbf{P}(t, t_0, \mathbf{x}(t)))$, for the local linearized observability gramian is computed for each value of ρ and $\dot{\rho}$ shown in Figure 3.1. The inverse of the machine epsilon value δ_m^{-1} is also plotted, which indicates that any $K(\mathbf{P}(t, t_0, \mathbf{x}(t))) > \delta_m^{-1}$ is essentially infinity due to the precision of the computer.

Then an additional observation is made 30 minutes after the initial set of observations. The additional observations are ingested by the particle filter and the updated observability gramian is computed. Figure 3.4 shows how the condition number for the observability gramian for each particle changes after the second observation is made. This change in condition number implies that the observability gramian becomes full rank after a second observation is made. At this point it is possible to transform the PDF expressed in terms of ρ and $\dot{\rho}$ into cartesian coordinates by direct application of Eqn. (3.3). Figure 3.5 shows the updated PDF after the second observation is made and can equivalently be expressed in cartesian coordinates by Eqn. (3.3).

To demonstrate the importance of Theorem 2 and Corollary 3, consider the process shown in Figure 3.3 by which the cartesian PDF for these observations can be determined. The original admissible region in ρ and $\dot{\rho}$ is represented by $\widetilde{\mathcal{R}}_{t_0}$ and after the second observation is made the PDF over the particles is given by $f_{\widetilde{\mathbf{x}}_u}(\widetilde{\mathbf{x}}_u)$. The admissible region given by \mathcal{R} represents the transformation of $\widetilde{\mathcal{R}}_{t_0}$ while the system is still unobservable. It has already been shown that this particular transformation does not satisfy Theorem 2, thus it is expected that the resulting PDF in cartesian space given by $f_{\mathbf{x}_u}(\mathbf{x}_u)$ will not be equal to the transformation of $f_{\widetilde{\mathbf{x}}_u}(\widetilde{\mathbf{x}}_u)$ into cartesian coordinates once the system is observable. This subtle difference in approach will generate two different PDFs for the particles resulting from the second observation and mathematically the PDF generated from the unobservable transformation is incorrect. Figure 3.6 shows the resulting PDF for the unobservable and observable transformations outlined in Figure 3.3. The PDFs shown are represented as the normalized histograms of the particles for each cartesian state after the resampling step in the particle filter. As can be seen, there is a slight bias in the particle filter results when instantiating the particle filter with an admissible region that has been transformed while unobservable. Note that these results do not imply that the particle filter will not converge to the correct state, but in certain cases, especially when the time between observations is short, there can be a noticeable bias in the particle filter. Figures 3.7 and 3.8 illustrate this

dependence on time. If the second observation is one hour after the initial observation, as seen in Figure 3.8, there is little difference between the PDFs because the particle filter eliminates the bias introduced by the different initial weighting of the transformed particles. However, if the second observation is only 10 minutes after the initial observation, as seen in Figure 3.7, there is a considerable difference in the PDF for the unobservable transformation.

This is exactly the reason for the application of the Principle of Transformation Groups; as the problem posed in both situations is identical, one would expect the resulting PDF to be identical as well. However, as these results show, for short observations, the prior probabilities determined for a given problem are not consistent across a general transformation even though the a priori information (the constraint hypotheses) is identical in both cases. Since initial orbit determination systems are often faced with short times between observations, it is important to ensure particle filters for initial orbit determination are instantiated properly. Thus, once a particle filter is instantiated in a given state space using an admissible region, the PDF must remain expressed in that state space until the system is observable when applying the Principle of Transformation Groups. The general exception to this are linear transformations which always satisfy the requirements of Theorem 2. The results of this work show that the consideration of the Principle of Transformation Groups does play a role in the formation of a PDF for a given admissible region. Ultimately, it is up to the user to determine if the effect this principle has on the PDF used for estimation is significant.

3.6 Conclusions

The probabilistic application of the admissible region to estimation is investigated under the Principle of Transformation Groups, which seeks to preserve prior probability in the lack of additional information on the system. The general theory of probability transformations is then presented and applied directly to the admissible region problem. It is found that

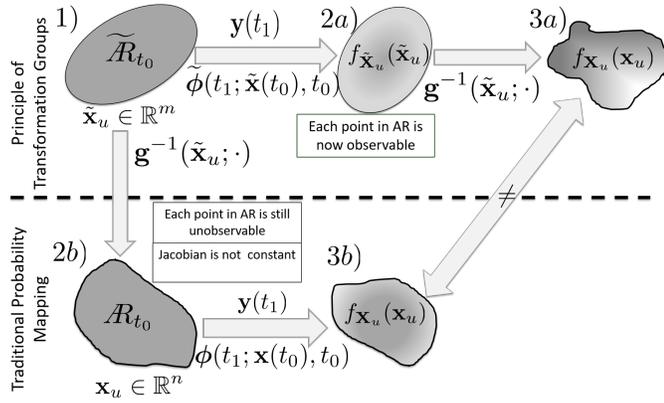


Figure 3.3: Outline of the two approaches for generating the PDF in cartesian coordinates

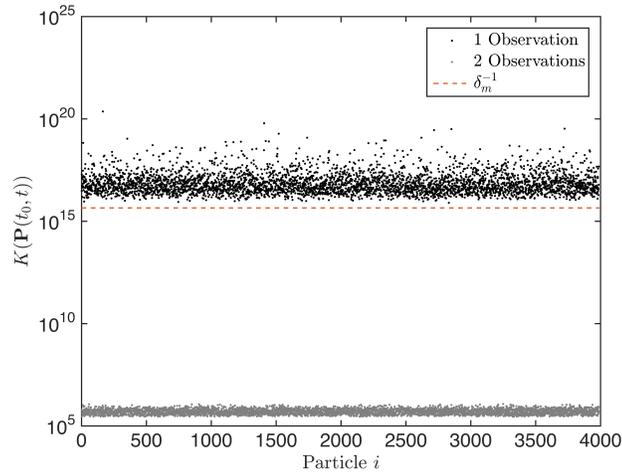


Figure 3.4: Condition number of $\mathbf{P}(t, t_0, \mathbf{x}(t))$ computed for each particle $\mathbf{x}_u(t)$

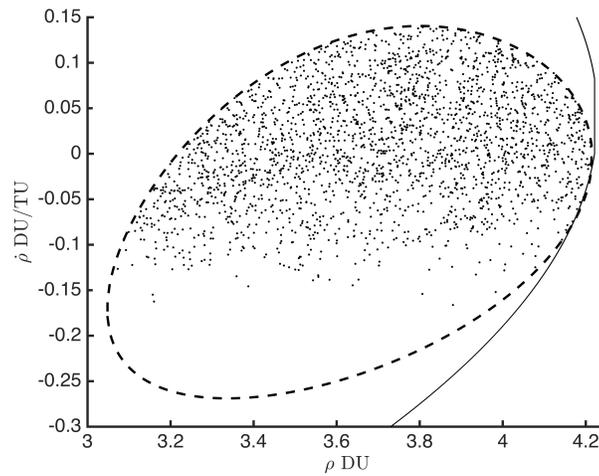


Figure 3.5: Admissible Region prior expressed in ρ and $\dot{\rho}$ 30 minutes after the initial observation

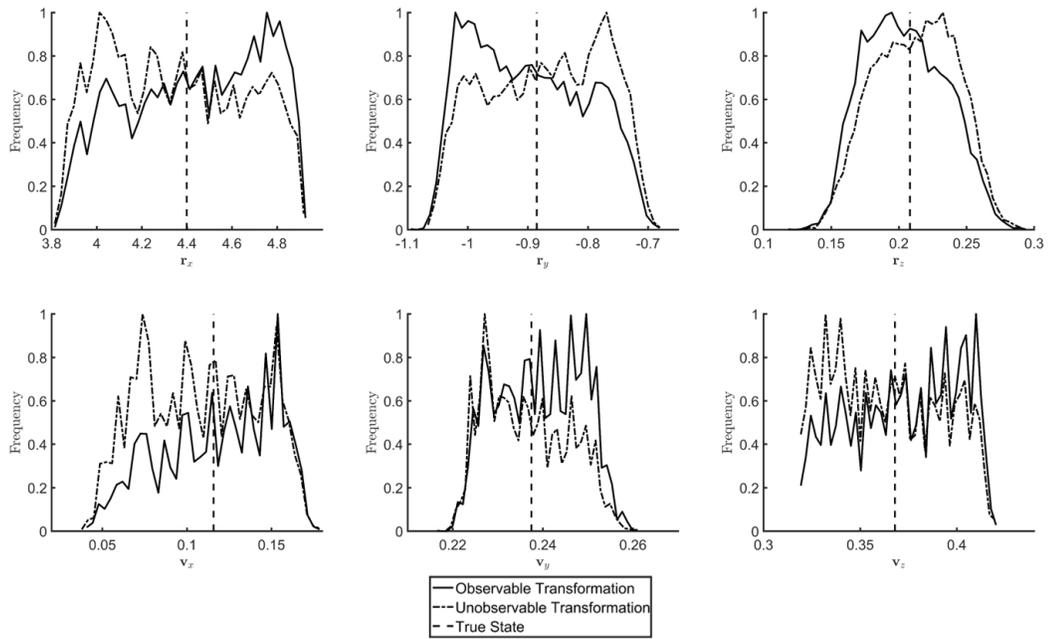


Figure 3.6: Difference between the cartesian PDFs, second observation is 30 minutes after the initial observation.

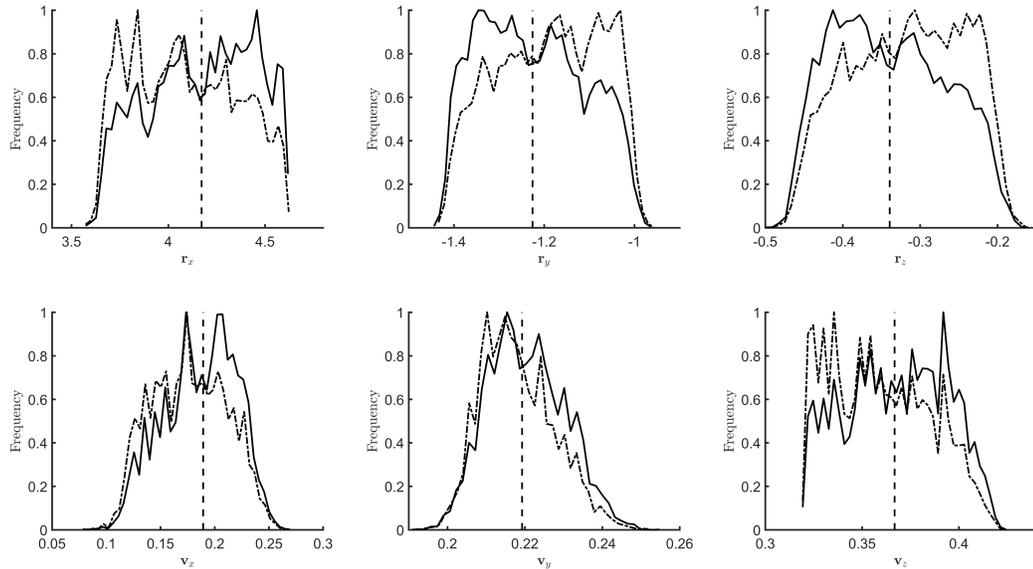


Figure 3.7: Difference between the cartesian PDFs, second observation is 10 minutes after the initial observation.

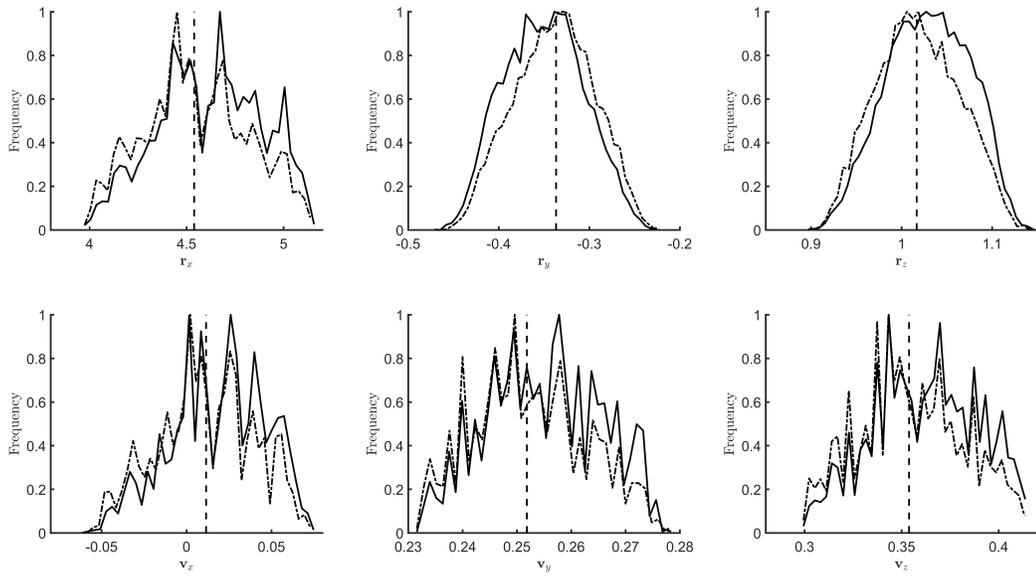


Figure 3.8: Difference between the cartesian PDFs, second observation is 60 minutes after the initial observation.

general probability transformations are invalid for admissible regions under the Principle of Transformation Groups since they do not preserve prior probability, thus a constraint on transformations for admissible region problems is defined. The constraint is shown to ensure the admissible region remains an uniform distribution regardless of the state space it is expressed in. This also shows that the statistical representation of the admissible region is consistent with Jeffreys' prior and satisfies the Principle of Transformation Groups. The results presented highlight the fact that a traditional probability mapping will tend to yield a different PDF than when the Principle of Transformation Groups is applied. As the observation time grows, the difference between the PDF is reduced, however over short time period observations it is shown the difference in the PDFs can be significant.

CHAPTER 4

EVIDENTIAL REASONING THEORY APPLIED TO THE ADMISSIBLE REGION

The findings of Chapter 3 show that the admissible region is not truly a general PDF due to the violation of fundamental probabilistic theorems. The solution offered in Chapter 3 treats the admissible region as an uninformative prior which preserves probability density across transformations, but still technically requires an assumption be made about the prior probability. What is desired is a methodology that can incorporate the fact that given a single set of unobservable measurements, the form of true underlying probability distribution is largely unknown, and thus any prior probabilities should not be assumed. Furthermore, it is desired for a methodology to also provide an indication as to when the hypothesized admissible region constraints are incorrect. This chapter introduces Dempster-Shafer evidential reasoning and the concept of ignorance as applied to the admissible region problem to provide such a methodology. As will be shown, this provides a powerful method by which not only are a priori assumptions on probability not necessary but also enables testing of validity of the assumptions used to construct the admissible region. The result is a more generalized estimation scheme which quantifies the lack of information, or rather the unobservability, as bounds on the true underlying probability distribution and enables the assignment of belief to states outside of the admissible region as well.

4.1 Dempster-Shafer theory

Traditional Bayesian probability is based on the pair (p, q) where p represents the probability that some hypothesis is true and q is the probability that some hypothesis is false [92].

Given some hypothesis h and state x , this is typically formulated as follows

$$p(x|h) \in [0, 1] \quad (4.1)$$

$$q(x|h) = 1 - p(x|h) \quad (4.2)$$

The limitation of traditional Bayesian probability is that the state of interest can only support or refute the hypothesis. However, in many real world applications there often exists states which do not inherently refute a hypothesis, but which also do not directly support the hypothesis. The Dempster Shafer (DS) approach involves utilizing the triple (p, q, r) which adds a quantification of plausibility, r , to address this imprecision in the system [52].

To introduce DS theory, first define Ω as the frame of discernment, the set which contains the states for over which evidence, or more specifically belief mass, should be assigned [10]. In traditional DS theory, Ω is defined as

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\} \quad (4.3)$$

where Ω is a mutually exclusive set of hypotheses for which exactly one hypothesis ω_i is true. The frame of discernment is also called the truth set since it contains the truth solution. DS theory then utilizes mass functions to assign mathematical probability to the power set of Ω which contains the $2^n - 1$ non-empty subsets of Ω (including Ω itself) as well as the empty set. Let $A \subset \Omega$ be a generic subset of the frame of discernment. The mass function is then defined as $m : 2^\Omega \rightarrow [0, 1]$, or equivalently the basic belief assignment (BBA) [55]. A given BBA must satisfy

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad (4.4)$$

A BBA could, for example, be defined as a plausibility measure on Ω . A simple form of such a plausibility measure is the membership function for $\omega \in A$ [55]. In general, many

different choices exist for the BBA and while it is often useful to define a more meaningful BBA based on the application, the selection of the BBA is subjective [53].

The mass function, or BBA, is used to define three useful quantities in DS theory,

$$\forall A \subseteq \Omega, \text{Bel}_i(A) = \sum_{\emptyset \neq B \subseteq A} m_i(B) \quad (4.5)$$

$$\forall A \subseteq \Omega, \text{Pl}_i(A) = \sum_{A \cap B \neq \emptyset} m_i(B) \quad (4.6)$$

$$\forall A \subseteq \Omega, \text{Q}_i(A) = \sum_{A \subseteq B} m_i(B) \quad (4.7)$$

$\text{Bel}_i(A)$ is the belief function and gathers evidence to support A , a given proposition. $\text{Pl}_i(A)$ is the plausibility function which gathers evidence that permits the occurrence of proposition A but does not necessarily support A directly. $\text{Q}_i(A)$ is called the commonality function by Shafer and is generally not used in a technical sense [10]. The belief and plausibility are related through duality

$$\forall A, \text{Pl}_i(A) + \text{Bel}_i(\bar{A}) = 1 - m_i(\emptyset) \quad (4.8)$$

where \bar{A} is the complement of A . If $m_i(\emptyset) = 0$ then it is implied that the solution must exist in Ω . Furthermore, it is only necessary to define one of the quantities, Bel, Pl, or m_i as each of the other quantities may be derived if the BBA is known or if either a belief or plausibility function is directly obtained. This is useful in applications where the BBA may be unknown but a plausibility function or belief function can be directly defined.

Another useful concept introduced from DS theory is the belief-plausibility gap, which is an indicator of ignorance in the system as utilized by Jaunzemis and Holzinger [93].

$$\text{Ig}_i(A) = \text{Pl}_i(A) - \text{Bel}_i(A) \quad (4.9)$$

Plausibility and belief are shown to be upper and lower bounds on the true probability by

Dempster [55]. The belief-plausibility gap can then be used as an indication of the amount of information in the system. It can be used to indicate when there is enough information for an under-determined state estimation problem to have a unique solution.

These summarized components of DS theory will be rigorously applied to the admissible region problem in the next section. This formulation avoids ambiguities in how to address the probabilistic nature of the admissible region for estimation by introducing belief and plausibility rather than just a probability or likelihood.

4.2 Belief Functions on Real Numbers

Traditional DS theory is derived for scenarios where Ω represents a finite set and the belief functions are defined over the power set of Ω . For general belief functions defined on real numbers, the belief functions are no longer defined over the power set of Ω . Let $\Omega = \mathbb{R}^n$ be a frame of discernment defined over the set of real numbers. Let $\mathcal{B}(\mathbb{R})$ denote the Borel sigma-algebra on the set \mathbb{R} and let $\mathcal{A} = \mathcal{B}^1(\mathbb{R}) \times \dots \times \mathcal{B}^n(\mathbb{R})$ be the cross product of n such Borel sigma-algebra. The belief density for $\Omega = \mathbb{R}^n$ is then defined as $m : \mathcal{A} \rightarrow [0, 1]$ satisfying

$$\int_{\mathcal{A}} m(z) dz = 1 \quad (4.10)$$

Note that in general the set \mathcal{A} includes both singleton and non-singleton subsets of \mathbb{R}^n . While Eqn. (4.10) poses no theoretical issues for a frame of discernment defined over real numbers, computationally the problem may become intractable, especially in higher dimensions. It is desired to define a restriction of the frame of discernment to reduce the computational requirements in application.

Consider $\Omega = \mathbb{R}^n$ where the subsets of \mathcal{A} are restricted to the singletons of Ω . The

basic belief assignment mass may then be defined as follows

$$m(A) = \begin{cases} m(\mathbf{a}) & A = \{\mathbf{a}\}, \mathbf{a} \in \mathbb{R}^n \\ 0 & \textit{otherwise} \end{cases}$$

which only assigns belief to the singletons of Ω [4]. Given that Ω is a countable set of points in \mathbb{R}^n then the summations in Eqns. (4.5) and (4.6) become infinite sums. Particularly this is useful in state estimation when a particular proposition $A \in \Omega$ cannot take on multiple values. However, Dubois and Prade [55] show that if the subsets of \mathcal{A} are restricted to the singletons of Ω then

$$\text{Bel}(A) = \text{Pl}(A), \quad \forall A \in \mathcal{A} \quad (4.11)$$

This implies that the basic belief assignment $m(\cdot)$ is simply a probability measure on Ω and the desired benefits gained through using DS theory are lost.

Thus, it is desired to have a more inclusive frame of discernment which permits both singleton and nonsingleton propositions while still remaining computationally tractable. Rogers and Costello show that, in general, it is sufficient to consider only a finite number of nonsingleton propositions [54]. The primary nonsingleton proposition which must be included in the frame of discernment is Ω itself, or the uncertainty proposition. The uncertainty proposition enables belief mass to be applied to the entire frame of discernment in addition to the individual singletons. The uncertainty proposition accounts for the fact that there may be situations in which a sensor may not reliably support any individual proposition, but given that the true proposition lies in Ω , it still supports assignment of all belief to Ω . The selection of the other nonsingleton propositions is assumed to be arbitrary, but chosen appropriately for the given problem.

Given that the initial orbit determination problem gives a frame of discernment over the real numbers, only a finite number of nonsingleton propositions should be defined to ensure

computational tractability. That is, only a finite number of sets from 2^Ω should be included as candidate for which belief mass can be assigned. Let the general frame of discernment defined for a belief function on real numbers as

$$\Omega = \{\{x\} \in \mathbb{R}^n\} \quad (4.12)$$

then the particular subset of 2^Ω of interest is

$$\Theta = \{\{x\} \in \mathbb{R}^n\} \cup \mathbb{R}^n \subset 2^\Omega \quad (4.13)$$

which is the union of a countable set of singleton state propositions and the full space \mathbb{R}^n and represents a computationally tractable set to define as opposed to the full power set. Let Θ be defined as the plausibility space, that is the space over which the plausibility and belief functions are defined. If it is possible to define a region \mathcal{X} inside of which the solution is hypothesized to lie such that $\mathcal{X} \subset \mathbb{R}^n$ then frame of discernment is given by

$$\Omega = \{\{x\} \in \mathcal{X}\} \quad (4.14)$$

and the plausibility space can be written as

$$\Theta = \{\{x\} \in \mathcal{X}\} \cup \mathcal{X} \quad (4.15)$$

Yet, Eqn. (4.15) does not account for the fact that evidence can be gathered to suggest the solution does not actually lie in \mathcal{X} as hypothesized. To fully include this possibility that the truth proposition does not lie in \mathcal{X} , let $\bar{\mathcal{X}}$ denote the concept of ‘none of the above’ (NOTA) for estimation. The frame of discernment is then defined as

$$\Omega = \{\{x\} \in \mathcal{X}\} \cup \bar{\mathcal{X}} \quad (4.16)$$

where $\bar{\mathcal{X}} = (\mathbb{R}^n \setminus \mathcal{X})$. The resulting plausibility space is then defined as

$$\Theta = \{\{x\} \in \mathcal{X}\} \cup \mathcal{X} \cup \bar{\mathcal{X}} \quad (4.17)$$

which implies that belief mass can be assigned, to states in \mathcal{X} , to \mathcal{X} itself, or outside of the subset in which the solution is thought to lie, giving the ability to identify changes to a system or, for instance, differentiate between objects under observation.

Note that the use of $\bar{\mathcal{X}}$ can instead be generalized into more useful alternative propositions. For instance, if all belief mass is assigned to $\bar{\mathcal{X}}$ given a set of measurements, it could indicate that the original assumptions which were used to construct the admissible region, and thus \mathcal{X} , are wrong. Thus, a potentially useful additional proposition is \mathcal{X}_2 which could represent the set of all potential states under a different hypothesis from the one used to create \mathcal{X} . This ability to attribute evidence to discriminate between correct or incorrect hypotheses is an example of the utility provided by implementing DS theory for the admissible region problem.

4.3 Combination of Evidence

The belief density defined in the previous section operates on single piece of evidence collected from a given source. An additional utility of DS theory is the flexibility in combining or fusing evidence from different sources. There exist many different forms of rules to combine evidence from different belief assignment functions [94] [95]. A general rule of combination for a given belief function is given by Dempster's combination rule

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{\eta}, \quad A \subseteq \Omega, A \neq \emptyset \quad (4.18)$$

$$\eta = 1 - \sum_{B \cap C = \emptyset} m_1(B)m_2(C) \quad (4.19)$$

where the belief functions m_1 and m_2 represent distinct pieces of evidence [55]. For instance, m_1 and m_2 could be a set of two sensors both providing independent evidence about the state of a system. The normalization factor η accounts for the degree of conflict between the two sources. Dempster's rule is a conjunctive rule that is both commutative and associative and thus can be used iteratively in estimation schemes to update belief assignment.

Dempster's rule is the subject of scrutiny due to potential issues such as Zadeh's paradox which yields a counterintuitive result if Dempster's rule is applied directly [96]. Dezert et. al. also presents arguments against the use of Dempster's rule of combination in certain situations [97]. In short, the problem arises when there is a source of evidence which get essentially treated as absolute truth, erasing the benefits gained by combining evidence provided by other sources. Problems also arise when the sources of information are not independent and a normalization factor must be included to correct for this dependence [98]. However, application of Dempster's rule to the admissible region problem does not suffer from any of these problems as it is assumed that any two given measurements are independent of one another. Zadeh's paradox is not an issue due to the treatment of the admissible region problem, since it is unobservable there is no belief mass being directly assigned to particular states, and as such it is unlikely for the evidence gained by a given source to be treated as absolute truth.

4.4 The Admissible Region BBA

The application of DS theory to the admissible region problem is initialized similarly to a traditional Bayesian approach. Since it is of interest to estimate the full state \mathbf{x} which is in \mathbb{R}^n the elements of the frame of discernment Ω must also be in \mathbb{R}^n for this admissible region problem. Following the construction of Ω outlined in the previous section, let $\mathcal{X} \subset \mathbb{R}^n$ represent the full set of admissible states defined by the admissible region and $\bar{\mathcal{X}} = \mathbb{R}^n \setminus \mathcal{X}$

represent the set of all inadmissible states. Mathematically, these sets are represented by

$$\mathcal{X} = \{(\mathbf{x}_u, \mathbf{x}_d) : \mathbf{x}_u \in \mathcal{R}\} \quad (4.20)$$

$$\bar{\mathcal{X}} = \{(\mathbf{x}_u, \mathbf{x}_d) : \mathbf{x}_u \notin \mathcal{R}\} \quad (4.21)$$

and represent the full n dimensional set of admissible states. Then Ω and Θ are fully defined as

$$\Omega = \{\mathbf{x} \in \mathcal{X}, \bar{\mathcal{X}}\} \quad (4.22)$$

$$\Theta = \{\mathbf{x} \in \mathcal{X}, \bar{\mathcal{X}}\} \quad (4.23)$$

where Θ is a fully exhaustive subset of the full power set of Ω , which must contain the solution. Note also that this is a fully generalizable problem formulation which can be applied to any unobservable system.

A simple outline of the posed problem is as follows, given a measurement \mathbf{y} it is desired to find a BBA of the form $m(\mathbf{x}|\mathbf{y})$ which assigns belief mass to elements of Θ . As noted, the state \mathbf{x} is partitioned into the determined state, which may be directly obtained from the measurements, and the undetermined state which may be unobservable. Assume that at time t_k a measurement \mathbf{y}_k is obtained. Through the independence property applied to belief functions [99], the BBA can be partitioned as

$$m(\mathbf{x}|\mathbf{y}_{0:k-1}) = m_u(\mathbf{x}_u|\mathbf{y}_{0:k-1})m_d(\mathbf{x}_d|\mathbf{y}_{0:k-1}) \quad (4.24)$$

since knowledge of \mathbf{x}_d does not impact the belief allocated to \mathbf{x}_u and vice versa. The determined states are directly observable through the measurements and as such it is known that the belief mass function is equivalently the probability mass function and Eqn. (4.24)

becomes

$$m(\mathbf{x}|\mathbf{y}_{0:k-1}) = m_u(\mathbf{x}_u|\mathbf{y}_{0:k-1})p_d(\mathbf{x}_d|\mathbf{y}_{0:k-1}) \quad (4.25)$$

where $p_k(\cdot)$ denotes a probability mass (or density) function. Similarly the plausibility function for the admissible region problem may be defined as

$$Pl(\mathbf{x}|\mathbf{y}_{0:k-1}) = Pl_u(\mathbf{x}_u|\mathbf{y}_{0:k-1})P_d(\mathbf{x}_d|\mathbf{y}_{0:k-1}) \quad (4.26)$$

where again the plausibility of the determined state is equal by definition to the belief and thus the probability. It is now of interest to determine the form of either the BBA $m_u(\mathbf{x}_u|\cdot)$ or the plausibility function. Note that the belief assignment and plausibility are both only conditioned on the sequence of measurements $\mathbf{y}_0, \dots, \mathbf{y}_k$. This contrasts the Bayesian instantiation of an estimation problem where the initial probability is conditioned both on the measurements as well as some a priori distribution. Since DS theory does not require any knowledge of this a priori distribution, it avoids altogether the issue that arises when this a priori distribution is either not known or uninformative.

The BBA for this problem is subject to several constraints from the problem formulation. Given that the states \mathbf{x}_u are undetermined, a single measurement does not offer evidence to substantiate any particular state being more valid than another. In a probabilistic sense, each state would be given a uniform probability however if there is no evidence for any of the states then belief should not be assigned to any of the states. This would give an essentially vacuous belief function which satisfies the following

$$Bel(\mathcal{X}) = 1 \quad (4.27)$$

$$Bel(V) = 0, \forall V \in \Theta \setminus \mathcal{X} \quad (4.28)$$

$$Pl(V) > 0, \forall V \in \Theta \setminus \mathcal{X} \quad (4.29)$$

A vacuous belief function is equivalently an indication that there is insufficient information to assign belief mass to any given state in the frame of discernment, but since Ω must contain the truth, the whole admissible region set is attributed all initial belief mass.

Given the problem formulation, the BBA $m_u(\mathbf{x}_u|\cdot)$ is a vacuous belief function given a single measurement. Through Eqn. (4.6) it is possible to show that there exists a concise linear relationship between Pl and m given the defined frame of discernment. Let this set of linear equations be represented by

$$\text{Pl}(\mathbf{x}|\mathbf{y}_{0:k-1}) = \mathbf{A}m(\mathbf{x}|\mathbf{y}_{0:k-1}) \quad (4.30)$$

$$\text{Pl} \begin{pmatrix} \mathbf{x}_1|\mathbf{y}_{0:k-1} \\ \mathbf{x}_2|\mathbf{y}_{0:k-1} \\ \vdots \\ \mathcal{X}|\mathbf{y}_{0:k-1} \\ \bar{\mathcal{X}}|\mathbf{y}_{0:k-1} \end{pmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} m \begin{pmatrix} \mathbf{x}_1|\mathbf{y}_{0:k-1} \\ \mathbf{x}_2|\mathbf{y}_{0:k-1} \\ \vdots \\ \mathcal{X}|\mathbf{y}_{0:k-1} \\ \bar{\mathcal{X}}|\mathbf{y}_{0:k-1} \end{pmatrix} \quad (4.31)$$

where the matrix $\mathbf{A} \in \mathbb{R}^{\ell \times \ell}$ where ℓ is the cardinality of Θ [70]. Note in general, if N is the number of points sampled from \mathcal{X} , then $\ell = 2^N + 1$ which accounts for each of the cross combinations of particles contained in \mathcal{X} . Again, this poses significant problems for computational tractability, especially for the matrix inverse operation [54]. A major simplification is to define a priori that the belief masses assigned to the cross terms are zero and excluding the resulting plausibilities, which gives $\ell = N + 2$. This matrix \mathbf{A} is defined based on the principle of least commitment, the idea behind which implies the BBA defined should never assign more belief mass than justified to elements of Ω [100, 101, 102].

Since \mathbf{A} is always an invertible matrix, there exists a direct form by which the belief masses can be found assuming the plausibility function is known and is simply given by

$$m(\cdot) = \mathbf{A}^{-1}\text{Pl}(\cdot) \quad (4.32)$$

assuming a suitable plausibility function may be found. Thus it is desired to determine a suitable candidate for the plausibility function, one of which is the probability of set membership in \mathcal{R}_i . Furthermore, since the combination of vacuous belief functions is still vacuous [4], it is desired to find a way to combine plausibility which may be used with Eqn. (4.32) to find a non-vacuous belief function.

An initial, potential candidate for a plausibility function is the probability of set membership. Let $P_i : \mathbb{R}^u \times \mathbb{R}^{z+1} \rightarrow \mathbb{R}^+$ denote the probability of set membership for a given state given the i^{th} hypothesized constraint

$$P_i[(\mathbf{x}_u \in \mathcal{R}_i)] \approx \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\|\mathbf{x}_u - \mathbf{x}_{u,B_\perp}\|}{\sqrt{2\operatorname{tr}\mathbf{P}_{\mathbf{x}_{u,B_\perp}}}} \right) \right] \quad (4.33)$$

where the method to determine $\mathbf{P}_{\mathbf{x}_{u,B_\perp}}$ is given by Worthy and Holzinger [76]. Then let this initial plausibility function for the admissible region problem be defined using Eqn. (2.54) as follows

$$P(\mathbf{x}_u \in \cap_{i=1}^c \mathcal{R}_i) \geq \left[\sum_i^c P_i(\mathbf{x}_u \in \mathcal{R}_i) \right] - (c - 1) \quad (4.34)$$

where again c is the total number of constraint hypotheses.

Referring back to Eqn. (4.6), by definition the plausibility of any state \mathbf{x} is lower bounded by $m(\mathcal{X}|\mathbf{y})$. Using Eqn. (4.34), a pseudo-plausibility function for a singleton hypothesis $\mathbf{x} \in \Theta$ may be defined as follows

$$\tilde{\text{Pl}}(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}_u \in \cap_{i=1}^c \mathcal{R}_i)p(\mathbf{x}_d|\mathbf{y}) \quad (4.35)$$

Eqn. (4.35) alone is not sufficient to define a plausibility function as while it may go to zero, plausibility of any state in \mathcal{X} must be lower bounded by the belief of \mathcal{X} itself. Thus a

well defined plausibility function is given by

$$\text{Pl}(\mathbf{x}|\mathbf{y}) = \tilde{\text{Pl}}(\mathbf{x}|\mathbf{y}) + m(\mathcal{X}|\mathbf{y}) \quad (4.36)$$

which gives the necessary lower bound on the plausibility of any given state.

It is then necessary to define how either plausibility or belief mass are assigned to \mathcal{X} as well as $\bar{\mathcal{X}}$. Let

$$m(\mathcal{X}|\mathbf{y}) = \frac{\int_{\mathcal{X}} \tilde{\text{Pl}}(\mathbf{x}|\mathbf{y}) d\mathbf{x}}{\int_{\mathcal{X}} d\mathbf{x}} \quad (4.37)$$

$$\text{Pl}(\bar{\mathcal{X}}|\mathbf{y}) = m(\bar{\mathcal{X}}|\mathbf{y}) = 1 - \max_{\mathbf{x} \in \mathcal{X}} \text{Pl}(\mathbf{x}|\mathbf{y}) \quad (4.38)$$

be the belief functions for these nonsingleton hypotheses. Direct application of Eqn. (4.6) gives a plausibility function for \mathcal{X} containing the sum of the belief masses attributed to all $\mathbf{x} \in \mathcal{X}$, as can be seen in the corresponding row of the matrix \mathbf{A} . Defining such a function, however, is difficult since a priori the belief masses, $m(\mathbf{x}|\mathbf{y})$, are unknown. If instead a well defined belief mass function for \mathcal{X} is available, then this issue can be avoided. Eqn. (4.37) represents such a belief mass function defined by the integral of the pseudo-plausibility function normalized by the volume of \mathcal{X} . It is straightforward to see that the belief function for $\bar{\mathcal{X}}$ is equivalent to the plausibility function and thus what is defined is the probability or likelihood that a given state is not in \mathcal{X} .

Given Eqns. (4.37) and (4.38), the linear relationship shown in Eqn. (4.31) is modified as follows

$$\begin{bmatrix} \text{Pl}(\mathbf{x}_1|\mathbf{y}_{0:k-1}) \\ \text{Pl}(\mathbf{x}_2|\mathbf{y}_{0:k-1}) \\ \vdots \\ m(\mathcal{X}|\mathbf{y}_{0:k-1}) \\ m(\bar{\mathcal{X}}|\mathbf{y}_{0:k-1}) \end{bmatrix} = \frac{1}{\zeta} \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} m \begin{pmatrix} \mathbf{x}_1|\mathbf{y}_{0:k-1} \\ \mathbf{x}_2|\mathbf{y}_{0:k-1} \\ \vdots \\ \mathcal{X}|\mathbf{y}_{0:k-1} \\ \bar{\mathcal{X}}|\mathbf{y}_{0:k-1} \end{pmatrix} \quad (4.39)$$

which provides a relationship between the computed pseudo-plausibilities of each state

coupled with the defined beliefs of the non-singleton propositions and the belief masses attributed to the states. The constant ζ ensures that the total belief function defined over Θ collectively integrates, or sums, to unity. Eqn. (4.39) provides a computationally tractable way to quantify plausibility and belief in the system, without considering all $2^N + 1$ combinations of propositions necessary if $\Theta = 2^\Omega$, by defining the belief masses for the desired non-singleton hypothesis directly and utilizing those belief masses to compute the resulting plausibilities of the states in \mathcal{X} . Furthermore, the plausibility of \mathcal{X} can be easily defined after taking the inverse relationship.

Eqn. (4.37) assigns plausibility to the collective set of states comprising the admissible region based on the pseudo-plausibilities of individual states in the admissible region. The plausibility of the proposition that the true state does not lie in the admissible region is then defined in Eqn. (4.38). This implies that given a single measurement, since the admissible region hypothesizes that the true lies in AR , the belief mass assigned to $\bar{\mathcal{X}}$ should be zero until evidence is gained that suggests otherwise.

With Eqns. (4.36), (4.37), and (4.38), a plausibility function is fully defined for Ω , and given the relationship given in Eqn. (4.32) the belief mass function can also be derived. The remaining concern is the combination of evidence from two independent observations. It is shown that given a single observation the BBA is vacuous and utilizing Eqn. (4.32) with the plausibility function just defined confirms this fact. The combination of two observations should yield a joint BBA which is no longer vacuous, however without knowledge of the form of the vacuous belief function the combination rule cannot be directly applied. It is desired to find an equivalent combination rule for plausibility, that is given $Pl_1(\cdot|\mathbf{y}_1)$ and $Pl_2(\cdot|\mathbf{y}_2)$ what is $Pl_{1\oplus 2}(\cdot|\mathbf{y}_1, \mathbf{y}_2)$. First consider Dempster's combination rule applied to the admissible region, for any singleton proposition $A \in \Omega$ the combination rule simplifies to

$$(m_1 \oplus m_2)(A) = \frac{m_1(A)m_2(A)}{\eta} \quad (4.40)$$

since $B \cap C = A$ is also a singleton [103]. Furthermore, if $B \cap C = \emptyset$ then by definition $m_1 = m_2 = 0$ so the normalization term η becomes 1. This is equivalently stating that there is no conflicting evidence being offered by the sensors. Thus, the combination rule as applied to propositions in the admissible region problem is simply given by

$$m_{1\oplus 2}(A) = m_1(A)m_2(A) \quad (4.41)$$

Applying Eqn. (4.6), the joint plausibility for the admissible region problem is given by

$$Pl_{1\oplus 2}(A|\mathbf{y}_1, \mathbf{y}_2) = \sum_{A \cap B \neq \emptyset} m_{1\oplus 2}(B) \quad (4.42)$$

which is simply the product of the individual plausibility functions.

$$Pl_{1\oplus 2}(A|\mathbf{y}_1, \mathbf{y}_2) = Pl_1(A|\mathbf{y}_1)Pl_2(A|\mathbf{y}_2) \quad (4.43)$$

Eqn. (4.43) now provides an iterative method by which plausibilities from independent measurements can be combined to create a joint plausibility. More importantly, through the use of Eqn. (4.32), once the joint plausibility is determined, the joint belief mass function can also be determined and thus both belief and plausibility can be found for a given state in Ω .

4.5 Observability and Ignorance

Since the condition $m(\mathcal{X}) = 0$ yields a probability over the rest of Θ by construction, this condition can be related to the observability of the system. Recall a system is observable if the state can be determined in finite time using only the measurements, or equivalently the output of the system [81]. By this definition, it is equivalent to stating when $m(\mathcal{X}) = 0$ the system must be observable since in a probabilistic sense there exist singleton propositions in Θ for which there is explicit support for being the truth state.

The $m(\mathcal{X}) = 0$ condition is related to ignorance by construction since it is only possible for $m(\mathcal{X}) = 0$ if $\text{Ig}(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$. Worthy and Holzinger show that the degree to which a given set of observation is observable is related to the following relationship

$$\Delta t = t_f - t_0 \ll \lambda_{\max}^{-1} \quad (4.44)$$

where t_f and t_0 are the times of the first and last observation respectively and λ is defined in [104]. For keplerian dynamics this condition is given by

$$\Delta t \ll \sqrt{\frac{\|\mathbf{r}\|^3}{2\mu}} \quad (4.45)$$

where $\mathbf{r} \in \mathbb{R}^3$ is the position of the satellite and if this condition is satisfied then the problem is essentially unobservable. This condition is essentially a lower bound on the order of magnitude of the observation time required for the state of a given object to become observable.

Combining these concepts, it can be shown that as long as the condition given by Eqn. (4.44) is satisfied, there will be significant ignorance in the system. The question is then when is there sufficient observability, or when is ignorance small enough to consider the system observable. In the ideal case when $\text{Ig}(\mathbf{x}) = 0 \forall \mathbf{x} \in \mathcal{X}$, then instead of utilizing a DS approach for sequential estimation, a traditional particle filter can be instantiated with the resulting probability distribution over \mathcal{X} . Recalling again that belief and plausibility can be treated as lower and upper bounds on probability, respectively,

$$m(\mathbf{x}) \leq P(\mathbf{x}) \leq \text{Pl}(\mathbf{x})$$

$$P(\mathbf{x}) = m(\mathbf{x}) + \Delta$$

$$\Delta \leq \text{Ig}(\mathbf{x})$$

By the above, $\text{Ig}(\mathbf{x})$ could be considered as the upper bound on the amount of error, de-

noted Δ , in the probability distribution used to instantiate a traditional Bayesian estimation scheme.

$$\text{tol} \geq \max_{\mathbf{x}}(\text{Ig}(\mathbf{x})) \quad (4.46)$$

Eqn. (4.46) then provides a user-defined parameter which indicates the maximum allowed error in the distribution used to instantiate a traditional Bayesian estimator, or equivalently the maximum allowed ignorance before the system is considered observable. As long as the tol parameter is exceeded, the DS filtering approach should continue to be used but once the condition is met a particle filter could be instantiated with the resulting belief function $m(\mathbf{x})$ defined over \mathcal{X} .

4.6 Application to Sequential Estimation

The direct application of Eqns. (4.43) for the admissible region problem is sequential estimation. In particular, the the particle filter is a standard estimation tool and it is desired to understand how to incorporate belief and plausibility into the particle filter formulation. There are several existing applications of DS theory to particle filtering which take advantage of either the belief assignment or plausibility functions as the primary weighting terms. Reineking rigorously applies the principles of DS theory to particle filtering and derives a general update equation for plausibility similar to Eqn. (4.43) [105]. Muños-Salinas et. al. demonstrate the application of DS theory to people tracking by instantiating multiple particle filters operating on belief, with initial belief mass distributions updated with Dempster's rule [98]. Rather than updating the belief functions directly, the approach implemented in this paper utilizes a plausibility update which then enables the corresponding belief value to be computed and then reverts to a traditional Bayesian estimator once the problem is observable. The sequential update for plausibility used in this paper is generated through

the use of Eqn. (4.43).

$$\text{Pl}_0(\mathbf{x}|\mathbf{y}_0) = P(\mathbf{x}_u \in \cap_{i=1}^c \mathcal{R}_i | \mathbf{y}_0) p(\mathbf{x}_d | \mathbf{y}_0) \quad (4.47)$$

$$\text{Pl}_k(\mathbf{x}|\mathbf{y}_{0:k}) = \text{Pl}_k(\mathbf{x}|\mathbf{y}_k) \text{Pl}_{k-1}(\mathbf{x}|\mathbf{y}_{0:k-1}) \quad (4.48)$$

The initialization of the belief assignment function is vacuous

$$m_0(\mathbf{x}|\mathbf{y}_0) = 0 \quad (4.49)$$

$$m_0(\mathcal{X}|\mathbf{y}_0) = 1 \quad (4.50)$$

and the update for the belief assignment function then comes through the inverse linear relationship defined by Eqn. (4.39)

$$m_k(\mathbf{x}|\mathbf{y}_{0:k}) = \zeta \mathbf{A}^{-1} \text{Pl}_k(\mathbf{x}|\mathbf{y}_{0:k}) \quad (4.51)$$

Eqn. (4.51) is also equal to $\text{Bel}(\mathbf{x}|\mathbf{y}_{0:k})$ by the definition of Θ . The DS particle filter methodology implemented for this work utilizes Eqns. (4.48) and (4.51) and is outlined in Algorithm 1.

4.7 Reduction to Bayesian Inference

The use of the concepts of belief and plausibility to instantiate a particle filter for the admissible region problem is a convenient way to avoid the ambiguities inherent in the direct application of Bayesian inference. However, it is still desired to determine when the concepts of belief and plausibility collapse back to standard Bayesian inference, that is when a traditional particle filter may be utilized. Consider again the linear relationship defined in Eqn. (4.31). It equivalently states that

$$\text{Pl}_k(\mathbf{x}|\cdot) = m_k(\mathbf{x}|\cdot) + m_k(\mathcal{X}|\cdot) \quad (4.52)$$

Algorithm 1 Uninformative Prior Dempster Shafer Particle Filter

```

1: procedure DEMPSTERSHAFFERFILTER( $\mathbf{y}_k, k, \Theta_{k-1}, \text{Pl}_{k-1}$ )
2:    $\Theta_k = \emptyset$ 
3:   if  $k = 0$  then ▷ Initialize belief, plausibility for  $k = 0$ 
4:     for  $i \leftarrow 1, N$  do
5:        $\Theta_k^i = \mathbf{x}^i, \mathbf{x}^i \in \mathcal{R}$  ▷ Sample admissible region, add propositions to  $\Theta_k$ 
6:        $\text{Pl}(\mathbf{x}^i) = 1$  ▷ (Eqn. (4.47))
7:        $m(\mathbf{x}^i) = 0$  ▷ (Eqn. (4.49))
8:        $\Theta_k^{i+1} = \mathcal{X}$  ▷ Add  $\mathcal{R}$  proposition
9:        $\text{Pl}(\mathcal{X}) = 1$ 
10:       $m(\mathcal{X}) = 1$  ▷ (Eqn. (4.50))
11:       $\Theta_k^{i+2} = \tilde{\mathcal{X}}$  ▷ Add ‘none-of-the-above’ proposition
12:       $\text{Pl}(\tilde{\mathcal{X}}) = 0$ 
13:       $m(\tilde{\mathcal{X}}) = 0$ 
14:   else
15:     for  $i \leftarrow 1, N$  do
16:        $\mathbf{x}^i = \phi(k; \mathbf{x}, k - 1), \mathbf{x} \in \Theta_{k-1}$  ▷ Draw from plausibility space
17:        $\Theta_k^i = \mathbf{x}^i$ 
18:        $\tilde{\mathbf{y}}_k = \mathbf{h}(\mathbf{x}^i, \mathbf{k}, k)$  ▷ (Eqn. (2.3))
19:        $\mathbf{x}_u^i, \mathbf{x}_d^i = \mathbf{g}(\mathbf{x}, \mathbf{k}, \cdot)$  ▷ Partition state vector
20:        $p(\mathbf{x}_d^i) = \exp -\frac{1}{2}(\mathbf{y}_k - \tilde{\mathbf{y}}_k)^T \mathbf{R}^{-1}(\mathbf{y}_k - \tilde{\mathbf{y}}_k)$ 
21:        $\text{Pl}(\mathbf{x}_u^i) = P(\mathbf{x}_u^i \in \mathcal{R}_k)$  ▷ Plausibility of  $\mathbf{x}_u^i$  (Eqn. (4.34))
22:        $\text{Pl}(\mathbf{x}^i) = \text{Pl}(\mathbf{x}_u^i)p(\mathbf{x}_d^i)$  ▷ Full state plausibility (Eqn. (4.36))
23:        $\text{Pl}_k(\cdot) = \text{Pl}(\cdot)\text{Pl}_{k-1}(\cdot)$  ▷ Update plausibility distribution (Eqn. (4.48))
24:        $\text{Pl}_k(\mathcal{X}) = \frac{\int_{\mathcal{X}} \tilde{\text{Pl}}(\mathbf{x}|\mathbf{y})d\mathbf{x}}{\int_{\mathcal{X}} d\mathbf{x}}$  ▷ Determine plausibility of  $\mathcal{R}$  (Eqn. (4.37))
25:        $\text{Pl}_k(\tilde{\mathcal{X}}) = 1 - \text{Pl}(\mathcal{X})$  ▷ Determine ‘none-of-the-above’ plausibility
26:        $m_k(\cdot) = \mathbf{A}^{-1}\text{Pl}_k(\cdot)$  ▷ Determine belief assignments (Eqn. (4.32))
27:   return  $m_k, \text{Pl}_k, \Theta_k$ 

```

for all the singleton propositions $\mathbf{x} \in \mathcal{X}$. Given that if $\text{Pl}(\mathbf{x}) = m(\mathbf{x})$ then $\text{Pl}(\mathbf{x}) = \text{Bel}(\mathbf{x}) = p(\mathbf{x})$ it is necessary that $m(\mathcal{X}) = 0$ for this construction of a DS particle filter to collapse to the traditional Bayesian implementation. However, the inclusion of the none-of-the-above hypothesis $\bar{\mathcal{X}}$ requires an additional condition to be considered. Since the operator is interested in the belief attributed to each of the states in Θ , it is necessary to include a check for which hypothesis has the most belief mass when $m(\mathcal{X}) = 0$. In particular if Eqn (4.53) is satisfied, then it suggests that, based on the plausibility function defined, there is more evidence suggesting that the truth solution does not lie in the original admissible region.

$$\max_{\mathbf{x} \in \mathcal{X}} m_k(\mathbf{x}|\cdot) < m_k(\bar{\mathcal{X}}|\cdot) \quad (4.53)$$

As such when this condition is met, it is necessary to then repeat the filtering process initialized with a different hypothesis for \mathcal{R} . If Eqn. (4.53) is not satisfied then it is sufficient to instantiate a particle filter with the resulting distribution over \mathcal{X} .

4.8 Simulation and Results

Several scenarios will be examined to demonstrate the utility of the application of DS theory to the admissible region problem. The first case demonstrates the use of the DS particle filter as outlined in Algorithm 1 and its reduction to essentially a traditional particle filter once enough observations are made as well as the indifference of DS theory to the problem formulation, elucidating the benefits gained by choosing to use DS theory to avoid ambiguities caused by different problem formulations in Bayesian theory. The second case demonstrates the utility gained by augmenting Ω with the ‘NOTA’ proposition, and highlights a potential area for future research. The last two cases demonstrates the use of the DS particle filter on real observation data.

4.8.1 Scenario 1

The observations utilized in the following example test cases are assumed to be captured from Atlanta, GA with an optical telescope with uncertainty parameters given as listed in Table 5.1. The observation scenario utilizes a measurement function of the form given in

Table 4.1: Parameters, Measurement Error and Parameter Uncertainty

Lattitude, Longitude, Altitude	[33.7490°N, 84.3880°W, 310m]
Right Ascension uncertainty, σ_α	10 arcseconds
Declination uncertainty, σ_δ	10 arcseconds
Timing error, σ_t	0.01 s
Position error (each axis), σ_o	1 m
Velocity error (each axis), $\sigma_{\dot{o}}$	1 cm/s

Eqn. (2.3)

$$\mathbf{y}_k = \mathbf{h}_u(\mathbf{x}_k; \mathbf{k}, t) \quad (4.54)$$

where \mathbf{k} is a set of parameters. Through Taylor series expansion of Eqn. (2.3), the measurement function may be approximated over short time periods as

$$\mathbf{y}_k(t) \approx \mathbf{y}_k(t_0) + \frac{\partial \mathbf{y}_k}{\partial t} \Delta t \quad (4.55)$$

where Δt represents a finite observation time. Given that Δt is not insignificantly small, Eqn. (4.55) enables both the angles and their time derivatives to be determined over the course of a short observation yielding

$$\mathbf{y}_k = \begin{bmatrix} \alpha \\ \delta \\ \dot{\alpha} \\ \dot{\delta} \end{bmatrix} \quad (4.56)$$

Table 4.2: True Orbit

Semi-major axis, a	6782.0 km
Eccentricity, e	.0007
Inclination, i	51.6°
RAAN, Ω	29.4°
Argument of perigee, ω	117.5°
True Anomaly, ν	20.0°

where α is the right ascension, δ is the declination of the space object and $\dot{\alpha}$ and $\dot{\delta}$ are the respective rates.

The initial state of the object used in this scenario is described in Table 4.2. The determined states are then given by

$$\mathbf{x}_d = \begin{bmatrix} \alpha \\ \delta \\ \dot{\alpha} \\ \dot{\delta} \end{bmatrix} \quad (4.57)$$

with undetermined states

$$\mathbf{x}_u = \begin{bmatrix} \rho \\ \dot{\rho} \end{bmatrix} \quad (4.58)$$

The measurements are provided by a series of 30 simulated observations at 20 second intervals initiated at 2017-1-17T08:00:00Z. The DS particle filter is implemented as described in Algorithm 1, with $\text{tol} = .01$ and is initialized from a sampling of the set membership function over the admissible region and the initial measurement and the belief and plausibility values are updated each time a new simulated observation is acquired. As described, the initial values of belief are zero for all propositions in Θ except for \mathcal{X} which is assigned belief of 1. The initial plausibility assigned is 1 for all propositions except for $\tilde{\mathcal{X}}$ the ‘NOTA’ proposition.

For comparison a traditional particle filter is instantiated with a purely uniform distribution over the admissible region and updated with each new measurement

$$p_{0,PF_1}(\mathbf{x}|\mathbf{y}_0) = \frac{1}{N} \quad (4.59)$$

where $N = 6000$ is the total number of samples.

To not only highlight the independence of DS from this initial distribution, but also to show that the particle filter eventually removes biases caused by arbitrary a priori probability assignment, the following arbitrary ‘PDF’ is selected for the initialization of the second particle filter

$$p_{0,PF_2}(\mathbf{x}|\mathbf{y}_0) = \frac{1}{C} \sin(\text{mod}(\rho, \pi)) \times \sin(\text{mod}(\dot{\rho}, \pi)) \quad (4.60)$$

where C ensures $p_{0,PF_2}(\cdot)$ integrates to 1. The use of this PDF is just for demonstration purposes only since the undetermined states are unobservable, thus in this lack of information any PDF is technically equally valid as long as it is not defined as a function of any determined states [84]. Each filter is updated with the measurements, and the first measurement \mathbf{y}_0 at time $t = 0$ is used to construct the admissible region from which the samples are drawn for the particle filter as well as the set from which samples of \mathcal{X} are drawn. Figure 4.1 shows the initial distributions for each filter.

As each measurement is ingested by the various filters, the PDFs or belief surfaces are updated accordingly. Figures 4.2-4.4 are included to convey how the plausibility and belief surfaces evolve over time compared to the probability density function of the two particle filters. These figures highlight two important aspects of the limitations inherent in attempting to assign prior probabilities in undetermined systems. First, both distributions shown in Figure 4.4c and Figure 4.4d seem to be converging to the truth solution appropriately even though they were initialized with very different prior probabilities. This is a direct result of the concept of ignorance, both initial distributions are essentially acceptable because

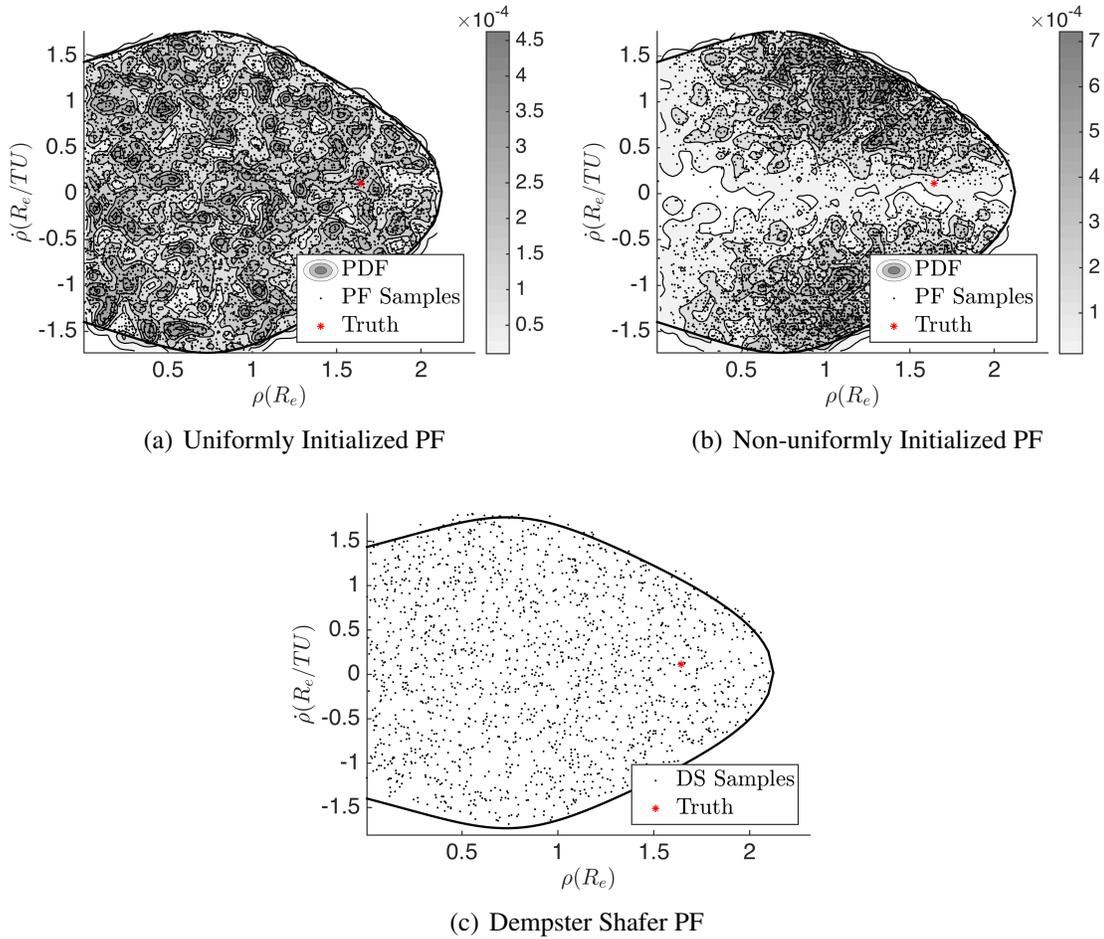


Figure 4.1: Initial distribution of states for Scenario 1 along with nominal \mathcal{A} constraint surface plotted.

without any a priori information all that can be truly stated is that the prior probabilities lie between 0 and 1 for all states. The strength of the DS particle filter, and DS theory in general, is that no prior probabilities need to be assigned. The prior information provided to the DS particle filter reflects the actual knowledge of the system, that is the unobservability (or similarly ignorance) of the system.

Another important implication of DS theory is that once there is no ignorance in the system, and the belief/plausibility functions and PDF are equivalent functions, the resulting PDF must be correct. The fact that the resulting distributions from initializing the particle filter with two different prior probabilities are different seemingly imply that one or both of these distributions could be incorrect given the initial unobservability of the system.

Plausibility and belief surfaces have the same general shape and as states gain additional belief mass they proportionately lose plausibility until belief and plausibility are equal and at this point traditional particle filter implementation is essentially equivalent. Figure 4.4 shows the near equality of the belief and plausibility values just before $m(\mathcal{X})$ goes to zero at time $t = 560$.

It is also of interest to examine how the belief mass attributed to both \mathcal{X} and $\bar{\mathcal{X}}$ change over the course of the simulation as well. Figure 4.7 displays these belief masses over the course of the simulation. As can be seen, as more observations are made, and equivalently more evidence is gathered, the evidence supporting assignment to the entire admissible region instead of individual states in \mathcal{R} is goes to zero. Note that while Figure 4.7 shows the belief and ignorance approaching 0 and 1 respectively, they do not attain these bounds. The algorithm utilized for the test scenarios continues to update the belief and plausibility values until the ignorance condition is met. If this condition is removed the algorithm would continue updating belief and plausibility, but the ignorance and maximum belief values would asymptotically approach 0 and 1 respectively as one would expect. Figure 4.6 demonstrates the true asymptotic nature of the ignorance (and equivalently the inverse asymptotic relationship with belief mass in \mathcal{X}). Figure 4.6 also shows the point in time where the $\text{tol} = 0.01$ condition is met and how ignorance changes after this condition is met. As expected, ignorance continues to fall even after the tolerance condition is met until it reaches a lower bound defined based on the errors and uncertainties in the system. Note that this is the general behavior of ignorance through the scenarios shown in this section, and this asymptotic behavior is also seen in the maximum belief mass assignment. The next example will demonstrate how it is possible to denote a potential incorrect admissible region hypothesis constraint assumption, or equivalently cases when the maximum belief is attributed to the NOTA proposition.

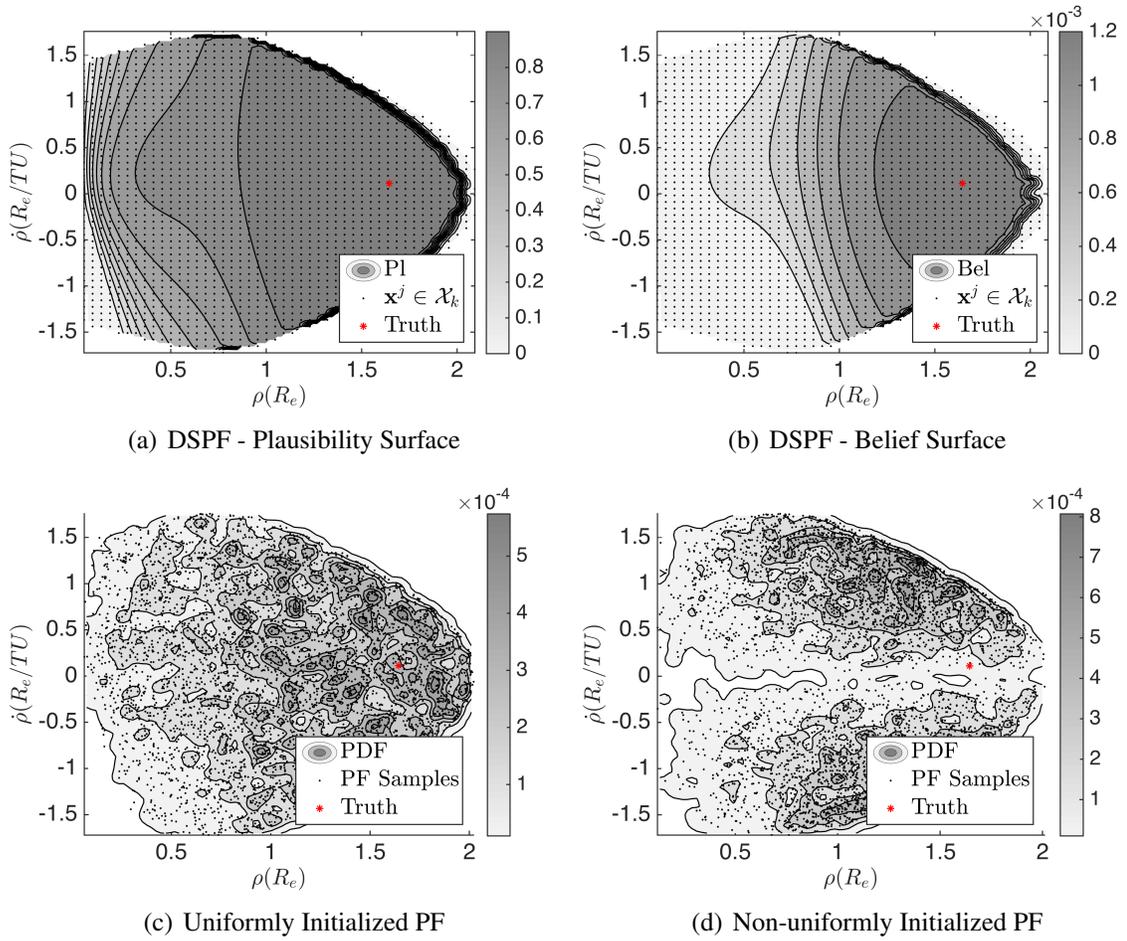


Figure 4.2: Belief, Plausibility, and Probability updates for $t = 40s$.

4.8.2 Scenario 2

Scenario 2 uses the same observation configuration as before but now with the object state given in Table 5.4. Note that this object is in a hyperbolic orbit, which implies that the admissible region constructed for all possible closed orbits will not contain the truth solution. Let \mathcal{A} represent the admissible region under the constraint hypothesis that the object has a closed orbit and let Ω follow from the definition presented in this paper. The set \mathcal{X} represents the set of all admissible closed orbits consistent with the measurements being captured, but it is known that the true state consistent with these measurements lies in $\bar{\mathcal{X}}$. The purpose of this case is to demonstrate that the belief mass assignments can provide

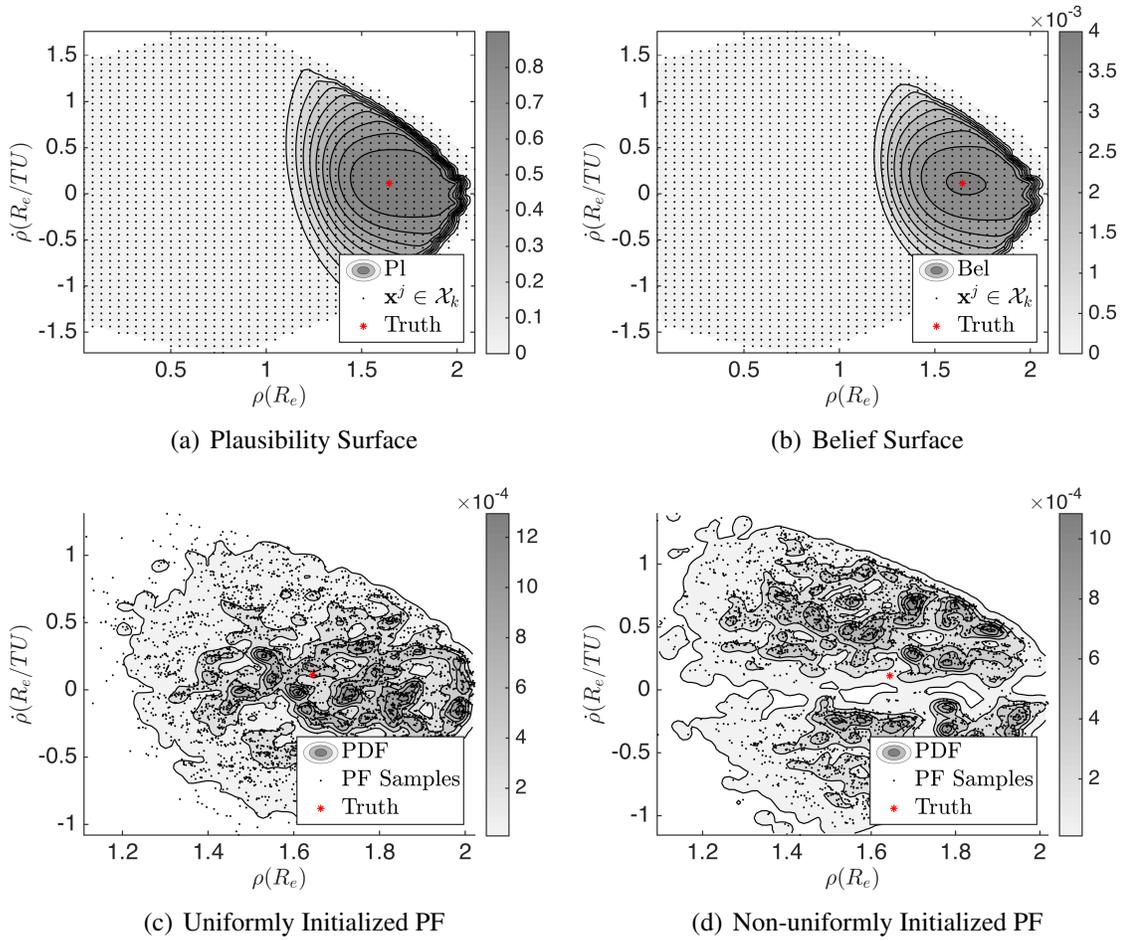


Figure 4.3: Belief, Plausibility, and Probability updates for $t = 120s$.

indications as to when the assumptions of the problem, namely the assumptions involved in defining \mathcal{R} , and thus \mathcal{X} and also $\bar{\mathcal{X}}$, are valid.

In lieu of showing the full evolution of the plausibility, belief, and probability surfaces over time, Figure 4.7 shows the belief, plausibility, and PF distributions at $t = 540$ and Figure 4.8 shows the evolution of the belief assignments to the admissible region, the none of the above (NOTA) set, and the maximum belief assigned to any state in the original admissible region similar to Figure 4.7. The desire is to show that if more belief mass is being assigned to the ‘NOTA’ set than any particular state in the admissible region, it may indicate that the truth solution does not lie in the original admissible region. While the ‘NOTA’ set in this example is simply defined as all states not in the admissible region, it

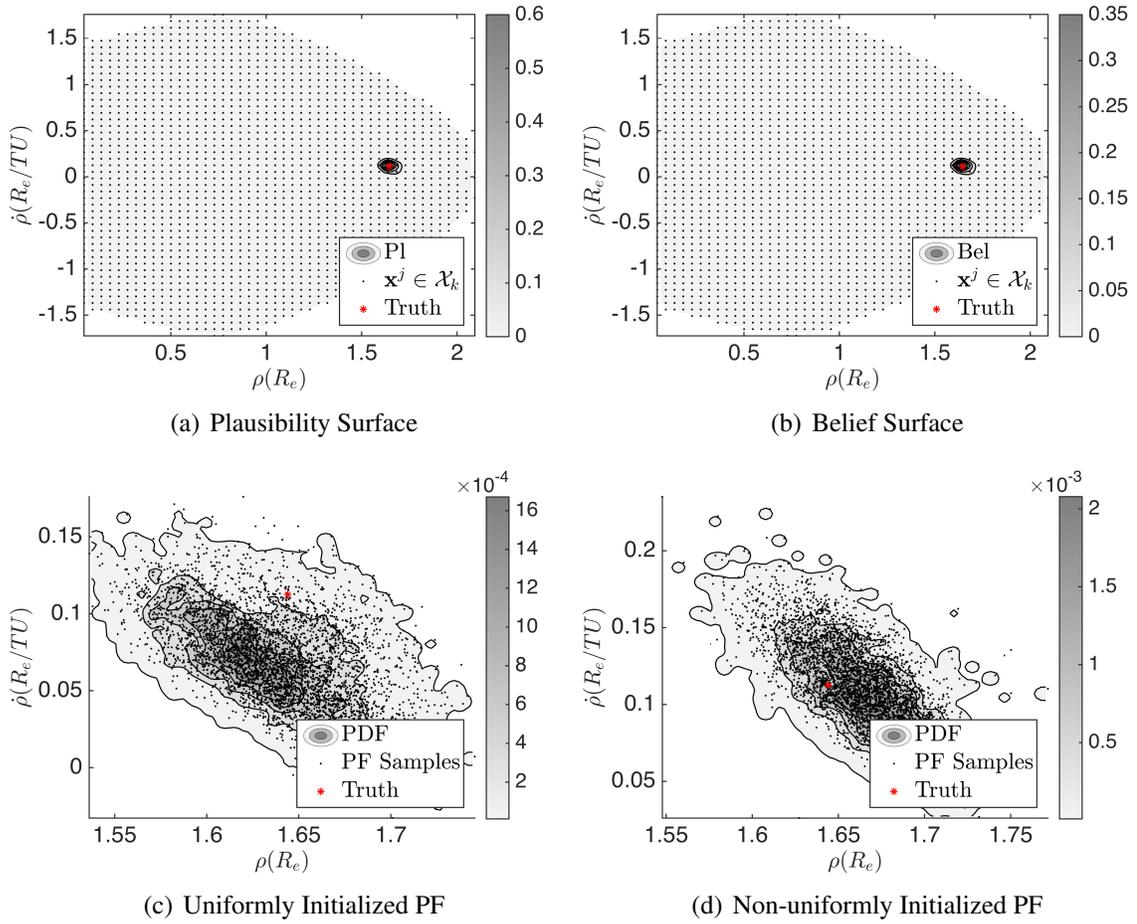


Figure 4.4: Belief, Plausibility, and Probability updates for $t = 480s$.

may also be more useful to define subsets of $\mathcal{X} \setminus \mathcal{R}$ which are valid alternative hypotheses which could be tested such that if belief masses are assigned to these not only is it an indication that the initial hypothesis is incorrect, but also provides an indication as to a correct alternative hypothesis. Figure 4.8 demonstrates that while there is a state in \mathcal{X} which appears to have support, or belief, there is more significant evidence suggesting that the true state is not in \mathcal{X} . This information could then be used to reinitialize the DS process with a different, more suitable hypothesis.

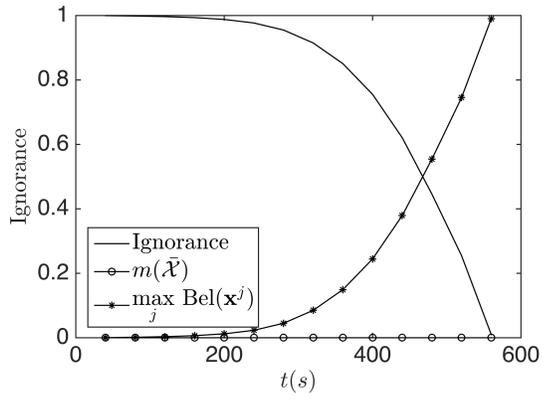


Figure 4.5: The belief mass assigned to the nonsingletons of Ω

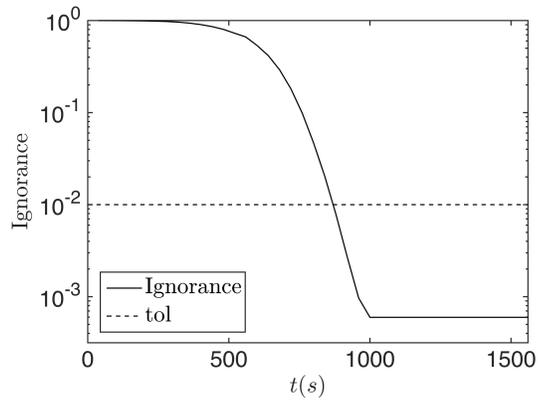


Figure 4.6: Asymptotic behavior of Ignorance for Scenario 1.

Semi-major axis, a	-20000 km
Eccentricity, e	1.5
Inclination, i	51.6°
RAAN, Ω	29.4°
Argument of perigee, ω	117.5°
True Anomaly, ν	20.0°

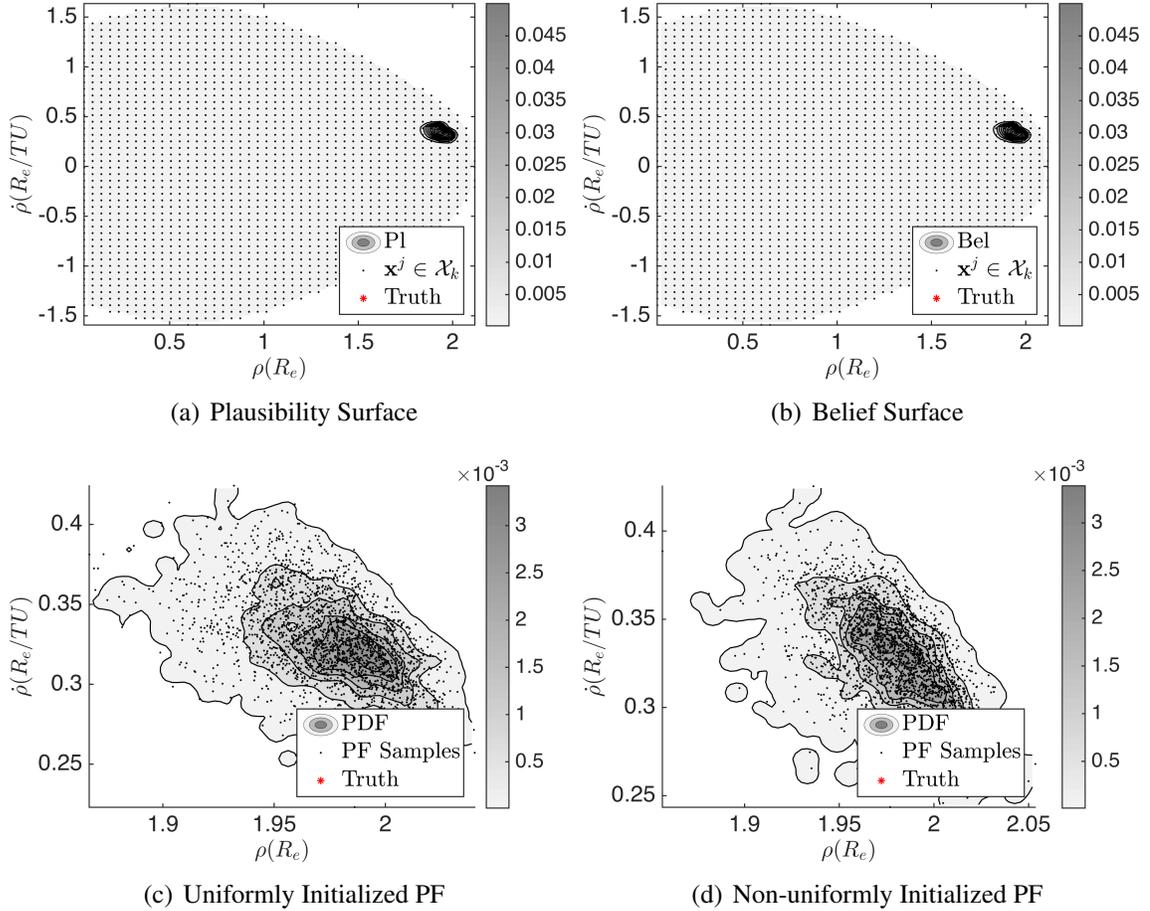


Figure 4.7: Belief, Plausibility, and Probability updates for $t = 540s$.

4.8.3 Scenario 3

This scenario deals with the application of the DS particle filter to real observation data from the Georgia Tech Space Object Research Telescope (GT-SORT) observatory. GT-SORT is a 0.5 meter, f/6 Raven class optical telescope mounted on a Paramount ME II

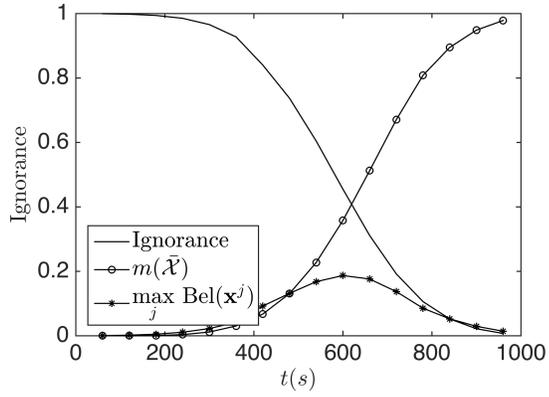


Figure 4.8: The belief mass assigned to the nonsingletons of Ω indicate the hypothesis may be incorrect.

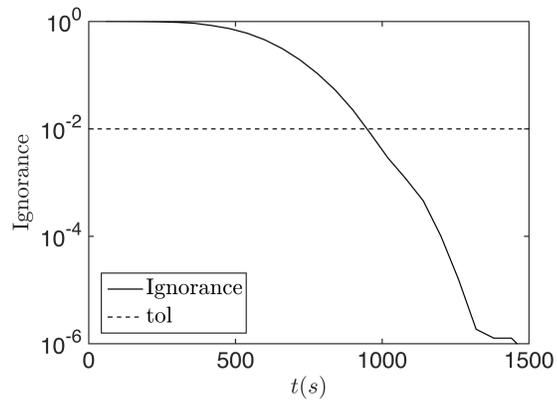


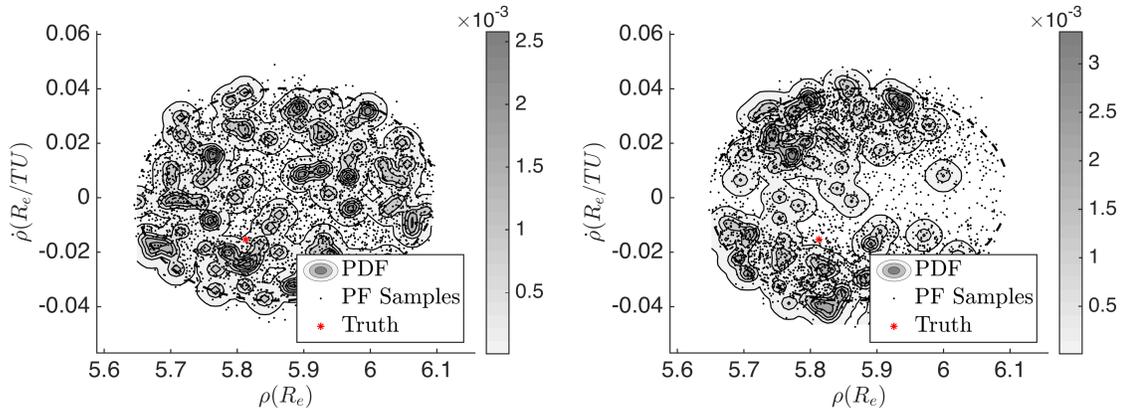
Figure 4.9: Asymptotic behavior of Ignorance for Scenario 2.

German equatorial mount¹ accompanied by a 60mm ZWO optics finder scope. Images are captured using the Grasshopper3², a 6MP, monochrome, CMOS electro-optical sensor on the telescope and a Zwo Asi174 MM CMOS sensor on the finder scope. This configuration provides GT-SORT 13.3×10.6 arcminute field of view and an instantaneous field of view of .3121 arcseconds. No filters were applied when the data was collected for the results shown in this section. The ANIK-F1 satellite (NORAD ID 28868) was selected to demonstrate a nominal case where the strict hypothesized constraints are true. The constraints used for the admissible region are the closed orbit energy constraint as well as an eccentricity constraint which bounds the admissible eccentricity for the object. Since ANIK-F1 is a known GEO object, the admissible region is initialized with $e = 0.1$ as the constraint on eccentricity.

Figure 4.10 shows the initial admissible region for the DS particle filter along with the initial samplings for the particle filters. Note that the true state of the object, estimated using TLE information, lies within the defined admissible region as expected. Measurements of ANIK-F1 were taken over a 24 hour period with initial measurements taken on 3 Mar 2017 and follow up observations taken on 4 Mar 2017. Figure 4.11 shows the belief and plausibility surfaces after the final measurement is ingested as well as the particle filter distributions. Note that in both the DS particle filter and the traditional particle filter the truth lies just outside the set of particles with non-zero belief or density respectively. This is due in some part to the errors in the TLE over the time period of interest as well as due to uncharacterized biases in the GT-SORT system and dynamics bias from the assumption of Keplerian dynamics. However, the purpose of this case is to show that the DS particle filter does indeed confirm the hypothesized constraint is valid once enough observation information is available as shown in Figure 4.12.

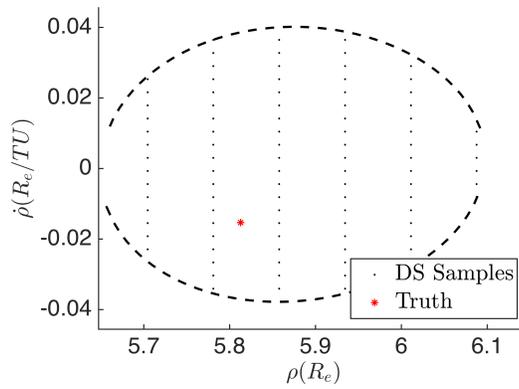
¹<http://www.bisque.com/sc/pages/ParamountMEII.aspx> last accessed on 12/12/2016

²<https://www.ptgrey.com/grasshopper3-usb3-vision-cameras> last accessed on 12/12/2016



(a) Uniformly Initialized PF

(b) Non-uniformly Initialized PF



(c) Dempster Shafer PF

Figure 4.10: Initial distribution of states for Scenario 3 along with nominal \mathcal{R} constraint surface and truth orbit plotted.

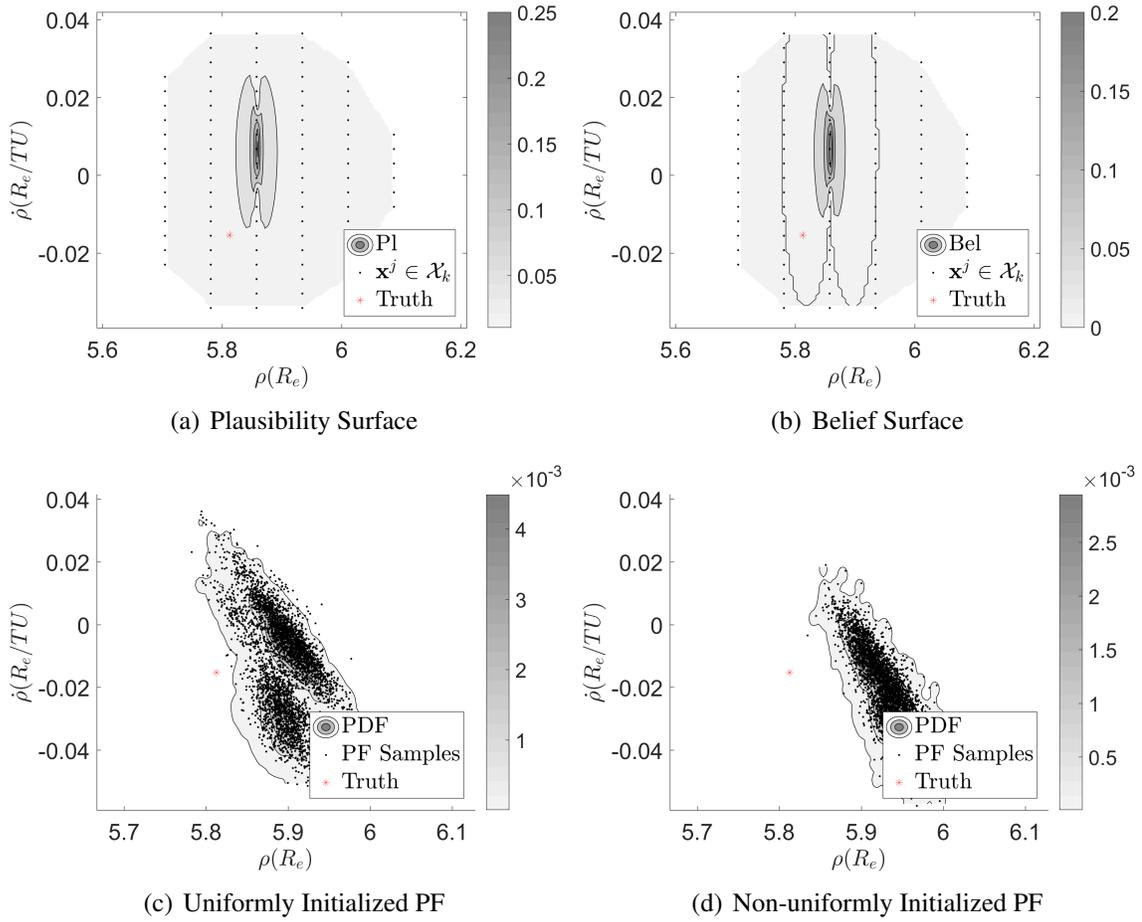


Figure 4.11: Belief, Plausibility, and Probability updates for $t = 2017-03-04T09:40:04$.

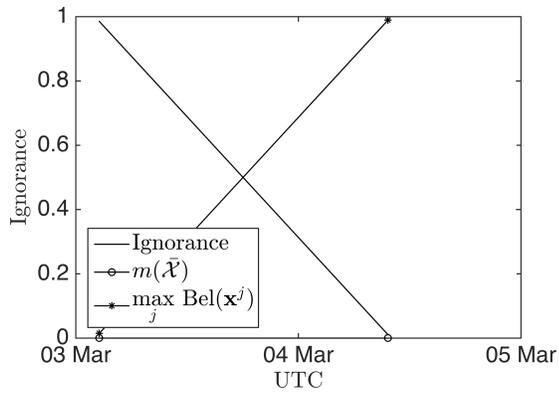


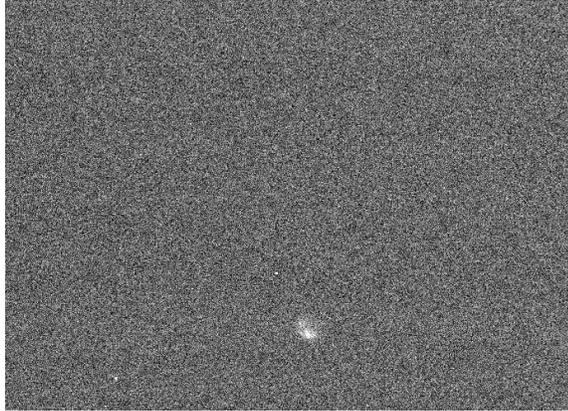
Figure 4.12: The belief mass assignments to the propositions in Θ .

4.8.4 Scenario 4

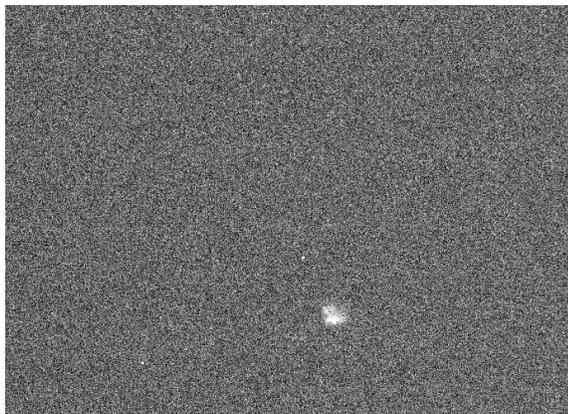
The final test case demonstrates the use of this approach on real observation data when the hypothesized constraints are too strict and exclude the true state. The object tracked is the defunct Cosmos 1247 satellite (NORAD ID 12303) with the initial observation taken at 2017-3-4T01:59:40Z. The observations were captured on the GT-SORT main sensor with an integration time of 1s and a frame rate of approximately 0.6 Hz for for 20 minutes. Figure 4.13 shows the tumbling nature of the object as the brightness varies over the different exposures of the object.

Similar to Scenario 3, the initial DS particle filter is instantiated with the assumption that the object observed is a GEO object and as such has an imposed eccentricity constraint of $e = 0.1$. Figure 4.14 shows the initial sampling for each of the filters with truth not contained within the initial admissible region. Figure 4.15 shows the belief and plausibility surfaces along with the particle filter distribution at $t = 160s$, the time step before all belief mass is attributed to the ‘NOTA’ hypothesis for $e = 0.1$. Since it is clear from these results that the initial constraint is likely incorrect, it is desired to see how changing the constraint affects the belief assignments.

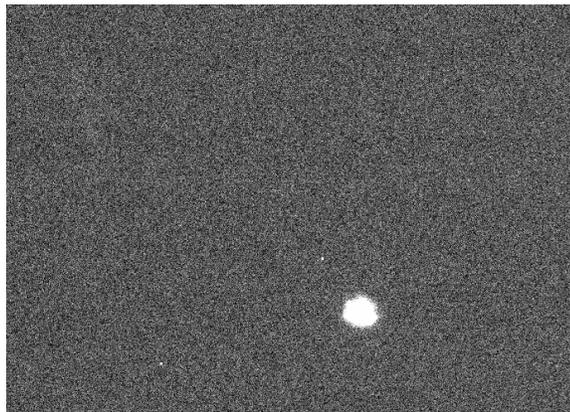
Figure 4.16 shows how the belief masses vary when the constraint is modified from a fairly strict assumption of $e = .1$ to a general closed orbit constraint $e = 1$. The true eccentricity of Cosmos 1247 is 0.5898033 which confirms what the DS particle filter results show for any constraints on eccentricity which exclude this value. Note that for the eccentricity constraints which fully encompass $e = 0.5898033$, the belief mass assigned to the ‘NOTA’ hypothesis remains very small. Note also that for this length of observation, even with the least restrictive assumption the problem isn’t fully observable after 20 minutes so while the DS filter is well posed for $e > .6$, it has not fully converged to a PDF after 20 minutes of observation.



(a) 2017-03-04T02:12:08.070Z



(b) 2017-03-04T02:12:14.310Z



(c) 2017-03-04T02:12:15.890Z

Figure 4.13: Images of the tumbling COSMOS 1247 object.

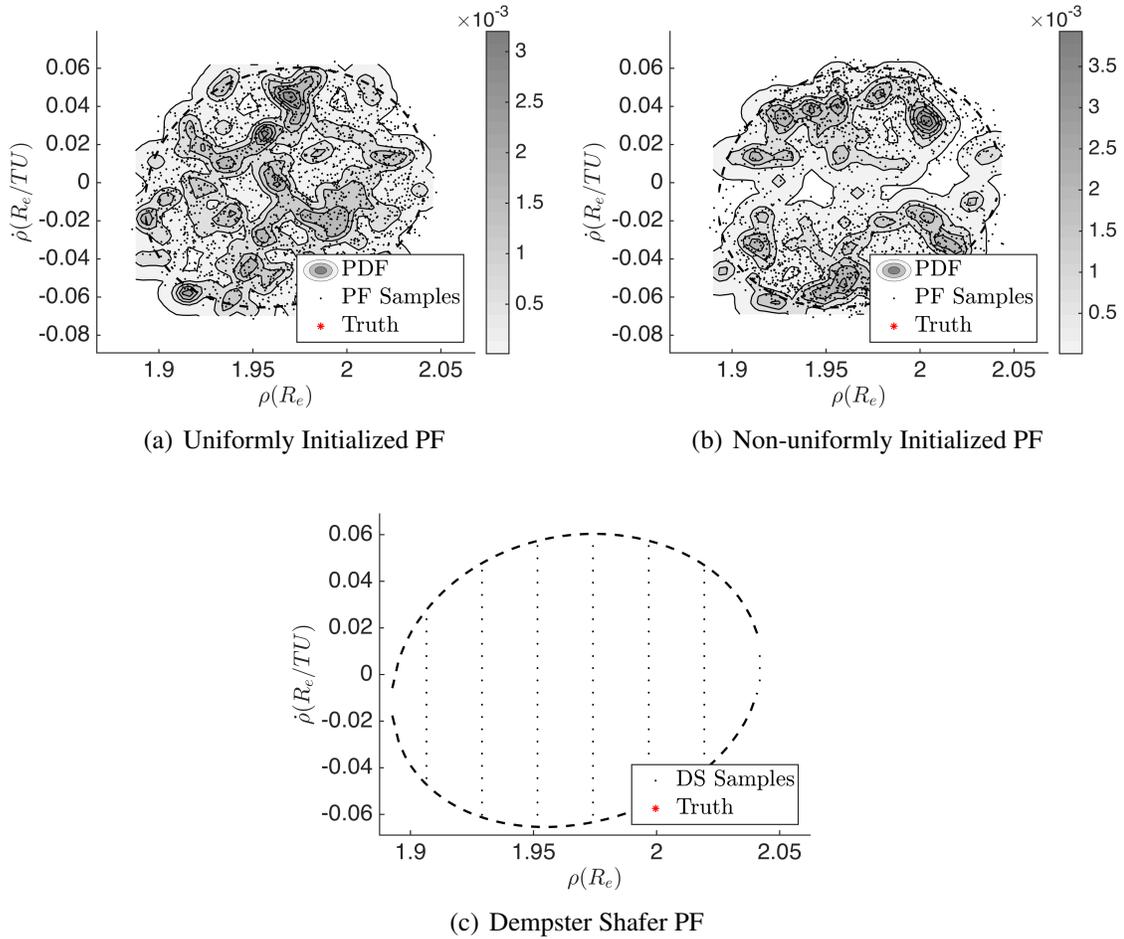


Figure 4.14: Initial distribution of states for Scenario 4 along with nominal \mathcal{A} constraint surface plotted.

4.9 Conclusions

This chapter introduces Dempster-Shafer (DS) theory and applies it to the admissible region problem. Due to the unobservable nature of the problem, there exists ambiguities in how probabilities are assigned to the states within the admissible region. DS theory avoids these ambiguities by utilizing plausibility and belief functions which are derived from a belief assignment which only assigns belief mass if there is direct evidence supporting an given state. Furthermore, it enables the assignment of belief mass not only to individual states, but also to sets of state solutions, and in particular the entire admissible region. Given a single observation, it is shown that the belief assignment function is vacuous for

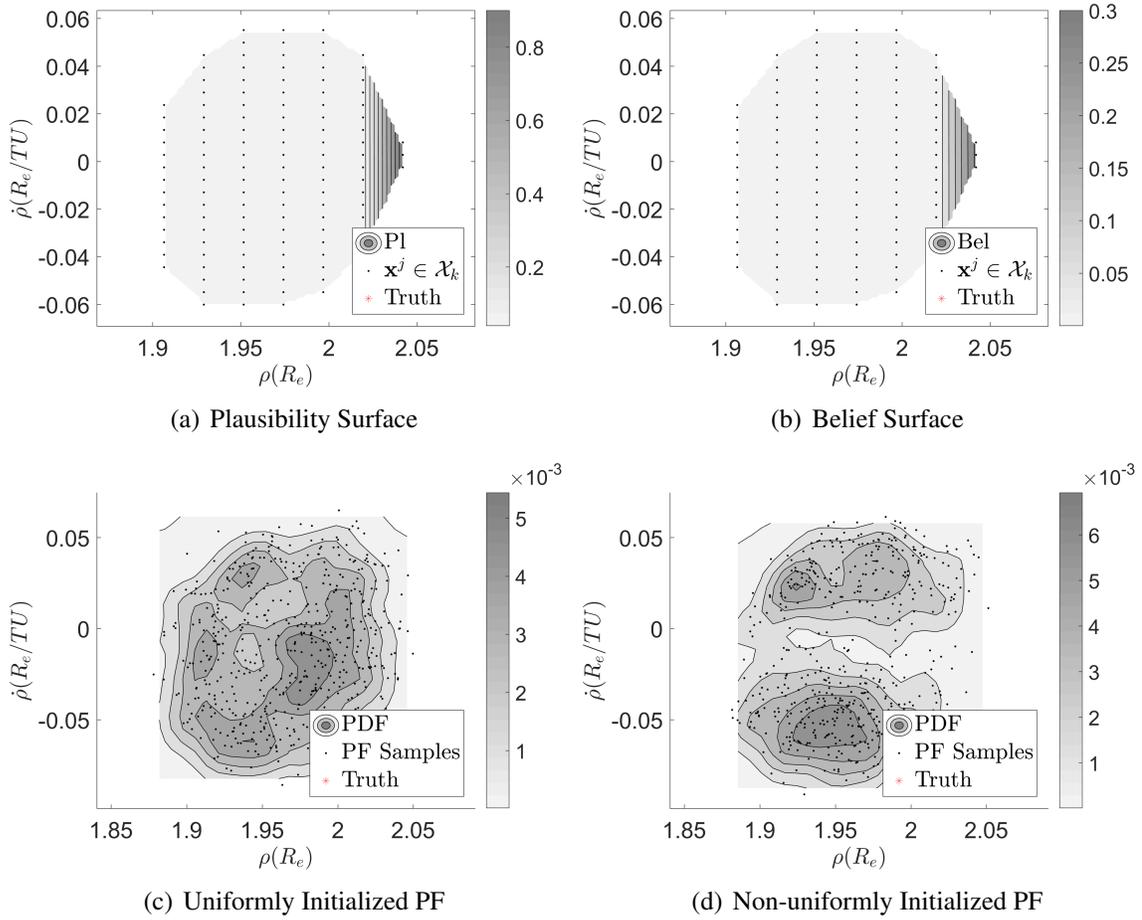


Figure 4.15: Belief, Plausibility, and Probability updates for $t = 160s$.

the admissible region problem and all belief mass is thus assigned to the full admissible region. A plausibility function is defined which assigns plausibility to each state in the admissible region, the admissible region itself, and the proposition that the state does not lie in the admissible region. It is shown that the combination of these plausibility functions enables a corresponding belief function to be defined through a linear relationship which upon sufficient observations collapses to traditional Bayesian inference. This DS particle filter is demonstrated on four example scenarios which convey the advantage of utilizing belief functions to initialize sequential estimation schemes for undetermined systems as opposed to traditional application of probability theory.

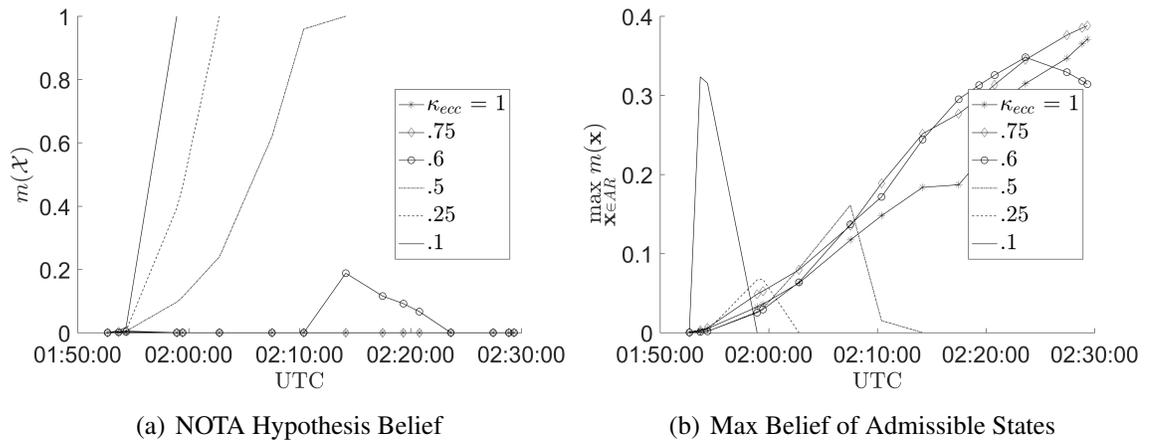


Figure 4.16: Results for how the belief assigned to $\bar{\mathcal{X}}$ changes with constraint strictness

CHAPTER 5

ASSOCIATION OF OBSERVATIONS VIA THE ADMISSIBLE REGION

The previous three chapters enable systemic uncertainties to be accounted for in the admissible region method and define how the admissible region should be treated probabilistically. However, it is known that if two observations are associated then the admissible regions, and furthermore the full n -dimensional manifold defined by combining \mathcal{R} with the determined states, must have a point (or several points) of intersection when propagated to a common epoch. Additionally if two or more observations can be shown to be associated and yield an observable combined set of measurements then it follows that the issue of unobservability presented in Chapter 3 can be altogether avoided. This chapter explores this idea further and uses the uncertain admissible region to define a computationally efficient optimization methodology which finds the point(s) of intersection of n -dimensional manifolds and uses hypothesis testing to define a probability of association for the given measurements. The result of the analytical gradient and Hessian information also results in a concise expression of the statistics about the resulting state estimate if a point(s) of intersection exists, and agrees well with a full particle filter implementation with significantly reduced computation time.

5.1 Association of Admissible Regions with Uncertainty

Park and Scheeres show that under Keplerian dynamics, admissible regions may be mapped forwards or backwards in time and the probability distribution remains invariant [89]. As such, this section will introduce a methodology to associate admissible regions using the topocentric spherical coordinates common to optical measurements of space objects, including the systemic uncertainties, for example measurement error, timing uncertainty, and observer position knowledge. The optimization based approach for associating observa-

tions is first defined in general. Then, the optimization problem formulation is simplified by partitioning the state space. Next, the uncertainties in the system are considered and a hypothesis testing method is introduced for associating the observations. Lastly, special cases of the optimization problem for association are discussed.

5.1.1 Optimization Approach for Set Intersection

Let \mathcal{R}_i now denote the joint admissible region generated from the i^{th} observation captured by an observer at time t_i . For each $\mathbf{x}_u \in \mathcal{R}_i$, a full state $\mathbf{x}(t_i)$ may be formed from the transformation

$$\mathbf{x}(t_i) = \mathbf{g}(\mathbf{x}_u, \mathbf{x}_d, \mathbf{k}) \quad (5.1)$$

where $\mathbf{k} \in \mathbb{R}^l$ is a set of parameters, which may also be propagated to any other time t . The set of all states $\mathbf{x}(t_i)$ that can be formed from $\mathbf{x}_u \in \mathcal{R}_i$ reside on a u -dimensional manifold in n -dimensional space (for optical observations of space objects this is a 2 dimensional manifold in a 6 dimensional space [66]). Let this time-varying, u -dimensional manifold for the i^{th} observation be defined at a time t by $\mathcal{X}_i(t)$ as follows

$$\mathcal{X}_i(t) = \{\mathbf{x}(t) : \mathbf{x}_u(t_i) \in \mathcal{R}_i\} \quad (5.2)$$

The set $\mathcal{X}_i(t) \subset \mathbb{R}^n$ contains all possible states, \mathbf{x} , at time t that could have originated from a particular undetermined state $\mathbf{x}_u \in \mathcal{R}_i$. Let \mathcal{R}_j now denote the admissible region generated from the j^{th} observation at time t_j . Likewise define,

$$\mathcal{X}_j(t) = \{\mathbf{x}(t) : \mathbf{x}_u(t_j) \in \mathcal{R}_j\} \quad (5.3)$$

and so forth for a given number of measurements. It is required that the number of measurements N satisfies $N \times d > n$ for there to be a solution to the general association problem.

Letting $N \times d = q$, then the association of a set of q measurements which yield the sets $\mathcal{X}_1(t_1) \subset \mathbb{R}^n, \dots, \mathcal{X}_q(t_q) \subset \mathbb{R}^n$ requires that there exist at least one state $\mathbf{x}(t)$ satisfying

$$\mathbf{x}(t_i) \in \mathcal{X}_i(t_i), \quad i = 1, \dots, q \quad (5.4)$$

Let $\mathcal{O}(t) \subset \mathbb{R}^n$ denote the set of such states, then the equivalent definition of this set is given by

$$\mathcal{O}(t) = \bigcap_q \mathcal{X}_i(t) \quad (5.5)$$

and it is required that

$$\mathcal{O}(t) \neq \emptyset \quad (5.6)$$

at any epoch time t for the set of q measurements to be associated.

For a case where $q = 2$, it is only necessary that there exist a state $\mathbf{x}(t)$ such that $\mathbf{x}(t) \in \mathcal{X}_i(t) \subset \mathbb{R}^n$ and $\mathbf{x}(t) \in \mathcal{X}_j(t) \subset \mathbb{R}^n$ at any common epoch time t . Let $\mathcal{O}_{ij}(t) \subset \mathbb{R}^n$ be defined as set of states residing in both $\mathcal{X}_i(t)$ and $\mathcal{X}_j(t)$.

$$\mathcal{O}_{ij}(t) = \mathcal{X}_i(t) \cap \mathcal{X}_j(t) \quad (5.7)$$

Then the condition for the two observations to be associated is given by

$$\mathcal{O}_{ij}(t) \neq \emptyset, \quad \forall t \quad (5.8)$$

The set $\mathcal{O}(t)$ satisfies $\text{card}(\mathcal{O}(t)) \in \mathbb{Z}^+$ if $\mathcal{O}(t) \neq \emptyset$ and thus $\mathcal{O}(t)$ represents the set of possible initial orbit state solutions for two observations as shown in [61]. This implies the measurements are associated only if the intersection of the sets \mathcal{X}_i and \mathcal{X}_j is non-empty and is the foundation of the intersection approaches discussed in [61] [60] [47] [59].

Note that these intersection based methods are typically defined for optical observations. The methods are well defined since the codimension of the admissible region for optical observations is 4. Specifically, the admissible region for optical observations is constructed through the fact that the angles and angle rates are known and only range and range-rate are undetermined. This implies for a given admissible region observation, $\mathbf{x}_d \in \mathbb{R}^4$ and $\mathcal{X}_i \in \mathbb{R}^6$. Thus, given two observations it is true that the general case of finding the intersection of \mathcal{X}_i and \mathcal{X}_j has either no solutions, or a single solution, because the problem is overdetermined. However, in a more general sense the intersection of two arbitrary sets \mathcal{X}_i and \mathcal{X}_j is not guaranteed to be an overdetermined problem. Consider a radar observer which provides range and range-rate measurements giving $\mathbf{x}_d \in \mathbb{R}^2$ but still $\mathcal{X}_i \in \mathbb{R}^6$. Given just two observations, the problem is now underdetermined and the general solution will find the intersection at an infinite set of points in \mathbb{R}^2 . Generalizing these intersection based methods for arbitrary sets requires then that, given two measurements it is required that

$$2u \geq n \tag{5.9}$$

to ensure the problem is overdetermined and thus ensure that there can exist a point solution at the intersection if it exists.

The intersection problem is a well posed optimization problem, where the point, or points, of intersection are the optimum of a formally defined distance metric [106]. Define the function $d(\mathbf{a}, \mathbf{b}) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+$ such that if an intersection exists then there exists $\mathbf{x}_i \in \mathcal{X}_i(t)$ and $\mathbf{x}_j \in \mathcal{X}_j(t)$ such that $d(\mathbf{x}_i, \mathbf{x}_j) = 0$. Further, if $d(\mathbf{a}, \mathbf{b})$ is nonnegative, symmetric, and satisfies the triangle inequality then d is a metric defined on $\mathcal{X}_i \times \mathcal{X}_j$ [107]. Figure 5.1 shows a representative example of how this distance metric is defined.

This distance metric may serve as the cost function for a constrained optimization approach to find the minimum distance between the sets \mathcal{X}_i and \mathcal{X}_j and is selected based on

the desired application. For example, a general distance metric is the Euclidean distance between the sets. The minima of the Euclidean distance directly identifies potential state solutions for the observations. The Mahalanobis distance between measurements is an additional distance metric that may be used to account for the measurement uncertainties in the system [108].

Alternatively, the control distance metric may be used in situations when attempting to associate measurements of potentially maneuvering spacecraft [109]. Intersection based observation measurement association methods have difficulty in associating measurements from maneuvering spacecraft due to the fact that, by maneuvering, the likelihood that these methods are able to find a consistent solution is greatly reduced. The control distance finds the optimal maneuvered trajectory for two given observations, and by utilizing it as the distance metric for association, one can determine the probability that two observations are associated even if the spacecraft is maneuvering.

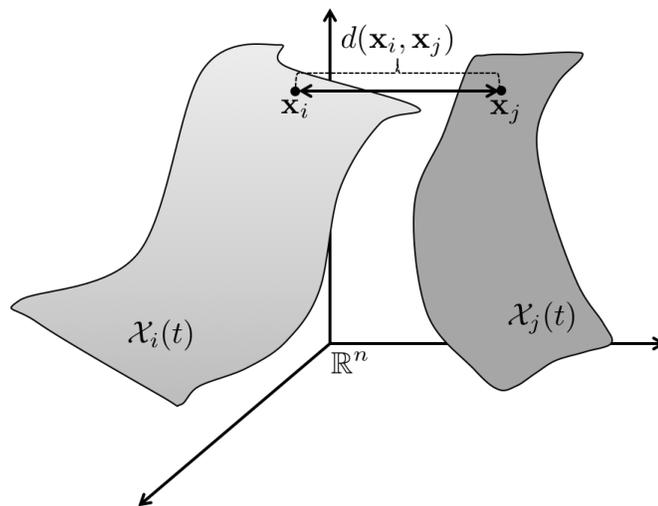


Figure 5.1: The distance metric d defined on $\mathcal{X}_i(t)$ and $\mathcal{X}_j(t)$

Regardless of the distance metric chosen, the general optimization problem is posed as

$$\begin{aligned} & \underset{\mathbf{x}_{u,i}, \mathbf{x}_{u,j}}{\text{minimize}} && d(\mathbf{x}_i(t), \mathbf{x}_j(t)) \\ & \text{subject to} && \mathbf{x}_i \in \mathcal{X}_i(t) \\ & && \mathbf{x}_j \in \mathcal{X}_j(t) \end{aligned}$$

which attempts to minimize the distance metric by changing the undetermined states of the i^{th} and j^{th} observations. This optimization problem directly yields the states in the set $\mathcal{O}_{ij}(t)$ which represent the potential hypothetical state solutions corresponding to the observations at a given time t . The optimization based approach posed by Siminski et. al is analogous to the generalized optimization problem defined above [59].

5.1.2 Simplified Optimization Approach for Set Intersection

The optimization problem defined in the previous section is of dimension $2u$ with $2n$ variables and $2d$ constraints. As such the optimizer must search for the optimum over an n -dimensional, non-convex space with $2u$ design variables. However, this problem can be simplified if the dimension of the design variables and the dimension of the state space over which the optimizer searches can be reduced, improving computational performance. The dimensionality of this the state space can be reduced by taking advantage of the ability to partition the state space. Every state $\mathbf{x}_i \in \mathcal{X}_i(t)$ may be partitioned into determined and undetermined states assuming knowledge of the observer dynamics. The set $\mathcal{X}_i(t)$ may also be partitioned into a set of determined states $\mathcal{X}_{d,i}(t)$ and a set of undetermined states $\mathcal{X}_{u,i}(t)$. A further simplification can be made by assuming that the common epoch time is either t_i or t_j , the time one of the observations is made. Assuming that the common epoch time is $t = t_j$, then $\mathcal{X}_j(t)$ does not need to be propagated in time. With this assumption, the entire set $\mathcal{X}_{u,j}(t)$ is simply given by the admissible region \mathcal{R}_j . More importantly, the set $\mathcal{X}_{d,j}(t)$ is simply defined as $\{\mathbf{x}_{d,j}\}$, a single point which is the set of determined states

that are obtained from the observation at time t_j . With this formulation of the problem, for every state $\mathbf{x}_i \in \mathcal{X}_i(t)$ such that $\mathbf{x}_{d,i}(t) = \mathbf{x}_{d,j}(t)$ and $\mathbf{x}_{u,i}(t) \in \mathcal{R}_j$ there exists a full state $\mathbf{x}(t)$ that lies in the intersection of $\mathcal{X}_i(t)$ and $\mathcal{X}_j(t)$. Note that since the determined states and the measurements have a one-to-one and invertible mapping, this form of the optimization problem is essentially posed in the measurement space. By these simplifications, the optimization problem may be reformulated as

$$\begin{aligned} & \underset{\mathbf{x}_{u,i} \in \mathcal{R}_i}{\text{minimize}} && \tilde{d}(\mathbf{x}_{d,i}(t), \mathbf{x}_{d,j}(t)) \\ & \text{subject to} && \mathbf{x}_{u,i} \in \mathcal{X}_{u,i}(t_j) \end{aligned}$$

where $\tilde{d}(\mathbf{a}, \mathbf{b}) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$ is the new distance metric and the only design variable is $\mathbf{x}_u(t_i)$. Again, this distance metric is general and could be chosen to be the Mahalanobis distance for example. Figure 2 illustrates how the distance metric is defined for the simplified optimization problem. Note that in this figure, the set $\mathcal{X}_{d,i}$ containing all the possible determined states, or equivalently the possible measurements, lie on a manifold of dimension \mathbb{R}^d . Choosing to use the determined states as the design space for the optimization problem

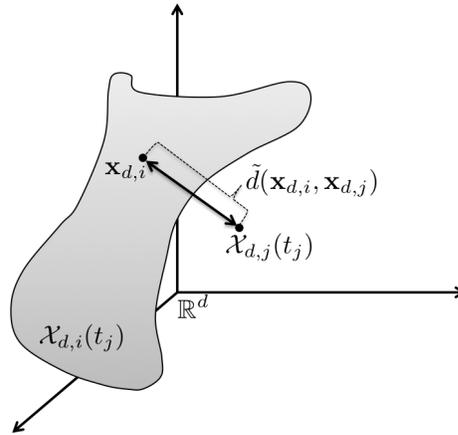


Figure 5.2: The distance metric defined on $\mathcal{X}_{d,i}(t)$ and $\mathcal{X}_{d,j}(t)$

yields a well defined, closed set over which the design variables, the undetermined states, must take values. This optimization problem attempts to then find the potential intersection

between the point $\mathcal{X}_{d,j}$ and the d dimensional manifold $\mathcal{X}_{d,i}$.

This simplified optimization approach minimizes the distance between the d -dimensional subspaces of $\mathcal{X}_{d,i}$ and $\mathcal{X}_{d,j}$. Note that since the dimension of the determined space is $d < n$, it is necessary that

$$Nd \geq n \quad (5.10)$$

where N is the number of observations. If Eqn. (5.10) is not satisfied, the optimization problem will be underdetermined. When Eqn. (5.10) is satisfied, it is important to note that N optimization problems must be solved to fully define $\mathcal{O}(t)$.

Let the solutions to the simplified optimization problem belong to the set

$$\mathcal{M}_i = \arg \min_{\mathbf{x}_{u,i} \in \mathcal{R}_i} \tilde{d}(\mathbf{x}_{d,i}(t_j), \mathbf{x}_{d,j}(t_j)) \quad (5.11)$$

Define the corresponding n -dimensional states in this set of solutions by

$$\mathcal{O}_{i \rightarrow j}(t_i) = \{\mathbf{g}(\mathbf{x}_u, \mathbf{x}_d, \mathbf{k}, t_i) : \mathbf{x}_u \in \mathcal{M}_i\} \quad (5.12)$$

This set represents the set of states at time t_i consistent with mapping $\mathcal{X}_{d,i}(t_i)$ from time t_i to time t_j . Because this optimization approach is underdetermined, that is it only finds an intersection of the sets $\mathcal{X}_{d,i}$ and $\mathcal{X}_{d,j}$, it is not sufficient to directly state that points in $\mathcal{O}_{i \rightarrow j}(t_i)$ are the potential state solutions. It is necessary to also find the solutions to the reverse optimization problem given by

$$\begin{aligned} & \underset{\mathbf{x}_{u,j} \in \mathcal{R}_j}{\text{minimize}} && \tilde{d}(\mathbf{x}_{d,j}, \mathbf{x}_{d,i}) \\ & \text{subject to} && \mathbf{x}_{u,j} \in \mathcal{X}_{u,j}(t_i) \end{aligned}$$

the solution to which yields

$$\mathcal{M}_j = \arg \min_{\mathbf{x}_{u,j}} \tilde{d}(\mathbf{x}_{d,i}(t_i), \mathbf{x}_{d,j}(t_j)) \quad (5.13)$$

$$\mathcal{O}_{j \rightarrow i}(t_j) = \{\mathbf{g}(\mathbf{x}_u, \mathbf{x}_d, \mathbf{k}, t_j) : \mathbf{x}_u \in \mathcal{M}_j\} \quad (5.14)$$

the set of states consistent with mapping $\mathcal{X}_{d,j}(t_j)$ to time t_i . Figure 5.3 illustrates the overall optimization problem following this simplified approach when $N = 2$. It is clear that for a

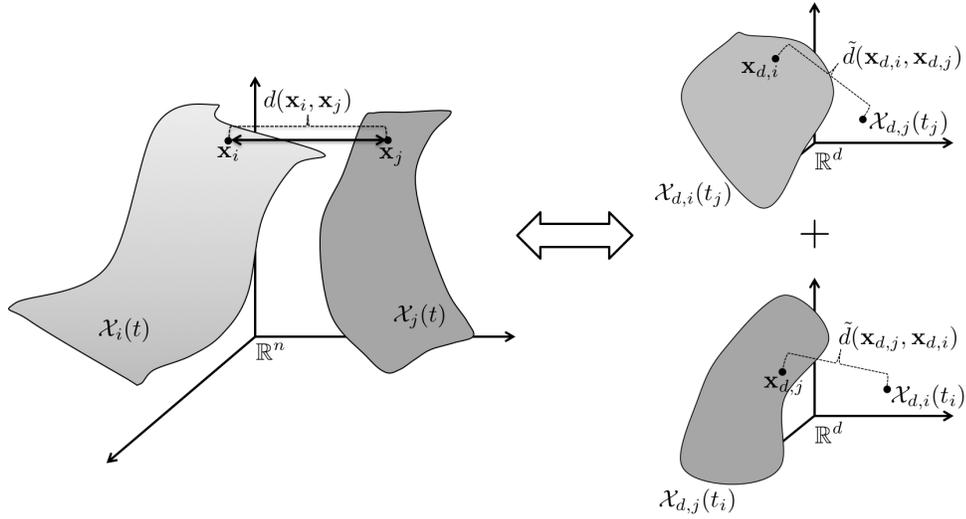


Figure 5.3: The simplified optimization problem requires the intersection to be found between pairwise sets of observations

potential state solution to exist, it must reside in $\mathcal{O}_{ij}(t) = \mathcal{O}_{j \rightarrow i}(t_i)|_t \cap \mathcal{O}_{i \rightarrow j}(t_j)|_t$, where $|_t$ indicates that each set is propagated to a common epoch time t . This generalizes for $N \geq 2$ as

$$\mathcal{O}_{ij}(t) = \bigcap_{i=1}^N \bigcap_{j=1, j \neq i}^N \mathcal{O}_{i \rightarrow j}(t_i)|_t \quad (5.15)$$

where each $\mathcal{O}_{i \rightarrow j}(t_i)$ is the set of solutions to the simplified optimization problem.

This condition ensures that the optimal solution when mapping from time t_i to time t_j is identical to the optimal solution when doing the reverse mapping for all pairwise sets of

N observations and can be equivalently expressed as

$$\begin{aligned} & \underset{\mathbf{x}_{u,i} \in \mathcal{R}_i}{\text{minimize}} && \sum_{i=1}^N \sum_{j=1, j \neq i}^N \tilde{d}(\mathbf{x}_{d,i}(t), \mathbf{x}_{d,j}(t)) \\ & \text{subject to} && \mathbf{x}_{u,i} \in \mathcal{X}_{u,i}(t_j) \end{aligned}$$

and is the foundation of the contributions of this work.

This optimization problem is similar to the boundary value approach presented by Siminski et. al. [59], but provides a more computationally efficient design space in which to optimize. Specifically, the optimization approach developed by Siminski et. al. is of dimension $u+n$ where the design variables are of dimension u and the objective space, which is essentially equivalent to a trajectory optimization problem, is of dimension n . Computationally, optimizing over an n dimensional space becomes more difficult as n increases. The proposed approach utilizes $N!$ optimization problems each with a u dimensional design variable in a m dimensional objective space. For $N = 2$ the dimensionality of both the proposed approach and the approach presented by Siminski et. al. are the same. However, it is more computationally tractable to perform the N required simplified optimization problems as it is easier to find all local minima in this non-convex, reduced dimension state space than in the full n dimensional, non-convex, objective space.

Note that while there is not commonly more than one point of intersection of the admissible region manifolds [61], it is possible that there exist several local minima found by the optimization algorithm. Figure 5.4 illustrates a notional case where there are two points such that

$$\tilde{d}_1 = \tilde{d}(\mathbf{x}_{d,1}(t_j), \mathbf{x}_{d,j}(t_j)) = \tilde{d}_2 = \tilde{d}(\mathbf{x}_{d,2}(t_j), \mathbf{x}_{d,j}(t_j)) \quad (5.16)$$

for $N = 2$. In this case, the optimization could present two local minima that could indicate the observations are associated. Without considering uncertainties in the problem,

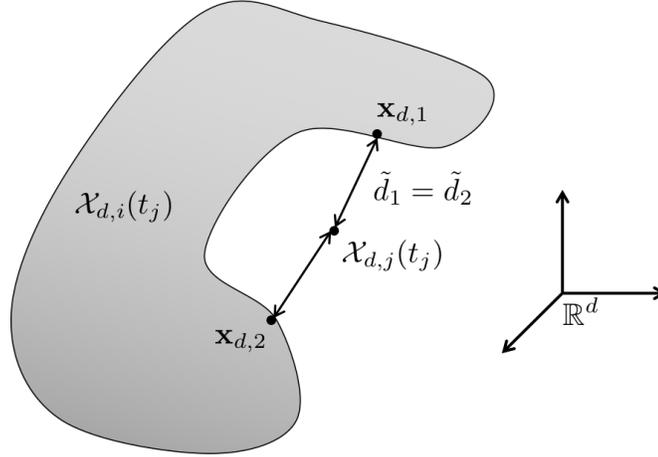


Figure 5.4: An notional example where there are two local minima.

the issue of local minima may be eliminated since if two observations are associated then there must exist at least one state such that the distance metric $\tilde{d}(\mathbf{x}_{d,j}(t), \mathbf{x}_{d,i}(t))$ is exactly zero. However, since there are uncertainties and errors in all observation systems, one is interested in the probability that the distance metric is zero, or equivalently, the probability that the two observations are associated. The case when uncertainties are considered in this optimization problem is discussed in the next sections.

5.1.3 Simplified Optimization Approach for General Phenomenologies

The simplified optimization approach described in Section 5.1.2 can be applied to various measurement phenomenologies. Figure 5.3 illustrates a case for optical observations where $d = 4$ and $n = 6$ and thus requiring $N = 2$ observations to find local minima which may indicate association. However this methodology is easily extensible to other observation methodologies, several of which are highlighted in this section. Radar observations enable accurate determination of range, range-rate, and line of sight however cannot accurately or reliably measure angular rates [64]. Again this is an example of an observation system with $d = 4$, and as such 2 radar observations are required in order to perform association via the simplified optimization approach. Magnetometer based observations have been recently proposed as a method for detecting objects in GEO [71]. These observations also provide

a set of 4 determined states which just as in the optical and radar cases require at least $N = 2$ observations to perform association. Signals of opportunity can give serendipitous measurements, specifically time difference of arrival (TDoA) measurements which provide pseudo-ranges and timing data which combined provide a position estimate for an object of interest [110]. Due to the lack of phase information, these observations do not provide any information on the velocity and as such these are the unobservable states giving $d = 3$ and again requiring $N = 2$ observations to perform the simplified optimization approach for association. The topology of the admissible region in this case is illustrated in [110] and lends to a well posed optimization problem even though the problem is not overdetermined.

The application of this simplified optimization approach for each of these measurement phenomenologies retains the computational attractiveness of performing 2 reduced order optimization problems, however there are more difficult cases in which d is of a lower dimension requiring significantly more optimization problems to be solved. For example, given a spurious observation of an object in a radar ranging system in which only the range and range-rate information is available the dimension of the determined states is $d = 2$ which implies there must be at least $N = 3$ such measurements to utilize this method for association. This then requires 6 optimization problems to be solved. If additionally the range-rate information is unavailable, then $d = 1$ and at least $N = 6$ such measurements are necessary for this association method resulting in 720 necessary optimization problems to be solved. While this method certainly generalizes to these phenomenologies, it is worth mentioning that the computational tractability of this method suffers when the number of observable states is sufficiently low. Furthermore, the topology of the optimization problem could pose problems in these cases since with only one or two observable states, it may be difficult to ensure that there exist unique points of minimum distance solutions. Thus, if the dimension of the determined states is sufficiently low such that the computational benefits offered from the reduced order optimization are outweighed by the total number of optimization problems that must be solved then it is an indication that a traditional particle

based, full dimension association method should be used.

5.1.4 Hypothesis Test for Observation Association

The optimal value of the distance metric defined in the optimization approach can only truly be zero in the case of perfect sensors and perfect knowledge of the observer dynamics. In general, there are uncertainties associated with the observation and errors associated with the observer state in addition to timing errors. Each of these effects must be taken into account when attempting to identify if an intersection exists between two observations. Figure 5.5 shows a general representation of the optimization problem considering uncertainty. With uncertainties, $\mathbf{X}_{d,i}$ and $\mathbf{X}_{d,j}$ are the random variables representing the dis-

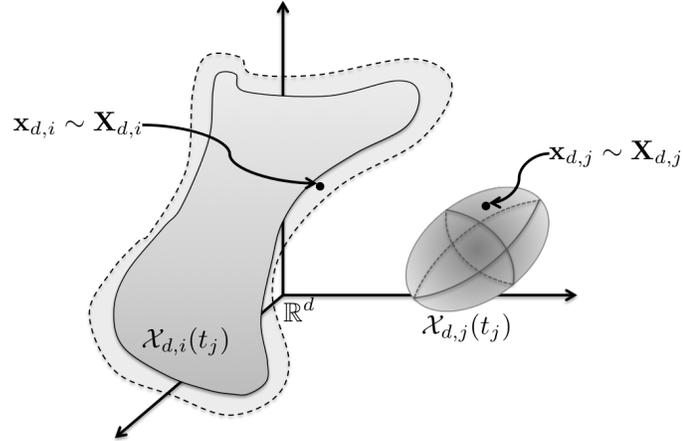


Figure 5.5: A representation of the sets $\mathcal{X}_{d,i}(t_j)$ and $\mathcal{X}_{d,j}(t_j)$ including uncertainties and errors.

tributions. These random variables, assuming Gaussian uncertainties, now imply the sets $\mathcal{X}_{d,i}(t)$ and $\mathcal{X}_{d,j}(t)$ are probabilistic

$$\mathcal{X}_{d,i}(t_i) = \{\mathbf{X}_{d,i}(t_i) : \mathbf{X}_{d,i}(t_i) \sim \mathcal{N}(\mathbf{x}_{d,i}(t_i), \Sigma_i), \mathbf{X}_{u,i}(t_i) \sim \mathcal{R}_i\} \quad (5.17)$$

$$\mathcal{X}_{d,j}(t_j) = \{\mathbf{X}_{d,j}(t_j) : \mathbf{X}_{d,j}(t_j) \sim \mathcal{N}(\mathbf{x}_{d,j}(t_j), \Sigma_j), \mathbf{X}_{u,j}(t_j) \sim \mathcal{R}_j\} \quad (5.18)$$

where \mathbf{x}_u is sampled from the i^{th} probabilistic admissible region generated by Eqn. (2.54) and $\Sigma_i \in \mathbb{R}^{d \times d}$ is the covariance matrix for the i^{th} observation. Following the simplified

optimization approach, the distribution about the single point $\mathbf{x}_{d,j} \in \mathcal{X}_{d,j}(t_j)$ is now an ellipsoid in \mathbb{R}^d centered at $\mathbf{x}_{d,j}$. The distance metric of the simplified optimization approach also becomes a random variable defined as

$$D = \tilde{d}(\mathbf{X}_{d,i}(t), \mathbf{X}_{d,j}(t)) \quad (5.19)$$

As shown in the previous section, if an intersection exists, or equivalently if the observations are associated, then there should exist a state $\mathbf{x}_{d,i}(t)$ such that Eqn. (5.19) equals zero. Let $l_i = \text{card}(\mathcal{M}_i)$ and $l_j = \text{card}(\mathcal{M}_j)$ and define

$$D_{i \rightarrow j}(\mathbf{x}_u(t_i)) = \tilde{d}(\mathbf{X}_{d,i}(t_j), \mathbf{X}_{d,j}(t_j)) \quad (5.20)$$

as the random variable of the distance metric mapping a state $\mathbf{x}_u \in \mathcal{M}_i$ from time t_i to time t_j where $\mathbf{x}_d(t_i) \sim \mathcal{N}(\mathbf{x}_{d,i}(t_i), \Sigma_i)$. Likewise, $D_{j \rightarrow i}(\mathbf{x}_u(t_j))$ is the random variable of the distance metric mapping a state $\mathbf{x}_u \in \mathcal{M}_j$ from time t_j to time t_i . Since D is now a random variable, it is desired to understand what distributions $D_{j \rightarrow i}(\mathbf{x}_u(t_j))$ and $D_{i \rightarrow j}(\mathbf{x}_u(t_j))$ are drawn from to determine if there is association. Let $f_0(s)$ be the distribution defined by imposing that the location of the intersection is at the location of a local minima. Equivalently,

$$f_0(s) = \{D_0 : d_0 = \tilde{d}(\mathbf{x}_{d,i}(t), \mathbf{X}_{d,j}(t) + \mathbf{c}), \mathbf{X}_{d,j}(t) \sim \mathcal{X}_{d,j}(t)\} \quad (5.21)$$

where $\mathbf{x}_{d,i}$ is a fixed point in $\mathcal{X}_{d,i}$ corresponding to a local minima and \mathbf{c} shifts $\mathcal{X}_{d,j}$ to be centered at $\mathbf{x}_{d,i}$ and $\mathbf{x}_u \in \mathcal{M}_i$. Figure 5.6 illustrates how this distribution is created by shifting the location of $\mathcal{X}_{d,j}$ by \mathbf{c} to the location of the local minima. This distribution represents the distribution that D is drawn from if this local minima is indeed a point of

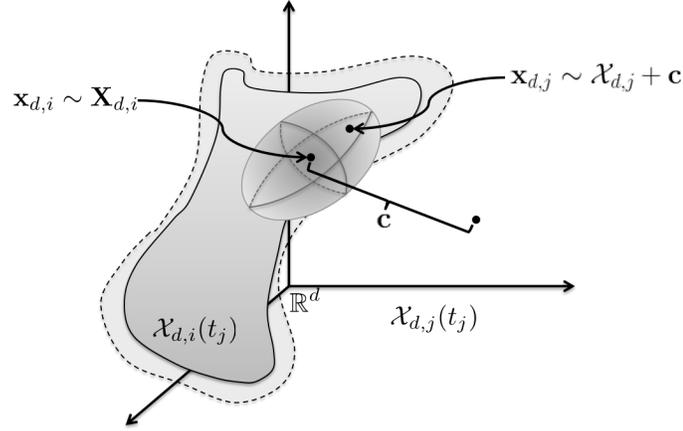


Figure 5.6: A representation of the how f_0 is constructed.

intersection. The actual distribution of D can then be represented by

$$f_1(s) = \{D_0 : d_0 = \tilde{d}(\mathbf{x}_{d,i}(t), \mathbf{X}_{d,j}(t) + \mathbf{c}), \mathbf{X}_{d,j}(t) \sim \mathcal{X}_{d,j}(t)\} + d_{ij, \mathbf{x}_u(t_i)} \quad (5.22)$$

$$= f_0(s) + d_{ij, \mathbf{x}_u(t_i)} \quad (5.23)$$

where since \mathbf{c} is a constant value, $d_{ij, \mathbf{x}_u(t_i)}$ represents a constant which shifts the location of distribution $f_0(s)$. Thus, $d_{ij, \mathbf{x}_u(t_i)}$ is only non-zero if the local minima is not a point of intersection. The scenario described to construct f_1 is illustrated by Figure 5.5.

Given these two distributions, it is possible to define a test to determine which distribution D is drawn from. By Eqn. (5.23) and Eqn. (5.10), for a set of N observations to be associated it is necessary that for at least one pair of points $\mathbf{x}_u \in \mathcal{M}_i$ and $\mathbf{x}_u \in \mathcal{M}_j$,

$$\left[\sum_{i=1}^N \sum_{j=1, j \neq i}^N d_{ij, \mathbf{x}_u(t_i)} \right] = 0 \quad (5.24)$$

Eqn. (5.24) takes into account the requirement that every pair of the d dimensional surfaces that can be formed from the N observations must all have an intersection. This results in

the following hypothesis tested to determine if Eqn. (5.24) is satisfied.

$$\mathcal{H}_0 : D_{i \rightarrow j}(\mathbf{x}_u(t_j)) \sim f_0(s) + d_{ij, \mathbf{x}_u(t_i)} \quad i, j = 1, \dots, N; i \neq j \quad (5.25)$$

$$\mathcal{H}_1 : D_{i \rightarrow j}(\mathbf{x}_u(t_j)) \sim f_0(s) + 0 \quad i, j = 1, \dots, N; i \neq j \quad (5.26)$$

The alternative hypothesis \mathcal{H}_1 assumes that all N observations are associated, and the null hypothesis \mathcal{H}_0 assumes that there is no association of one or more of the N observations. These hypotheses can be tested at all $l_i \times l_j$ pairwise combinations of solutions in \mathcal{M}_i and \mathcal{M}_j independently.

The binary hypothesis test is constructed by first determining the probability density functions associated with the null and alternative hypotheses, $f_0(s)$ and $f_1(s)$ respectively, which are defined by Eqns. (5.21) and (5.23) [111]. The Type I error of the binary hypothesis test, the probability of false association P_{FA} in this application, is determined by

$$P_{FA} = \int_c^\infty f_0(s) ds \quad (5.27)$$

The value of c is selected specifically to obtain a desired P_{FA} . The probability of association is then given by

$$P_A = \int_c^\infty f_1(s) ds \quad (5.28)$$

Additionally, the probability of a Type II error, the probability of missed association P_{MA} in this application, is the complement of P_A . While larger values of c reduce the probability of false association, it also reduces the probability of a correct association. Figure 5.7 illustrates the binary hypothesis testing setup. Since the hypotheses are tested at each pairwise combination of solutions, the result from this set of hypothesis tests is a $l_i \times l_j$ set of values of P_A .

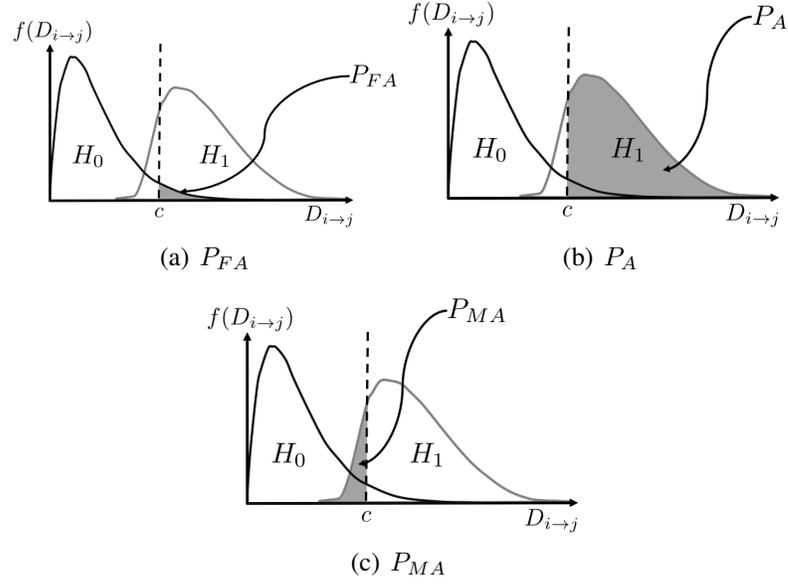


Figure 5.7: Binary Hypothesis Testing Setup

5.1.5 Solution at the Intersection of Admissible Regions

It is stated in the preceding optimization approach to associating admissible regions that there is not a general case in which there will be more than one local point of intersection of the two sets. It is clear that if $\mathcal{O}(t_j)$ consists of a single point \mathbf{x}_s , then not only are the observations yielding $\mathcal{X}_{d,i}$ and $\mathcal{X}_{d,j}$ associated, but also a single solution exists. Conceptually, this implies that the measurements taken at time t_i and t_j could only have been observed if the object's true state is \mathbf{x}_s at time t_j . Now consider a notional case where two intersections exist, the representative geometry of which is shown in Figure 5.8. In this notional case, there exist two unique states $\{\mathbf{x}_{u,1}(t_i), \mathbf{x}_{u,2}(t_i)\} \in \mathcal{R}_i$ that when mapped forwards to time t_i , produce exactly the same set of determined states $\mathbf{x}_d(t_j)$. This illustration is consistent with the multirevolution problem [61], wherein over long time periods, an observation may be consistent with more than one state due to ambiguity in the number of orbits completed between the observations.

In addition, when taking into account the systemic uncertainties in the system, there is the potential for spurious local minima to occur. Consider the illustration shown in Figure

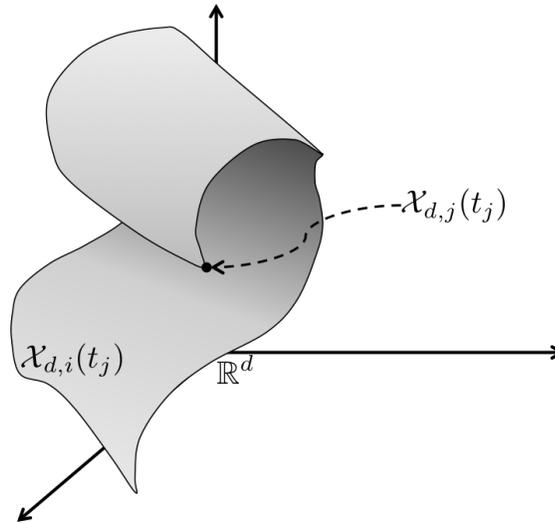


Figure 5.8: Notional case of two or more intersections of the sets $\mathcal{X}_{d,i}(t_j)$ and $\mathcal{X}_{d,j}(t_j)$

5.9 where each of the sets are now distributed probabilistically in \mathbb{R}^d . The uncertainties

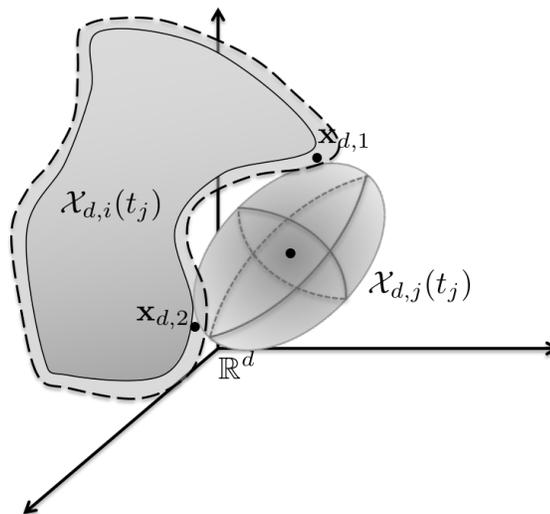


Figure 5.9: Ambiguous case where uncertainties creates additional potential points of intersection

about the sets $\mathcal{X}_{d,i}(t_j)$ and $\mathcal{X}_{d,j}(t_j)$ creates the potential for there to be more than one local minima found by the optimizer. Note that it is important for the optimizer to find all such local minima as it is not known a priori which minima is the truth solution. The question posed by this scenario, as well as the scenario from Figure 5.8, is how to handle the cases where several solutions are local minima and will be found to have a significant individual

value of P_A .

In general, while the solution at these optimal values are possibly the state of the object, the existence of a single, or multiple, intersections is fundamentally an indication that the two observations are associated. A positive association signifies that a traditional Bayesian update method can be performed to obtain the PDF over the object orbit estimate from ingesting the newly associated measurement.

Thus, the probability of association metric should simply denote when a positive association has occurred. As stated in the previous section, the probability of association computed for each pair of local minima yields a set of $l_i \times l_j$ values of P_A . Let these values of P_A be arranged into a $l_i \times l_j$ matrix and denote this matrix as \mathbf{P}_A . Let \mathcal{P}_A then represent the overall probability of association metric for the observations from time t_i and time t_j . Each element of \mathbf{P}_A gives the individual probability of association for a given pair of solutions for \mathbf{x}_u .

Note that each of these pairs of solutions are mutually exclusive, and as such the total probability that a particular pair is correct must sum to one. For this problem, the event probabilities can be assumed to be found by normalizing the individual probability of association values by the sum the values as follows

$$\mathbb{P}[(\mathbf{x}_{u,i}(t_i), \mathbf{x}_{u,j}(t_j))] = \frac{P_{A,(ij)}}{\sum_{i=1}^{l_i} \sum_{j=1}^{l_j} P_{A,(i,j)}} \quad (5.29)$$

The overall probability of association can then be found by taking the weighted norm of the probabilities of association values as follows

$$\mathcal{P}_A = \sum_{i=1}^{l_i} \sum_{j=1}^{l_j} \mathbb{P}[(\mathbf{x}_{u,i}(t_i), \mathbf{x}_{u,j}(t_j))] P_{A,(i,j)} \quad (5.30)$$

which gives a single metric quantifying the total probability that the observation taken at t_i is associated with the observation taken at time t_j .

5.1.6 Approximation of the Bayesian Update using Gaussian Mixtures

Ultimately, the association of a set of measurements is used along with Bayes rule to generate a full state estimate. Particle filtering or multiple hypothesis filtering methods are often used to incorporate a Bayesian update for initial orbit determination [62] [112]. These approaches can be computationally demanding due to the number of particles necessary to generate a good approximation of the PDF. While the particle filter approaches the exact PDF as the number of particles goes to infinity, it can be shown that a Gaussian Mixture approximation of this PDF can be derived from the optimization problem defined in this paper. Consider the distance metric defined in the previous section and the requirement given by Eqn (5.24). One may define

$$L(\mathbf{x}_{u,i} | \mathbf{x}_{d,j}, t) = \exp \left(- \sum_{i=1}^N \sum_{j=1, j \neq i}^N \tilde{d}(\mathbf{x}_{d,j}(t), \mathbf{x}_{d,i}(t)) \right) \quad (5.31)$$

as the general likelihood that the undetermined state is given by $\mathbf{x}_{u,i}$ at time t given a set of true measurements given by $\mathbf{x}_{d,j}, j = 1, \dots, N$. Note that Eqn. (5.31) has a form similar to the form of the Bayesian update used in particle filtering which enables this approximation [113]. The maximum likelihood estimate (MLE) for $\mathbf{x}_{u,i}$ is given as follows. [78]

$$\arg \max_{\mathbf{x}_{u,i} \in \mathcal{R}_i} L(\mathbf{x}_{u,i} | \mathbf{x}_{d,j}, t) \quad (5.32)$$

It is often more convenient to rewrite the natural log of Eqn. (5.32) as follows,

$$\arg \max_{\mathbf{x}_{u,j} \in \mathcal{R}_i} \ln [(L(\mathbf{x}_{u,i} | \mathbf{x}_{d,j}, t))] \quad (5.33)$$

which, for the likelihood defined by Eqn. (5.31), simply yields

$$\arg \max_{\mathbf{x}_{u,j}} - \sum_{i=1}^N \sum_{j=1, j \neq i}^N \tilde{d}(\mathbf{x}_{d,j}(t), \mathbf{x}_{d,i}(t)) \quad (5.34)$$

which is identical to optimization problem presented in the previous section, only posed as a maximization problem rather than a minimization. Thus, the set of solutions to this optimization problem, are equivalently derived from the set of MLEs for $\mathbf{x}_{u,i}$. Let $\text{card}(\mathcal{M}_i(t_i)) = \ell_i$ and let $\hat{\mathbf{x}}_{u,k}$ denote the k^{th} MLE for $\mathbf{x}_{u,i}$. As defined $\hat{\mathbf{x}}_{u,k}(t_i)$ is the $k^{\text{th}}(t_i)$ solution residing in $\mathcal{M}_i(t_i)$, the set of solutions to the optimization problem. Each $\hat{\mathbf{x}}_{u,k}$ satisfies the following

$$\mathbb{E} \left[\sum_{i=1}^N \sum_{j=1, j \neq i}^N \frac{\partial}{\partial \hat{\mathbf{x}}_{u,k}} \left(-\tilde{d}(\mathbf{x}_{d,j}(t), \mathbf{x}_{d,i}(t)) \right) \right] = \mathbf{0}_{u \times 1} \quad (5.35)$$

$$\mathbb{E} \left[\sum_{i=1}^N \sum_{j=1, j \neq i}^N \frac{\partial^2}{\partial \hat{\mathbf{x}}_{u,k}^2} \left(\tilde{d}(\mathbf{x}_{d,j}(t), \mathbf{x}_{d,i}(t)) \right) \right] = \mathcal{I}(\hat{\mathbf{x}}_{u,k})_{u \times u} \quad (5.36)$$

where Eqn. (5.35) follows directly from each $\hat{\mathbf{x}}_{u,k}$ being a MLE, since the derivative of the log-likelihood is zero at the maximum, and $\mathcal{I}(\hat{\mathbf{x}}_{u,k})$ is the Fisher information matrix [114]. Defining $\mathbf{P}_{u,k}$ as the covariance matrix associated with $\hat{\mathbf{x}}_{u,k}$, then by the Cramer-Rao bound

$$\mathbf{P}_{u,k} \geq \mathcal{I}(\hat{\mathbf{x}}_{u,k})^{-1} \quad (5.37)$$

[115]. Assume this bound is attained by the limiting properties of the MLE [116], then the statistics about the estimate is approximated by a Gaussian such that $\mathbf{x}_{u,i} \sim \mathcal{N}(\hat{\mathbf{x}}_{u,k}, \mathbf{P}_{u,k})$. Further, given that l_i is finite then the k^{th} full *a priori* estimate $\hat{\mathbf{x}}_k$ at time t_i is approximated

by

$$\hat{\mathbf{x}}_k(t_i) = \begin{bmatrix} \hat{\mathbf{x}}_{u,k} \\ \mathbf{x}_{d,i} \end{bmatrix} \quad (5.38)$$

$$\mathbf{P}_k(t_i) = \begin{bmatrix} \mathbf{P}_{u,k} & \mathbf{P}_{ud,k} \\ \mathbf{P}_{ud,k}^T & \mathbf{R}_i \end{bmatrix} \quad (5.39)$$

$$= \begin{bmatrix} \mathbf{P}_{u,k} & \mathbf{0}_{u \times d} \\ \mathbf{0}_{d \times u} & \mathbf{R}_i \end{bmatrix} \quad (5.40)$$

for each $\hat{\mathbf{x}}_k \in \mathcal{O}(t_i)$, where \mathbf{R}_i is again the determined state (or measurement) covariance matrix and $\mathbf{P}_{u \times d,k}$ is the cross covariance term. Given that the state space is partitioned and by definition the determined states \mathbf{x}_d are independent of the undetermined states \mathbf{x}_u (or equivalently the undetermined states are unobservable), it is clear that they are uncorrelated and $\mathbf{P}_{u \times d,k} = 0$ and the resulting covariance matrix for a given estimate \mathbf{x}_k has the block diagonal shown.

In the simplest case, $l_i = 1$, and $\hat{\mathbf{x}}(t_i)$ and $\mathbf{P}(t_i)$ represent the approximate distribution about the estimated state solution at time t_i . This process can likewise be used to approximate the PDF of the state estimate at time t_j or a traditional Kalman filter may be instantiated from $\hat{\mathbf{x}}(t_i)$ and $\mathbf{P}(t_i)$ to obtain the PDF of the state estimate at time t_j . If $l_i > 0$ then a Gaussian mixture model can be used to represent the distribution by

$$f(\mathbf{x}(t_i)) = \sum_{i=1}^{l_i} w_i \mathcal{N}(\hat{\mathbf{x}}_k(t_i), \mathbf{P}_k(t_i)) \quad (5.41)$$

and the weights w_i are given by

$$w_i = \sum_{j=1}^{l_j} \mathbb{P}[(\mathbf{x}_{u,i}(t_i), \mathbf{x}_{u,j}(t_j))] \quad (5.42)$$

where w_i represents the total event probability for $\mathbf{x}_{u,i}(t_i)$. The resulting Gaussian mixture

can then be used to instantiate a Gaussian mixture Kalman filter, for example, to obtain the PDF of the state estimate at time t_j . Likewise, an unscented particle filter could be instantiated from the Gaussian mixture given by Eqn. (5.41). The implementation of this MLE approach in addition to the optimization based approach to identify the probability of association combines the association and estimation processes into one step, enabling a significant reduction in computation time. However, there is an assumption of observability in the application of this approximation, and this limitation is discussed in the next section.

5.1.7 The Observability Condition

A limiting case of the proposed approximation of the Bayesian update is when the time between observations is very small. The approximate condition for near unobservability for Keplerian dynamics

$$(t_j - t_i) \ll \sqrt{\frac{\|\mathbf{r}\|^3}{3\mu}} \quad (5.43)$$

where $\mathbf{r} \in \mathbb{R}^3$ is the position of the satellite from the center of the central body and μ is the gravitational parameter of the central body [117]. Unobservability is an issue in the proposed optimization problem because the sets $\mathcal{X}_{d,i}$ and $\mathcal{X}_{d,j}$ are nearly tangential in this case. While the problem is unobservable the objective space for the optimization problem is then approximately constant with small values of $D_{i \rightarrow j}$, and it is unlikely for a well-defined local minimum to be found, especially when using numerical methods, and as such the assumption that the local minimum solution is the MLE breaks down. The question is when is the Gaussian mixture approximation appropriate for a given problem.

The computation of the Fisher information matrix (or equivalently the Hessian of the optimization problem) gives an indication as to how well posed is the optimization problem. If any of the eigenvalues of $\mathcal{I}_k(\cdot)$ are negative or zero then it is an indication that in the region of a given $\hat{\mathbf{x}}_u$, the optimization problem is not convex and thus this estimate is not an

MLE. If both eigenvalues are positive, then the condition number of the Fisher information, $\text{cond}(\mathcal{I}_k(\cdot))$, can indicate if the eigenvalues are sufficiently positive for optimality, and thus for the assumption that the estimate is a MLE to hold. Since $\mathcal{I}_k(\cdot)$ is a symmetric matrix, the condition number is given by

$$\text{cond}(\mathcal{I}_k(\cdot)) = \frac{\lambda_{\max}(\mathcal{I}_k(\cdot))}{\lambda_{\min}(\mathcal{I}_k(\cdot))} \quad (5.44)$$

which is the ratio of the maximum and minimum eigenvalues. If the condition number of this matrix is large for all local minima, it is equivalent to stating that the unobservability condition provided in Eqn. (5.43) is met. Thus the condition under which a Gaussian approximation is appropriate is given from rearranging Eqn. (5.43) as follows

$$\text{cond}(\mathcal{I}_k(\cdot)) < (t_j - t_i)^{-1} \sqrt{\frac{\|\mathbf{r}\|^3}{3\mu}}, \quad (5.45)$$

for each solution k . Under a limiting case when Eqn. (5.45) is not satisfied for each k , it would be necessary to use traditional methods, such as a particle filter, to perform a Bayesian update. However, the proposed general optimization based approach itself holds even for this limiting case since the probability of association can still be determined according to §5.1.4.

5.2 Simulation Results

The optimization based approach described in the previous sections is applied to the following simulated observations to demonstrate the effectiveness of the association method and to highlight key contributions of uncertainties in the observation systems. The observations are simulated over an object in low Earth orbit (LEO) with orbital elements randomly selected. The observer for the simulations is located in Atlanta, GA and is assumed to be operating in a fixed pointing mode so that simulated objects ‘streak’ across the simulate

image. For the purposes of this simulation, the field of view of the telescope is set to be $3^\circ \times 3^\circ$. This simulation environment is propagated forwards in time and as objects pass through the simulated field of view of the observer, the right ascension, α , and declination, δ , angles at the beginning and endpoint of the simulated streak are recorded. Using the Lagrange interpolation approach from [76], each simulated observation yields

$$\mathbf{z}_i = \left[\alpha_1 \quad \alpha_2 \quad \delta_1 \quad \delta_2 \quad t_1 \quad t_2 \quad \mathbf{o} \quad \dot{\mathbf{o}} \right]^T \quad (5.46)$$

where the subscripts 1 and 2 indicate the start and end point of the streak and $\mathbf{o} \in \mathbb{R}^3$ and $\dot{\mathbf{o}} \in \mathbb{R}^3$ are the observer position and velocity respectively. Each of these parameters has some associated uncertainty which will be taken into consideration when computing the probabilistic admissible region and during the hypothesis testing. The uncertainties used during the following simulations, unless otherwise stated, are defined in Table 5.1 using the GT-SORT as a baseline.

Table 5.1: Measurement Error and Parameter Uncertainty

Right Ascension uncertainty, σ_α	0.63 arcseconds
Declination uncertainty, σ_δ	0.57 arcseconds
Timing error, σ_t	0.0052 s
Position error (each axis), $\sigma_{\mathbf{o}}$	1 m
Velocity error (each axis), $\sigma_{\dot{\mathbf{o}}}$	1 cm/s

For optical observations,

$$\mathbf{x}_d = \left[\alpha \quad \dot{\alpha} \quad \delta \quad \dot{\delta} \right]^T \quad (5.47)$$

and for every observation set \mathbf{z}_i a probabilistic admissible region can be constructed over

$$\mathbf{x}_u = \left[\rho \quad \dot{\rho} \right] \quad (5.48)$$

given the uncertainties in Table 5.1 using the approach outlined in [76]. Eqn. (2.54) is used

to generate the joint probabilistic admissible region for the energy constraint identified in [76] and the periapse radius constraint. Each such admissible region generated from a given observation may be used with the optimization approach defined in this paper.

The distance metric in this section is defined to be the square of the Mahalanobis distance defined as

$$\tilde{d}(\mathbf{x}_{d,i}(t_j), \mathbf{x}_{d,j}(t_j)) = d_M(\mathbf{x}_{d,i}(t_j), \mathbf{x}_{d,j}(t_j))^2 \quad (5.49)$$

$$= (\mathbf{x}_{d,i}(t_j) - \mathbf{x}_{d,j}(t_j))^T \mathbf{R}^{-1} (\mathbf{x}_{d,i}(t_j) - \mathbf{x}_{d,j}(t_j)) \quad (5.50)$$

where \mathbf{R} is the measurement error covariance matrix. The hypothesis testing is set up by first running the optimizer over the sets $\mathcal{X}_{d,i}(t_i)$ and $\mathcal{X}_{d,j}(t_j)$ to find the local minima for the forwards and backwards mapping. Then, for each pairwise set of solutions, 500 values of $\mathbf{x}_{d,i}(t)$ and $\mathbf{x}_{d,j}(t)$ are sampled to generate a distribution of $D_{i \rightarrow j}$. The PDF for the alternative hypothesis is generated by these values of $D_{i \rightarrow j}$. Since the null hypothesis assumes the intersection exists, the PDF for the null hypothesis is generated by centering the distribution at the local minima solution. If the observations are associated, then this implies Eqn. (5.24) should hold. The binary hypothesis test is then performed using $P_{FA} = 5\%$. To achieve this, for every set of hypotheses constructed from Eqns. (5.25) and (5.26), the value of c is iterated until Eqn. (5.27) gives $P_{FA} = 5\%$. The probability of association threshold for these examples is set to $P_{\min} = 0.50$, which requires the largest probability of association to exceed this value in order for a Bayesian update to be performed on the newly associated measurement. The Bayesian update is performed using a particle filter by sampling \mathcal{R}_1 with 50000 points and ingesting the associated measurement at time t_2 and also by approximation using the method proposed in this work. Several test cases have been selected from the simulation described above to demonstrate the results of attempting to perform association with this method.

The details of the GT-SORT sensor used for the results shown in Cases 3 and 4 are

provided as follows. The telescope itself is a 0.5 meter, f/6 Raven class optical telescope mounted on a Paramount ME II German equatorial mount¹. The imaging sensor is the Grasshopper3², a 6MP, monochrome, CMOS electro-optical sensor. This configuration provides GT-SORT with a 13.3×10.6 arcminute field of view and an instantaneous field of view of .3121 arcseconds. No filters were applied when the data was collected for the results shown in this section.

5.2.1 Case 1: Single Solution

For the single solution case, the true state of the object at the beginning of the simulation is given in Table 5.4. The simulated initial observation for this case is initiated at time $t_1 =$

Table 5.2: True Orbit for Case 1

Semi-major axis, a	7034 km
Eccentricity, e	.001
Inclination, i	88°
RAAN, Ω	19°
Argument of perigee, ω	-134.6°
True Anomaly, ν	158.0°

02:00:00 UTC 01 Mar 2016 and the second observation is initiated at $t_2 = 02:10:41$ UTC 01 Mar 2016. For this observation, the admissible region probability of set membership and distance metric contours are shown in Figure 5.10. As can be seen, there is only one solution found through the optimization approach both in the forwards and reverse mapping. To convey the results of the actual optimization algorithm, Table 5.4 displays the probability of set association for this case. Since there is only one solution found, there are no other pairwise solutions to consider and $\mathcal{P}_A = 0.932$, indicating strong association between the two observations. Table 5.4 also displays the location of these solutions within the admissible regions at time t_1 and time t_2 . The determined states for both measurements

¹<http://www.bisque.com/sc/pages/ParamountMEII.aspx> last accessed on 12/12/2016

²<https://www.ptgrey.com/grasshopper3-usb3-vision-cameras> last accessed on 12/12/2016

are given below

$$\mathbf{x}_d(t_1) = \begin{bmatrix} 46.5^\circ & -36.3^\circ & 0.032^\circ/\text{s} & 0.075^\circ/\text{s} \end{bmatrix} \quad (5.51)$$

$$\mathbf{x}_d(t_2) = \begin{bmatrix} 171.8^\circ & 42.6^\circ & 0.089^\circ/\text{s} & -0.045^\circ/\text{s} \end{bmatrix} \quad (5.52)$$

where $\mathbf{x}_d = [\alpha, \delta, \dot{\alpha}, \dot{\delta}]$. Combining the solutions in Table 5.4 with the determined states above yields the full state estimate for the object at either time t_1 or t_2 .

For this case, a particle filter with 50,000 particles is used to approximate $p(\mathbf{x}_u(t_1))$, the numerical *a posteriori* pdf after ingesting the newly associated measurement. Figure 5.11b shows the updated PDF along with the location of the estimate at t_1 given in Table 5.4 as estimated by the particle filter. This PDF can also be approximated as a Gaussian about the MLE as described in Section 5.1.6. Figure 5.11 shows the comparison of the particle filter results with the MLE approximation described in Section 5.1.6. The approximation of the distribution is in good agreement with the numerically computed distribution of the particle filter. In particular, for Case 1 since the time between the observations is relatively short, the errors in the approximation method due to linearization are minimal and there is good agreement with the particle filter distribution. The estimated orbit solution for the MLE approximation is given by Table 5.3. However, the runtime for the particle filter is

Table 5.3: Estimated Orbit for Case 1

	Estimate	Cov. (3σ)	Error
Semi-major axis, a	7033.3 km	3.70 km	0.70 km
Eccentricity, e	.0011	4×10^{-4}	1×10^{-4}
Inclination, i	88.0°	0.006°	$4 \times 10^{-4^\circ}$
RAAN, Ω	19.0°	0.009°	$4 \times 10^{-4^\circ}$
Argument of perigee, ω	-134.7°	10.82°	0.1°
True Anomaly, ν	158.27°	10.89°	0.27°

on the order of several minutes whereas the runtime for the approximation is on the order of 8 seconds. This reduction in computation time is benefited further by the necessity of computing the probability of association even in the particle filter case. Since this approxi-

mation combines the association and estimation steps, it costs very little extra computation to produce the estimated PDF. For the next two examples, the *a posteriori* PDFs shown will also be generated from the MLE approximation. Note that in each of the figures comparing the MLE and particle filter approximations, the plots are centered about the MLE estimate and are plotted with equal axes to best compare the size and shape of the estimated distributions.

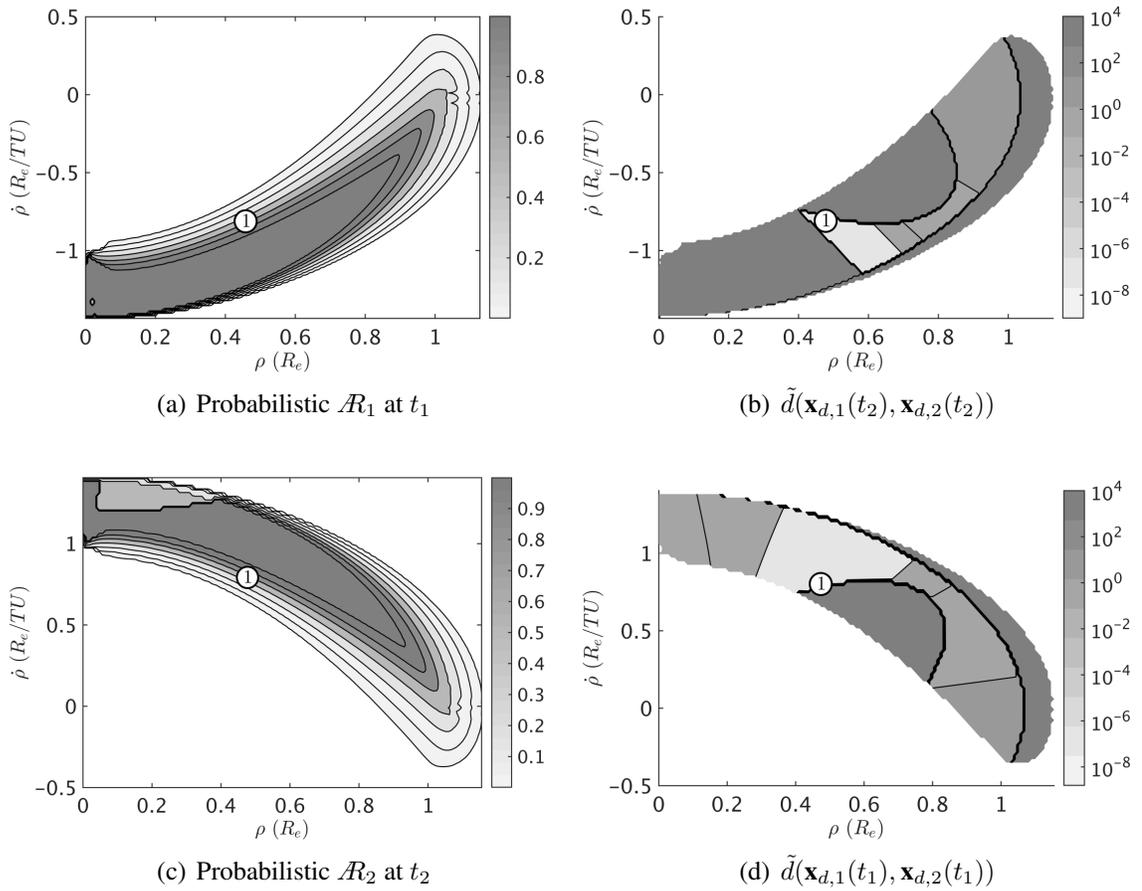


Figure 5.10: The contours of the probability of set membership and distance metrics for Case 1 with solutions labeled at time t_1 and t_2 .

Table 5.4: Case 1 Probability of association and optimization solutions

Solution	1
P_A	0.932
P_{FA}	0.05
$\rho(t_1) (R_e)$	0.4045
$\dot{\rho}(t_1) (R_e/TU)$	-0.7563
$\rho(t_2) (R_e)$	0.4067
$\dot{\rho}(t_2) (R_e/TU)$	0.7542

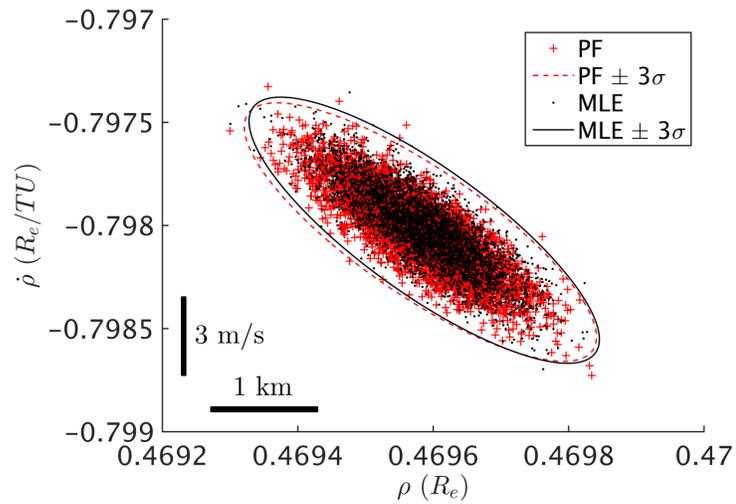


Figure 5.11: The numerical PDF after ingesting the newly associated observation and points sampled from the approximated PDF generated with the MLE approximation.

5.2.2 Case 2: Ambiguous Solutions

For this case, the time between the observations is selected to be large. The true state of the object at the beginning of the simulation is identical to Case 1 with a true anomaly now of $\nu = -156.2$. The initial observation for this case is initiated at time $t_1 = 16:44:1.52$ UTC 29 Feb 2016 and the second observation is initiated at $t_2 = 02:10:41$ UTC 01 Mar 2016. The time between the first and second observations in this case is about 9.5 hours and the determined states for both measurements are given below

$$\mathbf{x}_d(t_1) = \begin{bmatrix} 149.0^\circ & -11.3^\circ & 0.066^\circ/\text{s} & -0.107^\circ/\text{s} \end{bmatrix} \quad (5.53)$$

$$\mathbf{x}_d(t_2) = \begin{bmatrix} 171.8^\circ & 42.6^\circ & 0.089^\circ/\text{s} & -0.045^\circ/\text{s} \end{bmatrix} \quad (5.54)$$

Fujimoto et. al. describe the multi-rev solution problem in the intersection based association approach outlined in [61]. Ambiguity may arise when the observations of a space object may not be associated due to this multi-rev problem. The solutions presented below highlight this problem as Figure 5.13 shows for both the forwards and reverse mapping that there are several local minima that are selected as potential solutions. Figures 5.13b and d shows the expected case where the optimizer settles into multiple of the lowest valleys of the distance metric surface. Note that because a numerical optimizer is being used, there can be cases when an infeasible solution is selected, for example along the boundary of the admissible region. These solutions can safely be omitted because their probability of association value will be either exactly zero or very small. After omission of infeasible solutions, table 5.5 shows each of the pairs of solutions and the corresponding P_A values. The overall probability of association for this case is $\mathcal{P}_A = .5991$ indicating again that, even though there are more than two possible points of intersection, the two observations are likely associated. As such, the approximate PDF of the state estimate at time t_1 is shown in Figure 5.14.

The orbital elements associated with the three solutions for $\mathbf{x}_u(t_2)$ are given in Tables

5.6, 5.7, and 5.8. Note again the importance of considering all three local minima as possible solutions, while the probability of association value might be highest for Solution 1, the actual solution that matches best with truth is Solution 2. Note also that in Figure 5.14 the PDF comparison is only shown with the PDF for Solution 2 because in the particle filter approach the other local minima are eliminated by the full Bayesian update. The hypothesis test simply tries to determine if there is evidence that a point of intersection exists at each local minima. The three solutions shown in Table 5.5 all have significant evidence for an intersection but does not correspond to the weightings computed in the particle filter. Each of these solutions are indeed considered in the particle filter as they would each be assigned a non-zero weight, however the weightings assigned to Solutions 1 and 3 are significantly lower than those assigned to Solution 2, thus the particle filter only generates a PDF about Solution 2 with 50000 particles. Given an infinite number of particles, all three solutions would be present in the particle filter PDF as well. Figure 5.12 illustrates the three orbits that the optimizer finds to have significant probability of association.

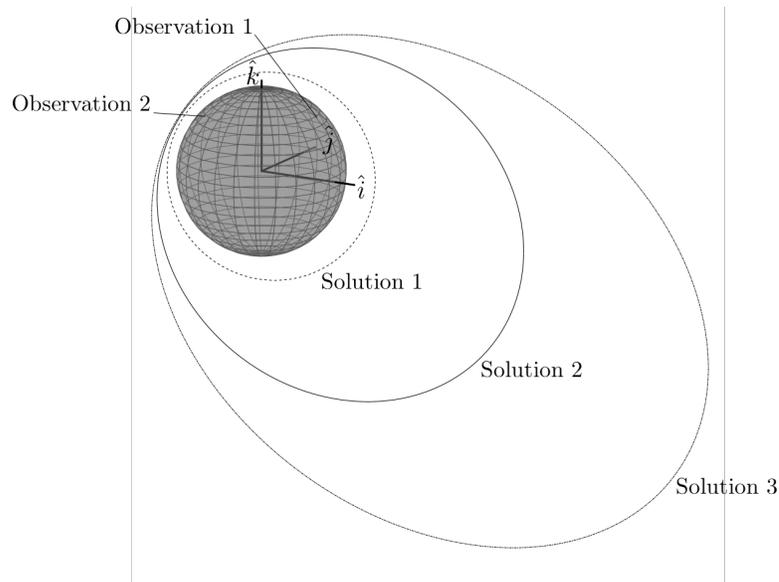


Figure 5.12: Illustration of three potential local minima solutions found by the optimizer

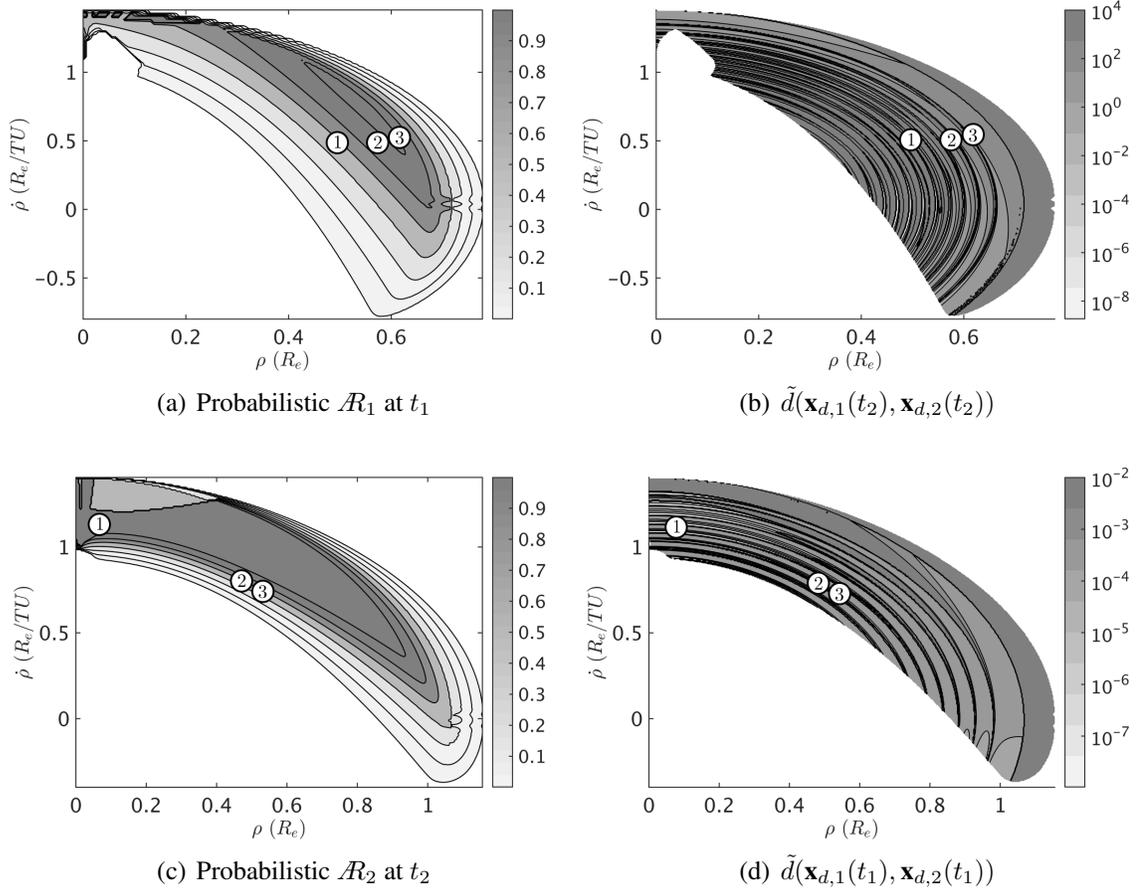


Figure 5.13: The contours of the probability of set membership and distance metrics for Case 2 with solutions labeled at time t_1 and t_2 .

Table 5.5: Case 2 Probability of association and optimization solutions

Solution	1	2	3
P_A	0.572	0.588	0.634
P_{FA}	0.05	0.05	0.05
$\rho(t_1)$ (R_e)	0.496	0.581	0.612
$\dot{\rho}(t_1)$ (R_e/TU)	0.489	0.507	0.531
$\rho(t_2)$ (R_e)	0.065	0.471	0.527
$\dot{\rho}(t_2)$ (R_e/TU)	1.135	0.800	0.730

Table 5.6: Estimated Orbit for Case 2 - Solution 1

	Estimate	Cov. (3σ)	Error
Semi-major axis, a	9334.1 km	1.77 km	2300.1 km
Eccentricity, e	.316	1×10^{-4}	0.315
Inclination, i	81.7°	0.009°	6.30°
RAAN, Ω	0.313°	0.010°	18.687°
Argument of perigee, ω	46.4°	0.074°	181.0°
True Anomaly, ν	3.88°	0.074°	154.12°

Table 5.7: Estimated Orbit for Case 2 - Solution 2

	Estimate	Cov. (3σ)	Error
Semi-major axis, a	7033.6 km	1.41 km	0.372 km
Eccentricity, e	.0011	2×10^{-4}	1.7×10^{-4}
Inclination, i	88.0°	0.011°	0.002°
RAAN, Ω	19.0°	0.016°	0.014°
Argument of perigee, ω	-135.7°	11.860°	1.145°
True Anomaly, ν	-154.1°	11.862°	2.092°

Table 5.8: Estimated Orbit for Case 2 - Solution 3

	Estimate	Cov. (3σ)	Error
Semi-major axis, a	7023.9 km	5.31 km	10.05 km
Eccentricity, e	.0431	5×10^{-4}	0.042
Inclination, i	88.2°	0.011°	0.194°
RAAN, Ω	22.6°	0.019°	3.561°
Argument of perigee, ω	-160.2°	0.696°	25.55°
True Anomaly, ν	-127.5°	0.705°	28.77°

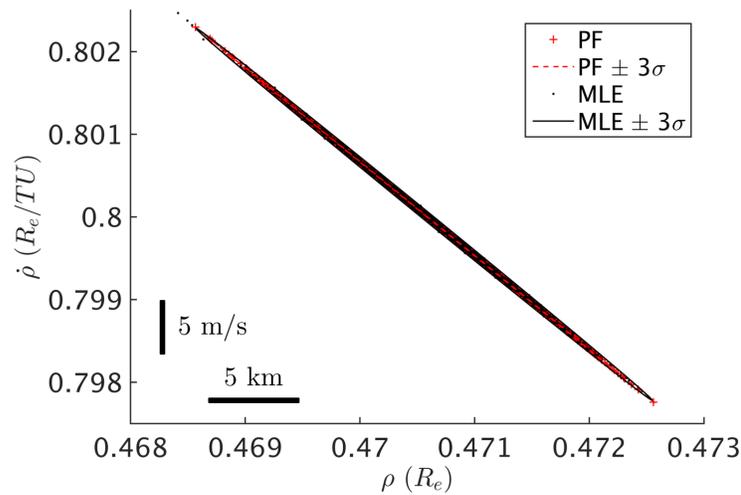


Figure 5.14: The numerical PDF after ingesting the newly associated observation and points sampled from the approximated PDF generated with the MLE approximation.

5.2.3 Case 3: GPS Satellite Observation

While the simulation shows good results from this optimization based approach, it is necessary to determine how the approach performs on real data. Using the Georgia Tech Space Object Research Telescope (GT-SORT), a Raven class telescope, a sequences of images were obtained of the GPS BIIR-11 satellite (NORAD ID 28190). The orbital elements for the GPS BIIR-11 satellite are determined from TLEs during this observation epoch are given in Table 5.9. For this configuration, the GT-SORT imager captured 10s exposures

Table 5.9: True Orbit for Case 3

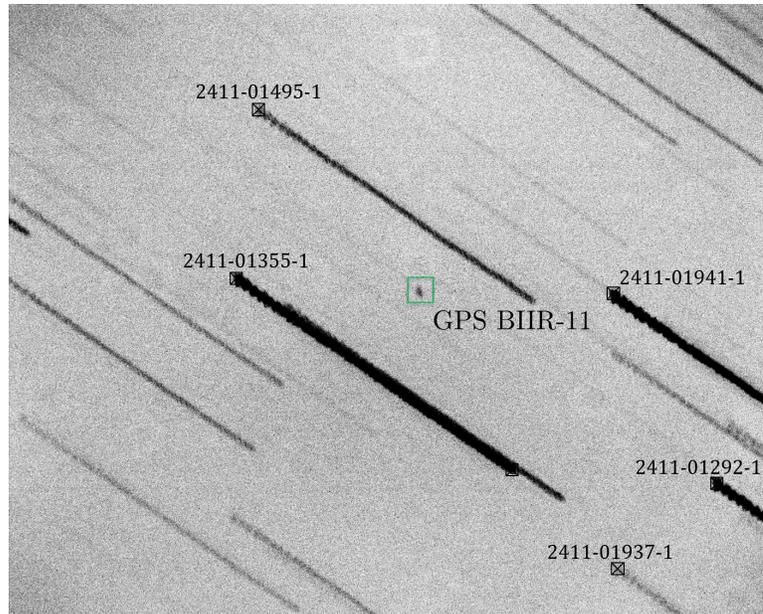
Semi-major axis, a	26559.7 km
Eccentricity, e	.01061
Inclination, i	55.7441°
RAAN, Ω	63.5728°
Argument of perigee, ω	40.9903°
True anomaly, ν	73.9°

with no filter. The first observation is captured at 02:14:52.590 UTC and the second observation is captured at 04:30:59.979 UTC on 30 Mar 2016. The images containing the GPS BIIR-11 object are shown in Figure 5.15. The star labels shown refer to the Tycho-2 star identification system. The total time between the observations is 2.268 hours, which is sufficient for an attempt at initial orbit determination since the unobservability criterion for GPS altitude is 1.06 hours. The measurements extracted from the images in Figure 5.15 by utilizing morphological operations to find the centroids as well as endpoints of the star streak geometries. The extracted measurements are described as follows

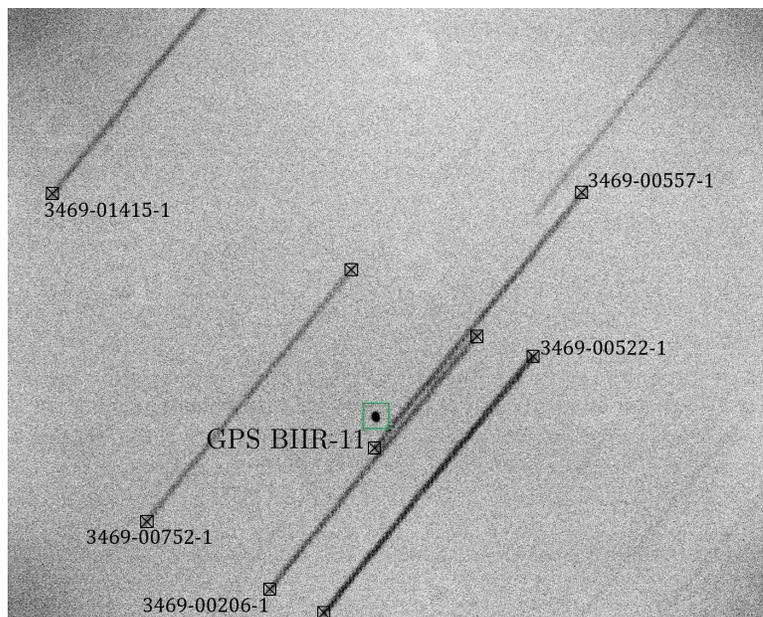
$$\mathbf{x}_d(t_1) = \begin{bmatrix} 82.0^\circ & 34.8^\circ & 0.007^\circ/\text{s} & 0.008^\circ/\text{s} \end{bmatrix} \quad (5.55)$$

$$\mathbf{x}_d(t_2) = \begin{bmatrix} -156.9^\circ & 51.4^\circ & 0.012^\circ/\text{s} & -0.006^\circ/\text{s} \end{bmatrix} \quad (5.56)$$

Figure 5.16 shows the admissible regions and distance metric contours for this example along with the location of the minima for this case. Table 5.10 shows the solutions and



(a) 30 Mar 2016 02:14:52.590 UTC at start of exposure



(b) 30 Mar 2016 04:30:59.979 UTC at start of exposure

Figure 5.15: Real images containing the tracked GPS BIIR-2 satellite. The streaks are the stars captured during the 10 second exposure. Field of view is 13.3×10.6 arc-minutes.

the individual probabilities of associations for each solution. In this case, the two pairs of solutions found by the optimizer are identical and $\mathcal{P}_A = .932$. Figure 5.17 shows the approximated updated PDF, which again can be shown to match reasonably well with the

particle filter results. The orbit state estimate is shown in Table 5.11. Note that the error in the state estimate solution is orders of magnitude higher in semi-major axis compared to the simulated test cases. While the method introduced in Chapter 2 enables systemic uncertainties to be taken into account, it does not directly take into account biases which may be present in the observation system. In the simulated cases, these biases are not of concern, however the data obtained from GT-SORT and the data reduction process itself may be a source of bias which affects the accuracy of this optimization based approach. Furthermore, there is an issue of dynamics bias introduced by assuming Keplerian dynamics in the approach. These biases contribute to the errors shown in Table 5.11.

One of the primary benefits of this methodology is that the hypothesis testing framework not only gives a probability of association for each pair of observations, but also provides a state estimate and estimated covariance directly eliminating the need to instantiate a particle filter. Each of the test cases show very good agreement between the full particle filter and the MLE approximation. Thus it is justifiable to state that this analytically approximated state and covariance estimate can be sufficiently used in place of the particle filter result. Furthermore, because this approximation is analytical it is very computationally efficient even considering the optimization required to find the local minima. To illustrate this fact, the run times for these test cases are shown in Table 5.12. Even in the cases where there are multiple solutions, the MLE approximation has a CPU run time at least an order of magnitude less than that of the particle filter. Note that the general association problem is, in worst case, factorial which implies that any computational improvements are also essentially factorial. The two order of magnitude reduction in computation time using this methodology would significantly outperform Particle Filter based association methods as the number of uncorrelated observations to associate increases.

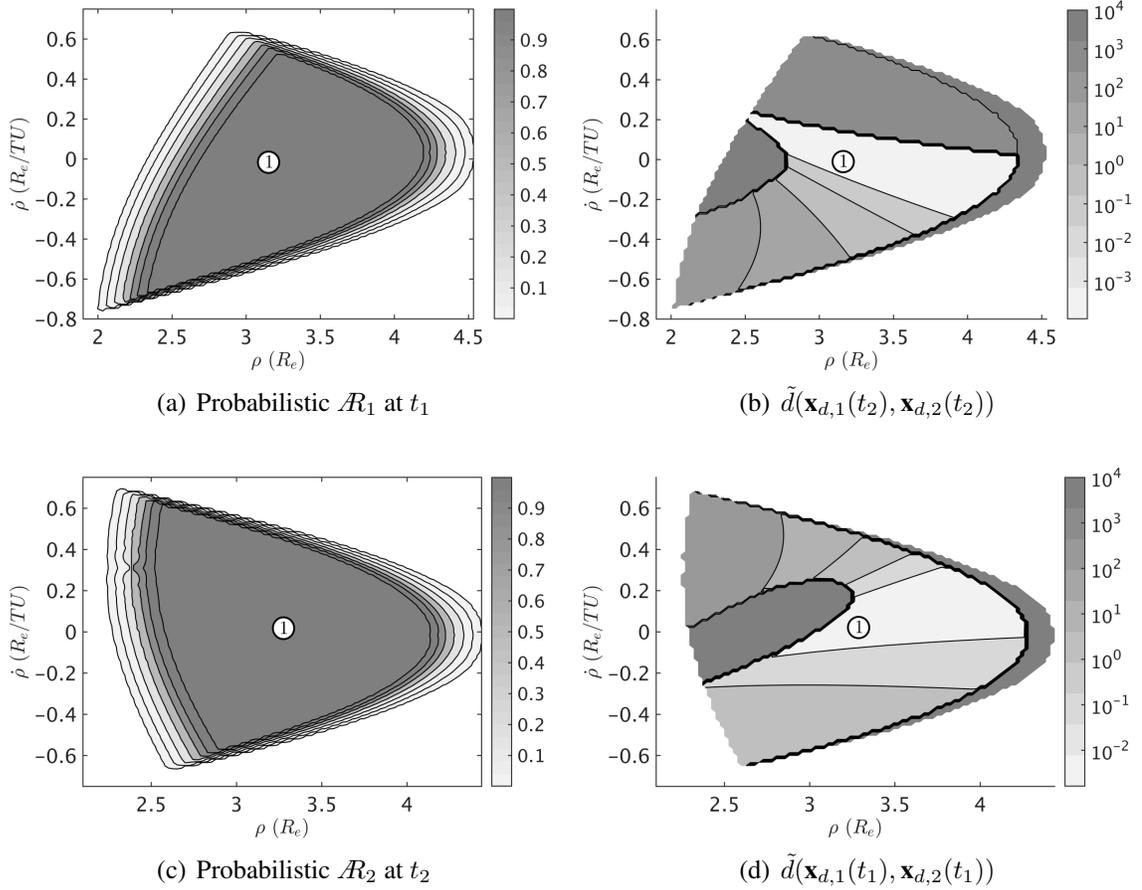


Figure 5.16: The contours of the probability of set membership and distance metrics for Case 3 with solutions labeled at time t_1 and t_2 .

Table 5.10: Case 3 Probability of association and optimization solutions

Solution	1
P_A	0.932
P_{FA}	0.05
$\rho(t_1) (R_e)$	3.151
$\dot{\rho}(t_1) (R_e/TU)$	0.0011
$\rho(t_2) (R_e)$	3.295
$\dot{\rho}(t_2) (R_e/TU)$	0.0207

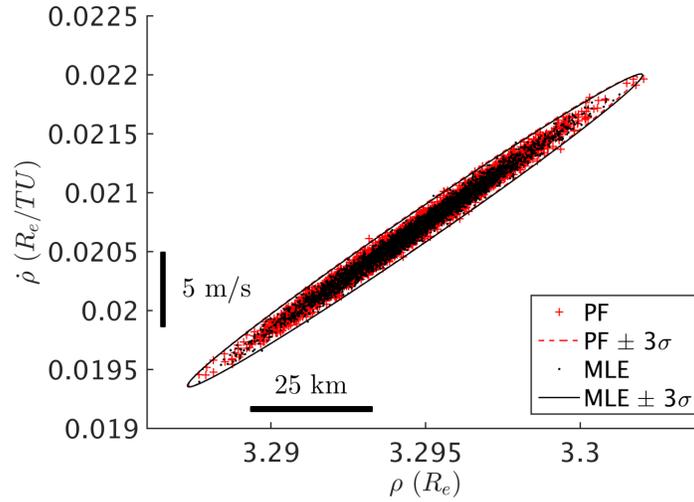


Figure 5.17: The updated approximate state PDF for Case 3 after ingesting the newly associated observation

Table 5.11: Estimated Orbit for Case 3

	Estimate	Cov. (3σ)	Error
Semi-major axis, a	27005.2 km	26.86 km	446.6 km
Eccentricity, e	.0270	3×10^{-4}	0.016
Inclination, i	55.8°	0.018°	0.130°
RAAN, Ω	63.4°	0.033°	0.012°
Argument of perigee, ω	43.88°	1.972°	2.147°
True Anomaly, ν	71.9°	1.963°	2.117°

Table 5.12: CPU Run Time Comparison

	MLE Approximation	Particle Filter
Case 1	7.392s	220.39s
Case 2	14.95s	220.25s
Case 3	7.692s	230.42s

5.2.4 Case 4: GEO Object Observation (Semi-Unobservable)

This case focuses on an observation of an object in GEO over a very short time period using GT-SORT. Two of the images from this sequence are shown in Figures 5.18 and 5.19. The black boxes outline the stars that were detected in the image. The exposure time is 5 seconds and the field of view is 13.3×10.6 arc-minutes.

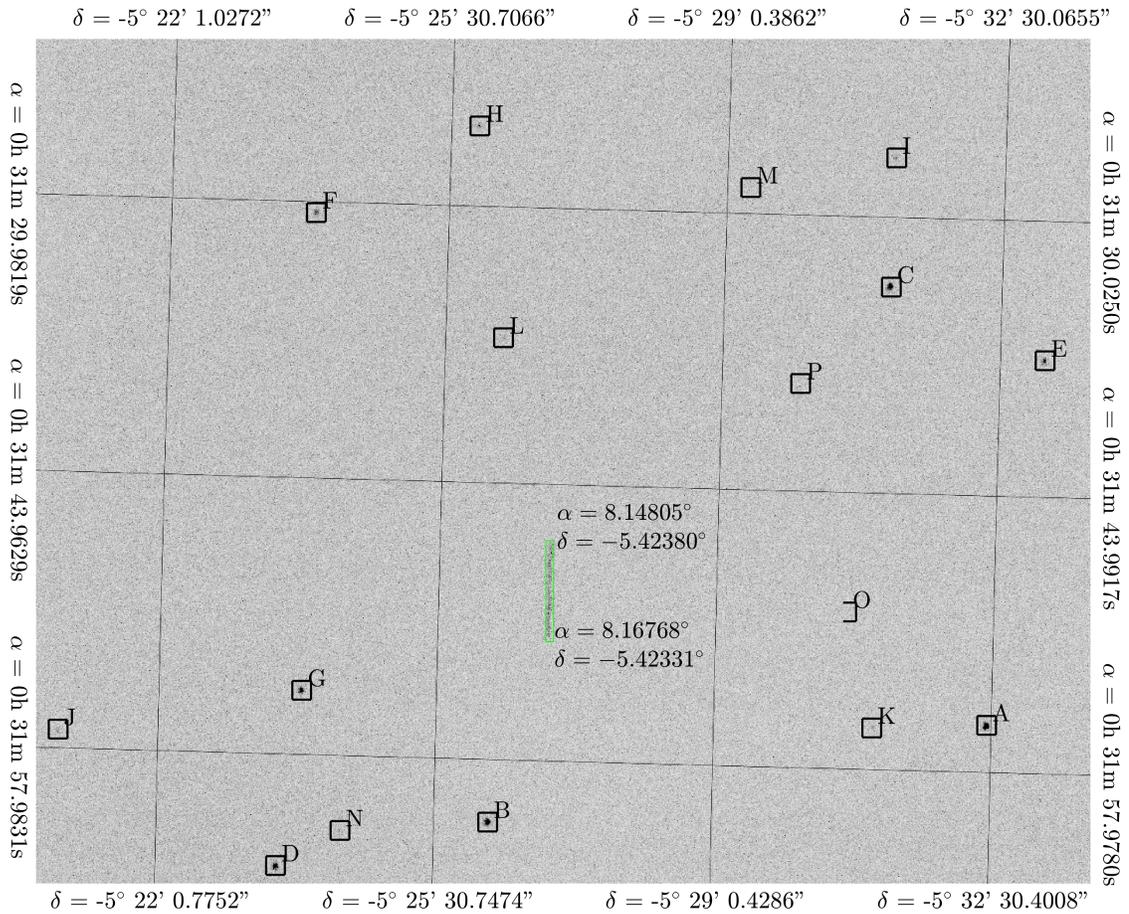


Figure 5.18: Real image containing an uncorrelated streak. 14 Jan 2016 01:02:21.429 UTC at start of exposure.

The streak captured from the object is highlighted by a green box with the right ascension and declination given in degrees. The inertial bearing angles given by the gridlines on superimposed on the image are generated from identifying the stars in the image using the

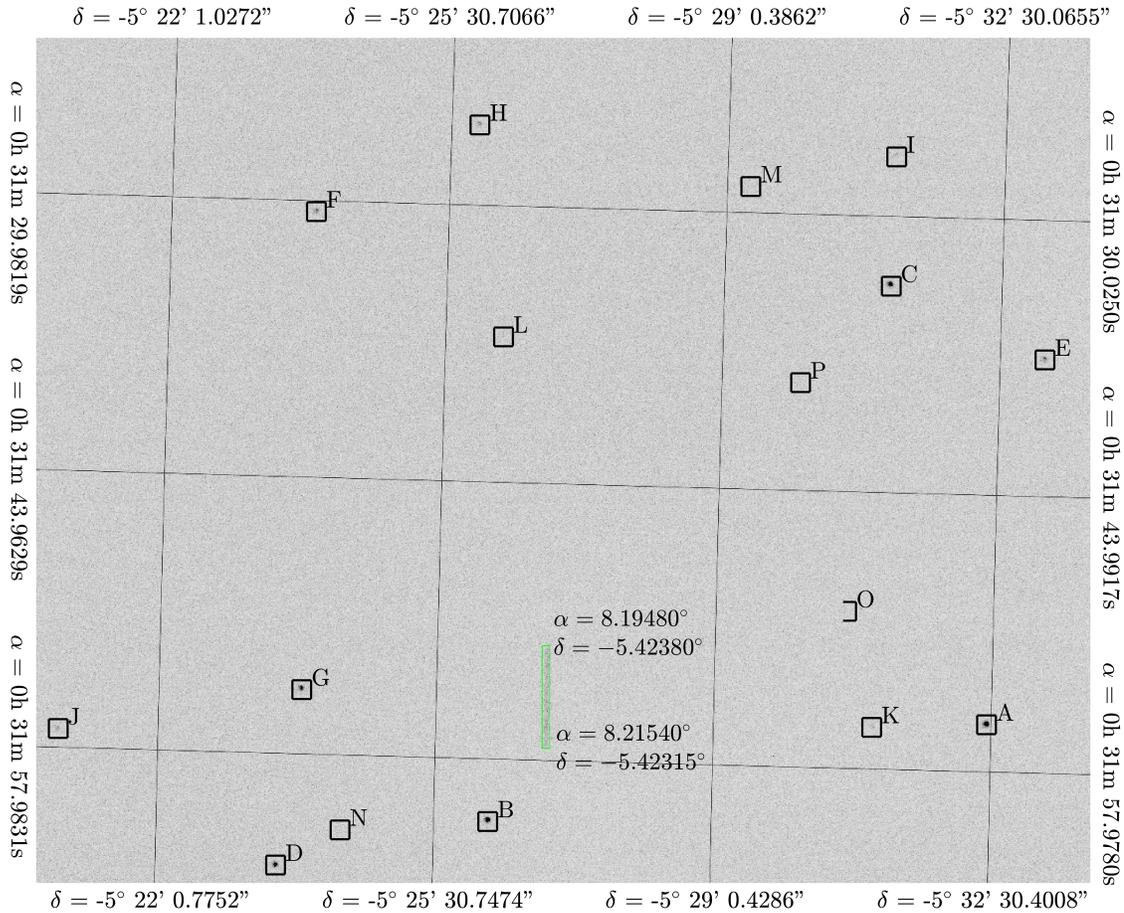


Figure 5.19: Real image containing an uncorrelated streak. 14 Jan 2016 01:02:32.549 UTC at start of exposure

astrometry.net suite³. The star IDs are included in Table 5.13 due to their length.

Figure 5.20 just shows closeup views of the streaks captured by GT-SORT. The streak shown in Figure 5.18a was captured at 01:02:21.429 UTC on 14 Jan 2016 with a 5 second integration time. The streak shown in Figure 5.18b was captured at 01:02:32.549 UTC on 14 Jan 2016 also with a 5 second integration time. The measurements extracted from the

³<http://nova.astrometry.net/>

Table 5.13: SDSS Star IDs

A	1237679077518803255
B	1237679077518868483
C	1237679077518802951
D	1237679322849542204
E	1237679077518802967
F	1237679322849476639
G	1237679322849542190
H	1237679077518802946
I	1237679077518802948
J	1237679322849542196
K	1237679077518802976
L	1237679077518802964
M	1237679077518802982
N	1237679322849542363
O	1237679077518803001
P	1237679077518802990

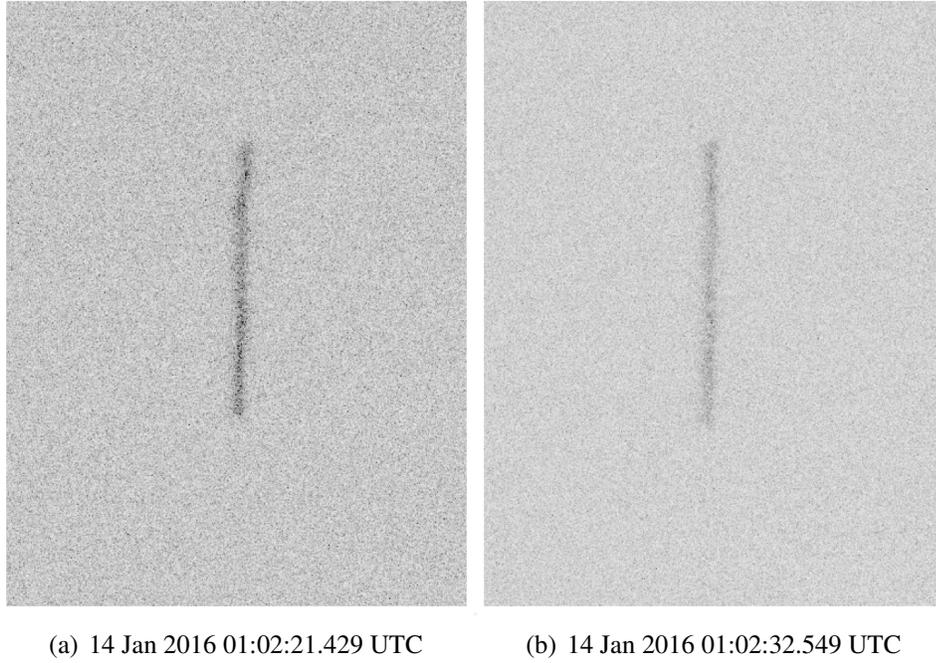


Figure 5.20: Closeup image of the streaks of the GEO object.

images in Figure 5.18 are given as follows

$$\mathbf{x}_d(t_1) = \begin{bmatrix} 8.15^\circ & -5.42^\circ & 0.0039^\circ/\text{s} & 9 \times 10^{-5} \cdot^\circ/\text{s} \end{bmatrix} \quad (5.57)$$

$$\mathbf{x}_d(t_2) = \begin{bmatrix} 8.19^\circ & -5.42^\circ & 0.0041^\circ/\text{s} & 0.0001^\circ/\text{s} \end{bmatrix} \quad (5.58)$$

The time between the observations is 11.1 seconds, which along with Eqn. (5.43) can be used to show that this observation sequence is unobservable. Let $\|\mathbf{r}\| = 6678\text{km}$, which is representative of a low LEO object and can be treated as a lower bound on the time required between observations for observability. Then the observability condition becomes

$$\sqrt{\frac{\|\mathbf{r}\|^3}{3\mu}} = \sqrt{\frac{6678\text{km}^3}{3 \cdot 3.986 \times 10^5}} = 499\text{s} \gg 11.1\text{s} \quad (5.59)$$

Thus, this observation scenario falls under the unobservable case discussed previously. Figure 5.21 shows the probabilistic admissible regions formed for both observations as well as a contour plot of P_A for each point sampled. For this unobservable scenario, the distance metric evaluated across \mathcal{R}_1 and \mathcal{R}_2 is approximately flat as can be seen in Figure 5.22. As such, Figures 5.21b and 5.21d just give examples of what the individual P_A value would be for points sampled within both \mathcal{R}_1 and \mathcal{R}_2 . Because the points found by the optimizer are unlikely to indicate the location of a solution producing a valid state estimate, the Table of solutions shown in the previous sections is omitted and instead the weighted norm of the probability of association is computed across the entire admissible region. In this regard, the weighted norm for the unobservable case can be treated as an ∞ -norm on the probability of association values. Applying the ∞ -norm to P_A in this case yields $\mathcal{P}_A = 0.903$ for these two observations, again indicating that association is very likely.

The Bayesian update in this case cannot be performed via the MLE approximation and a particle filter is used instead. Note that the updated PDF is nearly equivalent to the admissible region shown in Figure 5.21c. This is attributed to the unobservable nature of the observations, while it can be shown that they are associated, little information is gained after ingesting the associated observation into the particle filter.

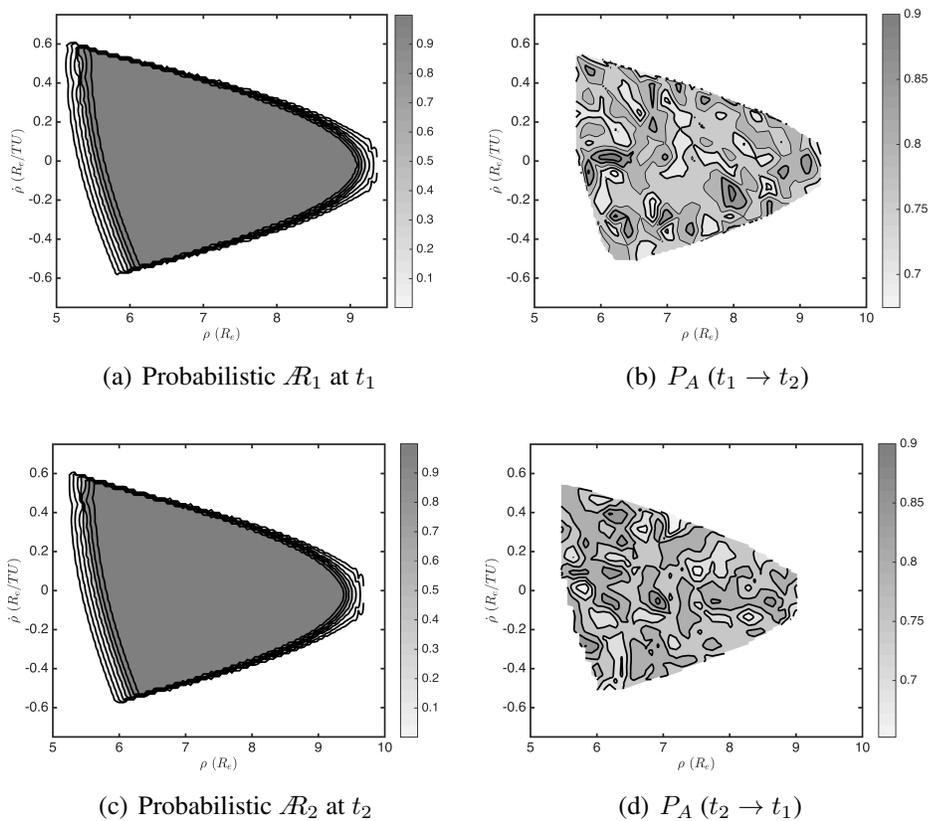


Figure 5.21: The contours of the probability of set membership and contours of P_A for Case 4.

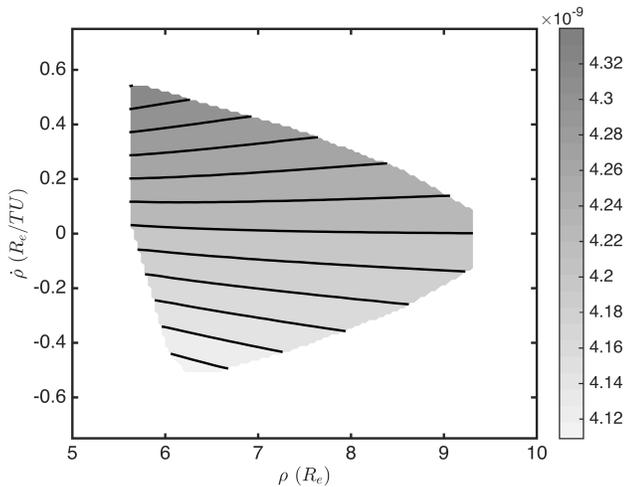


Figure 5.22: The distance metric surface computed for the unobservable case is essentially uniformly the point of intersection

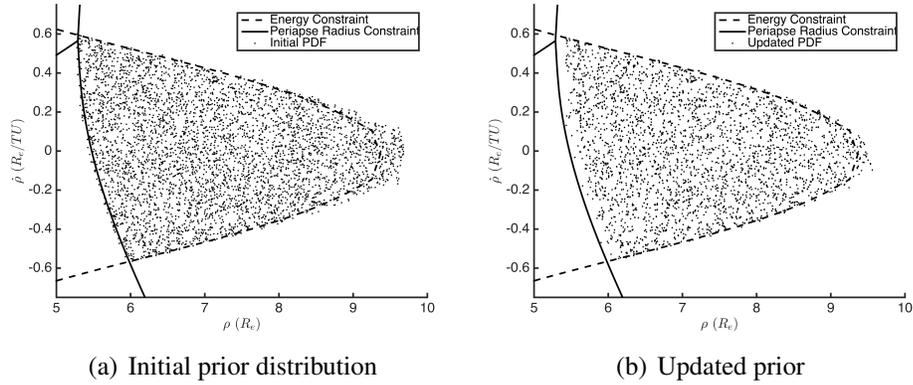


Figure 5.23: The updated state prior for Case 4 after ingesting the newly associated observation into a particle filter at time t_1

5.3 Conclusions

This chapter introduces an optimization based approach to perform observation association in the presence of systemic uncertainties. The general optimization methodology is based on minimizing the distance between the possible solutions over the full state space. The novel approach in this work optimizes over a partitioned state space. It is shown that the minimum distance between the partitioned state spaces is an equivalent method by which association may be determined. Conditions for the existence of possible state solutions are presented following the definition of the reduced order optimization problem. This methodology is then expanded to include systemic uncertainties. The hypothesis testing method is shown to be useful for associating observations in single solution cases. Large uncertainties and long time intervals between observations may give rise to optimization solutions with several local minima. The results shown utilize a single probability of association metric encompassing all such local minima solutions. This chapter also introduces an approximation which improves computational efficiency by combining the association and estimation by assuming the optimum solutions are the maximum likelihood estimator. This enables the distribution to be approximated as Gaussian and directly given by the inverse of the Hessian evaluated at the optimum. Several simulated test cases as well as real data

cases from the Georgia Tech Space Object Research Telescope are given to demonstrate the effectiveness of the methodology.

CHAPTER 6

FUTURE WORK

The contributions of the thesis work to improve state estimation methods for unobservable systems. However, there is still room for improvement and further development of the concepts presented. While Contributions 1 and 2 both are essentially complete treatments of uncertainty in the admissible region problem and necessary conditions for the admissible region prior respectively, Contributions 3 and 4 really only introduce new methodologies by which these state estimation problems may be addressed. Summarized below are potential areas of future work that will further develop the state of the art for initializing state estimation with uninformative priors as well as the association of short arc measurements.

6.1 Resampling DS Particle Filter

Chapter 4 details the Dempster-Shafer particle filter for initializing state estimation for unobservable dynamical systems based on admissible region theory. However, the algorithm presented in Chapter 4 does not utilize resampling of the plausibility space, Θ . As demonstrated, the presented algorithm works well to not only instantiate a Bayesian estimation scheme without making an assumption on the prior, but also to identify when an admissible region hypothesized constraint may be incorrect. At a theoretical level, the concept of resampling may violate the construction of Θ and Ω since the frame of discernment is assumed to be a mutually exclusive and also exhaustive list of hypotheses. Thus, the resampling algorithm developed must respect this constraint. The primary question then, is how to perform the resampling step when each particle has both a plausibility and belief value, when members of Θ to be sampled are non-singletons, and when the NOTA non-singleton set is unbounded. Future work will investigate how to best resample the DS particle filter in light of these concerns.

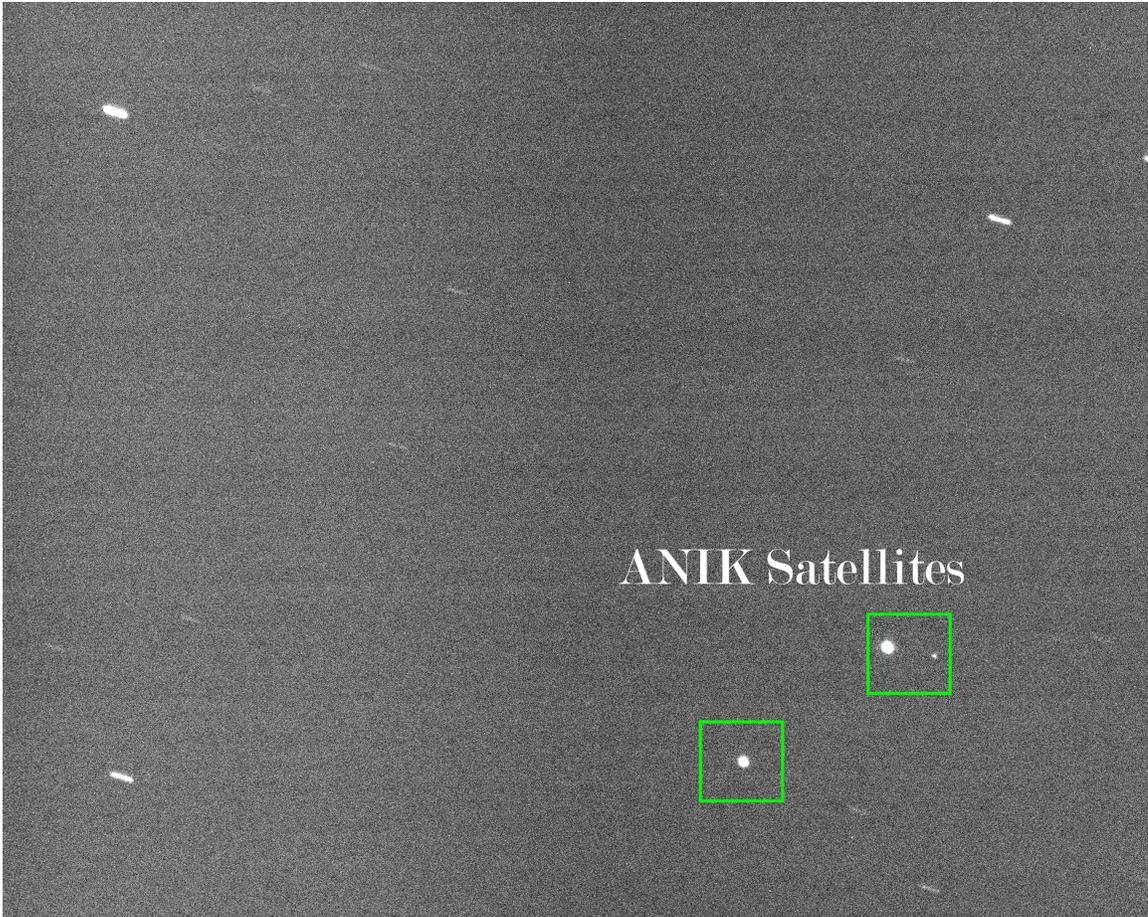


Figure 6.1: Three satellites in the ANIK-F1 Cluster at GEO

6.2 Evidence Based Association and Tracking

Chapter 5 builds on the admissible region theory by developing an optimization based approach to association of uncorrelated tracks. The probability of association metric developed works well to associate disparate observations for the simulated test cases and real data cases demonstrated. However, it is expected that for ambiguous cases where there are multiple satellites in very similar orbits or for formation flying clusters the association metric can give inconclusive results. Figure 6.1 shows a dataset from GT-SORT where these ambiguities actually pose a problem.

This problem may be remedied by combining the concepts introduced in Chapters 4

and 5 to develop an evidential reasoning based multiple hypothesis filter for association and tracking of objects. Future work should attempt to connect these methodologies for a more robust filter which takes into account that fact that belief mass could be assigned to the NOTA set when in reality the measurement fed to the DS particle filter is not associated with the already processed measurements.

6.3 Characterization and Tracking

Evidence based methods are often used to handle both characterization and tracking simultaneously. This may also be used to augment the capabilities of state estimation schemes for SSA by utilizing photometric data to further characterize the observed space objects. There are existing approaches which utilize multiple-model adaptive estimation schemes to tackle this problem [118], however the extension of DS theory to this problem might prove to be a more generalized approach. Future work will identify other hypothesis to add to the frame of discernment which can offer more information on the state of the object, for instance whether the object is tumbling or is stable.

6.4 High Area to Mass Ratio Objects

The association methodology presented in Chapter 5 uses the Mahalanobis distance as the cost function for the optimization problem. This can be substituted for a more appropriate cost function depending on the type of object being observed. For HAMR objects where the dynamics evolve with drag (if in LEO) and/or solar radiation pressure, using the Mahalanobis distance to associate two observations of the same HAMR object does not guarantee association. A more appropriate distance metric might be the control distance metric proposed by Jaunzemis and Holzinger [109]. Utilizing this methodology and attempting to find the optimal maneuver, or control cost, connecting two observations of HAMR objects may enable association to be carried out with more confidence than offered by current methods [56, 30]

6.5 Characterization of Dependent Errors

The association methodology presented in Chapter 5 takes into account the errors in the knowledge of the observer state as well as the timing error of the observer system. These are treated as independent error sources in the methodology presented, however it is known that these errors are not necessarily independent and as such it is a point of future work to take into account dependent sources of error in this association approach. Furthermore, it is a topic of future work to account for the potential underlying biases present in the observation systems and how these biases affect the optimization based association methodology.

CHAPTER 7

CONCLUSIONS

Initializing state estimation in underdetermined or unobservable systems is a difficult problem when posed with a single observation or a short sequence of unobservable observations. Particularly in orbit dynamics, the traditional state estimation methods, such as Gauss' method, fail when the measurement arc is too short. Short arc measurements will continue to be a part of the data obtained by sensors in the SSN, in addition to a widely growing network of commercial sensors, as objects continue to be launched into orbit and sensor technologies and efficiencies improve. This reinforces the need to improve estimation techniques instantiated on these too short arc measurements.

The admissible region method was developed initially to utilize hypothesized constraints to bound a set in the unobservable subspaces of the dynamical system of interest. Chapter 2 builds upon the admissible region method further by allowing for systemic uncertainties to be taken into account in the problem. The hypothesized constraints essentially form a set membership function, and the traditional admissible region method results in a binary membership function. The probability of set membership function introduced in Chapter 2 makes the admissible region instead a fuzzy set where the set membership for points near the nominal constraint boundary is dependent on the uncertainties in the system.

Chapter 3 details foundational principles probability transformation theory and then applies it to the admissible region problem. The admissible region itself is useful because it provides a bounded set from which a sequential estimator may be instantiated. However, there is no concise definition in the literature for how the admissible region should be treated probabilistically. Chapter 3 offers some insight on why the admissible region prior distribution cannot be truly treated as a PDF until the problem becomes observable. Fur-

thermore, it defines an observability condition that, when satisfied, implies there is enough observability for a true PDF to be defined.

But the contributions of Chapter 3 still impose some kind of assumed prior for the admissible region. What is truly desired is a method which does not require any kind of assumed prior and which can provide an indication of when the assumed constraint hypotheses are incorrect. This is provided with Dempster-Shafer theory as introduced in Chapter 4. The DS particle filter derived utilizes the vacuous nature of the belief mass assignment for a given short arc measurement to instantiate a DS filter with perfect plausibility of all states in the plausibility space and zero belief for each of states except for the entire admissible region itself. This filter then ingests measurements and updates the plausibility and belief values appropriately and collapses down to a traditional Bayesian PDF once the system is sufficiently observable. As demonstrated, this method not only requires no a priori assumption on the probabilities but also collapses down to match the distribution yielded from a traditional particle filter.

Finally, Chapter 5 details a methodology by which a set of uncorrelated tracks may be probabilistically associated with one another utilizing the admissible region method. The method builds on the previous work which shows that if a point of intersection may be found between the admissible regions of disparate observations, then the observations are likely associated. The optimization based approach described utilizes multiple reduced dimension sub-problems to identify a point of intersection in the d -dimensional subspaces of the admissible region. Furthermore, through a maximum likelihood estimator assumption, the solution and the Hessian at the solution can directly approximate the true distribution one would obtain from a traditional particle filter. The efficacy of the method is demonstrated not only on simulated test data but also on real observation data provided by GT-SORT.

To conclude, a methodology is first presented by which uncertainties in the observation systems may be incorporated in the sets defined in the admissible region method. Then,

various methods are explored which utilize the resulting fuzzy sets for initialization of state estimation methods. First, necessary conditions on when this fuzzy set may be treated as a PDF rather than an uninformative prior for the purposes of are shown. Then, an evidential reasoning based approach is developed which takes uninformative priors and uses plausibility and belief to bound the true PDF value without assuming an a priori distribution until the system is observable, at which point plausibility and belief collapse to the true PDF. Finally, an association method is develop which takes disparate observations and probabilistically associates them, and if a pair of observations are fully observable then provides a full state estimate in addition to the corresponding uncertainty distribution.

Appendices

APPENDIX A
CONSTRAINT DERIVATIVES

The approximate analytic equations for probability of set membership for the hypothesized energy and periapse radius constraints are based on partial derivatives of the constraint functions with respect to the combined parameter and determined state vector \mathbf{z} as defined in Chapter 2. Described below are the analytic derivatives used in the results section.

For the periaipse constraint:

$$\begin{aligned}
\frac{\partial \kappa_2}{\partial \mathbf{o}_x} &= 0 \\
\frac{\partial \kappa_2}{\partial \mathbf{o}_y} &= 0 \\
\frac{\partial \kappa_2}{\partial \mathbf{o}_z} &= 0 \\
\frac{\partial \kappa_2}{\partial \dot{\mathbf{o}}_x} &= 0 \\
\frac{\partial \kappa_2}{\partial \dot{\mathbf{o}}_y} &= 0 \\
\frac{\partial \kappa_2}{\partial \dot{\mathbf{o}}_z} &= 0 \\
\frac{\partial \kappa_2}{\partial \delta_2} &= -\frac{|\dot{\rho}|(2\delta - 2\delta_2)}{2 \left(\frac{(\delta - \delta_2)^2}{(t - t_2)^2} + \frac{\cos(\delta)^2(\alpha - \alpha_2)^2}{(t - t_2)^2} \right)^{3/2} (t - t_2)^2} \\
\frac{\partial \kappa_2}{\partial \delta} &= \frac{|\dot{\rho}| \left(\frac{2\delta - 2\delta_2}{(t - t_2)^2} - \frac{2 \cos(\delta) \sin(\delta)(\alpha - \alpha_2)^2}{(t - t_2)^2} \right)}{2 \left(\frac{(\delta - \delta_2)^2}{(t - t_2)^2} + \frac{\cos(\delta)^2(\alpha - \alpha_2)^2}{(t - t_2)^2} \right)^{3/2}} \\
\frac{\partial \kappa_2}{\partial \alpha_2} &= -\frac{\cos(\delta)^2 |\dot{\rho}| (2\alpha - 2\alpha_2)}{2 \left(\frac{(\delta - \delta_2)^2}{(t - t_2)^2} + \frac{\cos(\delta)^2(\alpha - \alpha_2)^2}{(t - t_2)^2} \right)^{3/2} (t - t_2)^2} \\
\frac{\partial \kappa_2}{\partial \alpha} &= \frac{\cos(\delta)^2 |\dot{\rho}| (2\alpha - 2\alpha_2)}{2 \left(\frac{(\delta - \delta_2)^2}{(t - t_2)^2} + \frac{\cos(\delta)^2(\alpha - \alpha_2)^2}{(t - t_2)^2} \right)^{3/2} (t - t_2)^2} \\
\frac{\partial \kappa_2}{\partial t_2} &= \frac{|\dot{\rho}| \left(\frac{2(\delta - \delta_2)^2}{(t - t_2)^3} + \frac{2 \cos(\delta)^2(\alpha - \alpha_2)^2}{(t - t_2)^3} \right)}{2 \left(\frac{(\delta - \delta_2)^2}{(t - t_2)^2} + \frac{\cos \delta^2(\alpha - \alpha_2)^2}{(t - t_2)^2} \right)^{3/2}} \\
\frac{\partial \kappa_2}{\partial t} &= -\frac{|\dot{\rho}| \left(\frac{2(\delta - \delta_2)^2}{(t - t_2)^3} + \frac{2 \cos(\delta)^2(\alpha - \alpha_2)^2}{(t - t_2)^3} \right)}{2 \left(\frac{(\delta - \delta_2)^2}{(t - t_2)^2} + \frac{\cos \delta^2(\alpha - \alpha_2)^2}{(t - t_2)^2} \right)^{3/2}}
\end{aligned}$$

For the energy constraint:

$$\begin{aligned}
\frac{\partial \kappa_1}{\partial \mathbf{o}_x} &= \frac{\mu(2\mathbf{o}_x + 2\rho \cos(\alpha) \cos(\delta))}{(\rho(2\mathbf{o}_z \sin(\delta) + 2\mathbf{o}_x \cos(\alpha) \cos(\delta) + 2\mathbf{o}_y \cos(\delta) \sin(\alpha) + \mathbf{o}_x^2 + \mathbf{o}_y^2 + \mathbf{o}_z^2 + \rho^2)^{3/2}} \\
\frac{\partial \kappa_1}{\partial \mathbf{o}_y} &= \frac{\mu(2\mathbf{o}_y + 2\rho \sin(\alpha) \cos(\delta))}{(\rho(2\mathbf{o}_z \sin(\delta) + 2\mathbf{o}_x \cos(\alpha) \cos(\delta) + 2\mathbf{o}_y \cos(\delta) \sin(\alpha) + \mathbf{o}_x^2 + \mathbf{o}_y^2 + \mathbf{o}_z^2 + \rho^2)^{3/2}} \\
\frac{\partial \kappa_1}{\partial \mathbf{o}_z} &= \frac{\mu(2\mathbf{o}_z + 2\rho \sin(\delta))}{(\rho(2\mathbf{o}_z \sin(\delta) + 2\mathbf{o}_x \cos(\alpha) \cos(\delta) + 2\mathbf{o}_y \cos(\delta) \sin(\alpha) + \mathbf{o}_x^2 + \mathbf{o}_y^2 + \mathbf{o}_z^2 + \rho^2)^{3/2}} \\
\frac{\partial \kappa_1}{\partial \dot{\mathbf{o}}_x} &= 2\dot{\mathbf{o}}_x - \rho \left(\frac{2 \cos \delta \sin \alpha (\alpha - \alpha_2)}{t - t_2} + \frac{2 \cos \alpha \sin \delta (\delta - \delta_2)}{t - t_2} \right) + 2\dot{\rho} \cos \alpha \cos \delta \\
\frac{\partial \kappa_1}{\partial \dot{\mathbf{o}}_y} &= 2\dot{\mathbf{o}}_y - \rho \left(\frac{2 \cos \delta \cos \alpha (\alpha - \alpha_2)}{t - t_2} + \frac{2 \sin \alpha \sin \delta (\delta - \delta_2)}{t - t_2} \right) + 2\dot{\rho} \sin \alpha \cos \delta \\
\frac{\partial \kappa_1}{\partial \dot{\mathbf{o}}_z} &= 2\dot{\mathbf{o}}_z + 2\dot{\rho} \sin \delta + \frac{2\rho \cos \delta (\delta - \delta_2)}{t - t_2} \\
\frac{\partial \kappa_1}{\partial \delta_2} &= \frac{\rho(\dot{\mathbf{o}}_x \cos \alpha \sin \delta - \dot{\mathbf{o}}_z \cos \delta + \dot{\mathbf{o}}_y \sin \alpha \sin \delta)2}{t - t_2} - \frac{\rho^2(2\delta - 2\delta_2)}{(t - t_2)^2} \\
\frac{\partial \kappa_1}{\partial \delta} &= \rho^2 \left(\frac{2(\delta - \delta_2)}{(t - t_2)^2} - \frac{2 \cos \delta \sin \delta (\alpha - \alpha_2)^2}{(t - t_2)^2} \right) - \dot{\rho}(2\dot{\mathbf{o}}_x \cos \alpha \sin \delta - 2\dot{\mathbf{o}}_z \cos \delta + 2\dot{\mathbf{o}}_y \sin \alpha \sin \delta) \\
&\quad - \rho \left(\frac{(2\dot{\mathbf{o}}_x \cos \alpha \sin \delta - 2\dot{\mathbf{o}}_z \cos \delta + 2\dot{\mathbf{o}}_y \sin \alpha \sin \delta)}{t - t_2} + \frac{(2(\delta - \delta_2)(\dot{\mathbf{o}}_z \sin \delta + \dot{\mathbf{o}}_x \cos \alpha \cos \delta + \dot{\mathbf{o}}_y \cos \delta))}{t - t_2} \right) \\
&\quad + \rho \left(\frac{(2\dot{\mathbf{o}}_y \cos \alpha \sin \delta - 2\dot{\mathbf{o}}_x \sin \alpha \sin \delta)(\alpha - \alpha_2)}{t - t_2} \right) \\
&\quad - \frac{\mu\rho(2\mathbf{o}_x \cos \alpha \sin \delta - 2\mathbf{o}_z \cos \delta + 2\mathbf{o}_y \sin \alpha \cos \delta)}{(\rho(2\mathbf{o}_z \sin(\delta) + 2\mathbf{o}_x \cos(\alpha) \cos(\delta) + 2\mathbf{o}_y \cos(\delta) \sin(\alpha) + \mathbf{o}_x^2 + \mathbf{o}_y^2 + \mathbf{o}_z^2 + \rho^2)^{3/2}} \\
\frac{\partial \kappa_1}{\partial \alpha_2} &= -\frac{2\rho(\dot{\mathbf{o}}_y \cos \alpha \sin \delta - \dot{\mathbf{o}}_x \cos \delta \sin \alpha)}{t - t_2} - \frac{\rho^2 \cos \delta^2 (2(\alpha - \alpha_2))}{(t - t_2)^2} \\
\frac{\partial \kappa_1}{\partial \alpha} &= \dot{\rho}(2\dot{\mathbf{o}}_y \cos \alpha \cos \delta - 2\dot{\mathbf{o}}_x \cos \delta \sin \alpha) - \rho \left(\frac{2(\alpha - \alpha_2)(\dot{\mathbf{o}}_y \cos \alpha \cos \delta - \dot{\mathbf{o}}_x \cos \delta \sin \alpha)}{t - t_2} \right) \\
&\quad - \rho \left(\frac{2\dot{\mathbf{o}}_y \cos \alpha \cos \delta - 2\dot{\mathbf{o}}_x \cos \delta \sin \alpha}{t - t_2} + \frac{2(\dot{\mathbf{o}}_y \cos \alpha \sin \delta - \dot{\mathbf{o}}_x \sin \alpha \cos \delta)(\delta - \delta_2)}{t - t_2} \right) + \frac{\rho^2 \cos \delta (2(\alpha - \alpha_2))}{(t - t_2)^2} \\
&\quad + \frac{\mu\rho(2(\mathbf{o}_y \cos \alpha \cos \delta - \mathbf{o}_x \cos \delta \sin \alpha))}{(\rho(2\mathbf{o}_z \sin \delta + 2\mathbf{o}_x \cos \alpha \cos \delta + 2\mathbf{o}_y \cos \delta \sin \alpha + \mathbf{o}_x^2 + \mathbf{o}_y^2 + \mathbf{o}_z^2 + \rho^2))^{3/2}} \\
\frac{\partial \kappa_1}{\partial t_2} &= \left(\frac{2(\delta - \delta_2)^2}{(t - t_2)^3} + \frac{2 \cos \delta^2 (\alpha - \alpha_2)^2}{(t - t_2)^3} \right) \rho^2 + \rho \frac{(2(\alpha - \alpha_2)(\dot{\mathbf{o}}_y \cos \alpha \cos \delta - \dot{\mathbf{o}}_x \cos \delta \sin \alpha))}{(t - t_2)^2} \\
&\quad - \rho \frac{(2(\delta - \delta_2)(\dot{\mathbf{o}}_x \cos \alpha \sin \delta - \dot{\mathbf{o}}_z \cos \delta + \dot{\mathbf{o}}_y \sin \alpha \sin \delta))}{(t - t_2)^2} \\
\frac{\partial \kappa_1}{\partial t} &= - \left(\frac{2(\delta - \delta_2)^2}{(t - t_2)^3} + \frac{2 \cos \delta^2 (\alpha - \alpha_2)^2}{(t - t_2)^3} \right) \rho^2 - \rho \frac{(2(\alpha - \alpha_2)(\dot{\mathbf{o}}_y \cos \alpha \cos \delta - \dot{\mathbf{o}}_x \cos \delta \sin \alpha))}{(t - t_2)^2} \\
&\quad + \rho \frac{(2(\delta - \delta_2)(\dot{\mathbf{o}}_x \cos \alpha \sin \delta - \dot{\mathbf{o}}_z \cos \delta + \dot{\mathbf{o}}_y \sin \alpha \sin \delta))}{(t - t_2)^2}
\end{aligned}$$

REFERENCES

- [1] C. R. Rojas, G. C. Goodwin, J. S. Welsh, and A. Feuer, “Optimal experiment design with diffuse prior information,” in *2007 European Control Conference (ECC)*, 2007, pp. 935–940.
- [2] R. L. Winkler, “The assessment of prior distributions in bayesian analysis,” *Journal of the American Statistical Association*, vol. 62, no. 319, pp. 776–800, 1967.
- [3] H. Jeffreys, “An invariant form for the prior probability in estimation problems,” *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 186, no. 1007, pp. 453–461, 1946.
- [4] P. Smets and B. Ristic, “Kalman filter and joint tracking and classification based on belief functions in the tbm framework,” *Information fusion*, vol. 8, no. 1, pp. 16–27, 2007.
- [5] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, “Multisensor data fusion: A review of the state-of-the-art,” *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.
- [6] M. K. Pitt and N. Shephard, “Filtering via simulation: Auxiliary particle filters,” *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.
- [7] S. J. Julier and J. K. Uhlmann, “New extension of the kalman filter to nonlinear systems,” in *AeroSense’97*, International Society for Optics and Photonics, 1997, pp. 182–193.
- [8] G. Welch and G. Bishop, “An introduction to the kalman filter,” 1995.
- [9] C. P. Robert, *Monte carlo methods*. Wiley Online Library, 2004.
- [10] G. Shafer *et al.*, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.
- [11] A. P. Dempster, “A generalization of bayesian inference,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no. 2, pp. 205–247, 1968.
- [12] L. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

- [13] R. P. Mahler, ““ statistics 101” for multisensor, multitarget data fusion,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 53–64, 2004.
- [14] H Klinkrad, P Beltrami, S Hauptmann, C Martin, H Sdunnus, H Stokes, R Walker, and J Wilkinson, “The {esa} space debris mitigation handbook 2002,” *Advances in Space Research*, vol. 34, no. 5, pp. 1251 –1259, 2004, Space Debris.
- [15] U. N. C. on the Peaceful Uses of Outer Space, *Space Debris Mitigation Guidelines of the Committee on the Peaceful Uses of Outer Space*. United Nations, 2010.
- [16] *Air Domain Surveillance and Intelligence Integration Plan*. Department of Homeland Security, 2007.
- [17] *National Maritime Domain Awareness Plan for the National Strategy of for Maritime Security*. 2013.
- [18] C. D. Wickens, “Situation awareness: Review of mica endsley’s 1995 articles on situation awareness theory and measurement,” *Human factors*, vol. 50, no. 3, pp. 397–403, 2008.
- [19] “Joint publication 3-14, space operations,” US Joint Chiefs of Staff, Washington, D.C., 2013.
- [20] “Technology horizons,” vol. 1, Office of the US Air Force Chief Scientist, Library of Congress, 2011, ISBN: AF/ST-TR-10-01-PR.
- [21] “National space policy of the united states of america,” Office of the President of the United States, Washington, D.C., 2010.
- [22] “Report of the scientific and technical subcommittee on its fifty-second session,” vol. 10(19), United Nations General Assembly, 2015, ISBN: A/AC.105/1088.
- [23] S. Kan, “China’s anti-satellite weapon test,” Congressional Research Service, Library of Congress order code RS22652, Report, 2007.
- [24] T. Wang, “Analysis of debris from the collision of the cosmos 2251 and the iridium 33 satellites,” *Science and Global Security*, vol. 18, no. 2, pp. 87–118, 2010.
- [25] J.-C. Liou, “Modeling the large and small orbital debris populations for environment remediation,” NASA Orbital Debris Program Office, Tech. Rep., 2014.
- [26] D. J. Heimerdinger, “Orbital debris and associated space flight risks,” in *Reliability and Maintainability Symposium, Alexandria, VA, IEEE*, 2005.

- [27] “Update on 3 major debris clouds,” vol. 14(2) *Orbital Debris Quarterly*, NASA Orbital Debris Program Program Office, 2010.
- [28] E. Christiansen, J. Hyde, and R. Bernhard, “Space shuttle debris and meteoroid impacts,” *Advances in Space Research*, vol. 34, no. 5, pp. 1097–1103, 2004.
- [29] A. J. Rosengren and D. J. Scheeres, “Long-term dynamics of high area-to-mass ratio objects in high-earth orbit,” *Advances in Space Research*, vol. 52, no. 8, pp. 1545–1560, 2013.
- [30] K. DeMars, M. Jah, D. Giza, and T. Kelecy, “Orbit determination performance improvements for high area-to-mass ratio space object tracking using an adaptive gaussian mixtures estimation algorithm,” in *21st International Symposium on Space Flight Dynamics*, 2009.
- [31] E. P. Chatters and B. J. Crothers, “Chapter 19: Space surveillance network,” in *Air University Space Primer*, US Air Force, 2009, ch. 19, pp. 249–258.
- [32] NASA, *Handbook for limiting orbital debris*, 2008.
- [33] W. Flury, A. Massart, T. Schildknecht, U. Hugentobler, J. Kuusela, and Z. Sodnik, “Searching for small debris in the geostationary ring,” *ESA Bulletin*, vol. 104, 2000.
- [34] D. Monet, T. Axelrod, C. Claver, T. Blake, R. Lupton, E. Pearce, R. Shah, and D. Woods, *Rapid cadence collections with the space surveillance telescope*, Conference Paper, 2012.
- [35] J. Shell, “Optimizing orbital debris monitoring with optical telescopes,” in *Advanced Maui Optical and Space Surveillance Technologies Conference*, Sep. 2010.
- [36] J. L. Foster, J. R. Benbrook, and E. G. Stansbery, “Detection of small radar cross-section orbital debris with the haystack radar,” in *Space Debris*, L. Anselmo, Ed., ser. *Advances in Space Research*. Kidlington: Pergamon-Elsevier Science Ltd, 2005, vol. 35, pp. 1210–1213, ISBN: 0273-1177.
- [37] A. A. Grometstein, *MIT Lincoln Laboratory Technology in Support of National Security*. Lexington, MA: Lincoln Laboratory, 2011, ISBN: 978-0-615-42880-2.
- [38] M. Morton and T. Roberts, “Joint space operations center (jspoc) mission system (jms),” in *AMOS Technologies Conference, Maui, HI*, 2011.
- [39] Committee on Space Debris, *Orbital debris: A technical assessment*, 1995.
- [40] R. F. Teehan, *Responsive space situation awareness in 2020*, Pamphlet, Maxwell Air Force Base, 2007.

- [41] W. Bauer, O. Romberg, C. Wiedemann, G. Drolshagen, and P. Vörsmann, “Development of in-situ space debris detector,” *Advances in Space Research*, vol. 54, no. 9, pp. 1858–1869, 2014.
- [42] N. R. Council, *Continuing Kepler’s Quest: Assessing Air Force Space Command’s Astrodynamics Standards*. Washington, DC: The National Academies Press, 2012.
- [43] L. G. Taff, “On initial orbit determination,” *Astronomical Journal*, vol. 89, pp. 1426–1428, Sep. 1984.
- [44] J. J. Degnan, “Satellite laser ranging: Current status and future prospects,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-23, no. 4, pp. 398–413, 1985.
- [45] A. Milani, G. F. Gronchi, M. d. Vitturi, and Z. Knežević, “Orbit determination with very short arcs. i admissible regions,” *Celestial Mechanics and Dynamical Astronomy*, vol. 90, no. 1-2, pp. 57–85, 2004.
- [46] A. Milani, G. F. Gronchi, Z. Knežević, M. E. Sansaturio, and O. Arratia, “Orbit determination with very short arcs: Ii. identifications,” *Icarus*, vol. 179, no. 2, pp. 350–374, 2005.
- [47] K. Fujimoto, D. J. Scheeres, J. Herzog, and T. Schildknecht, “Association of optical tracklets from a geosynchronous belt survey via the direct bayesian admissible region approach,” *Advances in Space Research*, vol. 53, no. 2, pp. 295–308, 2014.
- [48] K. Fujimoto, J. Maruskin, and D. Scheeres, “Circular and zero-inclination solutions for optical observations of earth-orbiting objects,” *Celestial Mechanics and Dynamical Astronomy*, vol. 106, no. 2, pp. 157–182, 2010.
- [49] K. DeMars and M. K. Jah, “Probabilistic initial orbit determination using gaussian mixture models,” *Journal of Guidance Control, and Dynamics*, vol. 36, no. 5, pp. 1324–1335, 2013.
- [50] I Hussein, C Roscoe, M Wilkins, and P Schumacher, “Probabilistic admissible region for short-arc angles-only observations,” in *Advanced Maui Optical and Space Surveillance Technologies Conference*, vol. 1, 2014, p. 76.
- [51] G. Shafer, “Belief functions and possibility measures,” *Analysis of fuzzy information*, vol. 1, pp. 51–84, 1987.
- [52] A. Dempster, “The dempster–shafer calculus for statisticians,” *International Journal of Approximate Reasoning*, vol. 48, no. 2, pp. 365–377, 2008, In Memory of Philippe Smets (1938–2005).

- [53] P. Smets and R. Kennes, “The transferable belief model,” *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [54] J. Rogers and M. Costello, “Smart projectile state estimation using evidence theory,” *Journal of Guidance, Control, and Dynamics*, vol. 35, no. 3, pp. 824–833, 2012.
- [55] D. Dubois and H. Prade, “A set-theoretic view of belief functions,” in *Classic Works of the Dempster-Shafer Theory of Belief Functions*, R. R. Yager and L. Liu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 375–410, ISBN: 978-3-540-44792-4.
- [56] C. Früh and T. Schildknecht, “Variation of the area-to-mass ratio of high area-to-mass ratio space debris objects,” *Monthly Notices of the Royal Astronomical Society*, vol. 419, no. 4, p. 3521, 2012.
- [57] A. Milani, G. Tommei, D. Farnocchia, A. Rossi, T. Schildknecht, and R. Jehn, “Correlation and orbit determination of space objects based on sparse optical data,” *Monthly Notices of the Royal Astronomical Society*, vol. 417, no. 3, pp. 2094–2103, 2011.
- [58] A. Milani and G. Gronchi, *Theory of orbit determination*. Cambridge University Press, 2010.
- [59] J. Siminski, O. Montenbruck, H. Fiedler, and T. Schildknecht, “Short-arc tracklet association for geostationary objects,” *Advances in Space Research*, vol. 53, no. 8, pp. 1184–1194, 2014.
- [60] J. M. Maruskin, D. J. Scheeres, and K. T. Alfriend, “Correlation of optical observations of objects in earth orbit,” *Journal of Guidance, Control, and Dynamics*, vol. 32, no. 1, pp. 194–209, 2009.
- [61] K. Fujimoto and D. Scheeres, “Correlation of optical observations of earth-orbiting objects and initial orbit determination,” *Journal of guidance, control, and dynamics*, vol. 35, no. 1, pp. 208–221, 2012.
- [62] K. DeMars, M. Jah, and P. Schumacher, “Initial orbit determination using short-arc angle and angle rate data,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2628–2637, 2012.
- [63] K. Fujimoto and K. T. Alfriend, “Optical short-arc association hypothesis gating via angle-rate information,” *Journal of Guidance, Control, and Dynamics*, pp. 1–12, 2015.

- [64] G. Tommei, A. Milani, and A. Rossi, “Orbit determination of space debris: Admissible regions,” *Celestial Mechanics and Dynamical Astronomy*, vol. 97, no. 4, pp. 289–304, 2007.
- [65] A. Milani, G. Gronchi, M. Vitturi, and Z. Knežević, “Orbit determination with very short arcs. i admissible regions,” *Celestial Mechanics and Dynamical Astronomy*, vol. 90, no. 1-2, pp. 57–85, 2004.
- [66] K. Fujimoto and D. J. Scheeres, “Applications of the admissible region to space-based observations,” *Advances in Space Research*, vol. 52, no. 4, pp. 696–704, 2013.
- [67] J. M. Maruskin, D. J. Scheeres, and K. T. Alfriend, “Correlation of optical observations of objects in earth orbit,” *Journal of Guidance, Control, and Dynamics*, vol. 32, no. 1, pp. 194–209, 2009.
- [68] J. A. Siminski, O. Montenbruck, H. Fiedler, and T. Schildknecht, “Short-arc tracklet association for geostationary objects,” *Advances in Space Research*, vol. 53, pp. 1184–1194, Apr. 2014.
- [69] B. Flury, *A First Course in Multivariate Statistics*, ser. Springer Texts in Statistics. Springer, 1997, pp. 1367–1375, ISBN: 9780387982069.
- [70] N. Kami, “Integrated formulation of the theory of belief functions from a linear algebraic perspective,” in *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, 2016, pp. 641–647.
- [71] M. J. Holzinger, “Using magnetometers for space object characterization in space situational awareness applications,” *Journal of Guidance, Control, and Dynamics*, vol. 37, no. 5, pp. 1397–1405, 2014.
- [72] K. Fujimoto and D. J. Scheeres, “Correlation of optical observations of earth-orbiting objects and initial orbit determination,” *Journal of Guidance, Control, and Dynamics*, vol. 35, no. 1, pp. 208–221, 2012.
- [73] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
- [74] F. Roberts and B. Tesman, *Applied combinatorics*. CRC Press, 2009.
- [75] A. Björklund, T. Husfeldt, and M. Koivisto, “Set partitioning via inclusion-exclusion,” *SIAM Journal on Computing*, vol. 39, no. 2, pp. 546–563, 2009.

- [76] J. L. Worthy and M. J. Holzinger, “Incorporating uncertainty in admissible regions for uncorrelated detections,” *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 9, pp. 1673–1689, 2015.
- [77] J. Galambos, “Bonferroni inequalities,” *The Annals of Probability*, pp. 577–581, 1977.
- [78] D. Montgomery and G. Runger, *Applied Statistics and Probability for Engineers*, 5th. John Wiley & Sons, 2010, ISBN: 9780470053041.
- [79] A. Chorin and O. Hald, *Stochastic Tools in Mathematics And Science*, ser. Stochastic Tools in Mathematics and Science. Springer, 2006, ISBN: 9780387280806.
- [80] R. Hogg, J. McKean, and A. Craig, *Introduction to Mathematical Statistics*. Pearson, 2013, ISBN: 9780321795434.
- [81] W. Brogan, “Modern control theory,” p. 382, 1991.
- [82] R. Hermann and A. J. Krener, “Nonlinear controllability and observability,” *Automatic Control, IEEE Transactions on*, vol. 22, no. 5, pp. 728–740, 1977.
- [83] M. Hwang and J. Seinfeld, “Observability of nonlinear systems,” *Journal of Optimization Theory and Applications*, vol. 10, no. 2, pp. 67–77, 1972.
- [84] A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, ser. McGraw-Hill series in electrical engineering: Communications and signal processing. Tata McGraw-Hill, 2002, ISBN: 9780070486584.
- [85] G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*. John Wiley & Sons, 2011, vol. 40.
- [86] C. Jauffret, “Observability and fisher information matrix in nonlinear regression,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 43, no. 2, pp. 756–759, 2007.
- [87] D.-J. Lee and K. T. Alfriend, “Sigma point filtering for sequential orbit estimation and prediction,” *Journal of Spacecraft and Rockets*, vol. 44, no. 2, pp. 388–398, 2007.
- [88] S. Julier, “The scaled unscented transformation,” in *American Control Conference, 2002. Proceedings of the 2002*, vol. 6, 2002, 4555–4559 vol.6.
- [89] R. S. Park and D. J. Scheeres, “Nonlinear mapping of gaussian statistics: Theory and applications to spacecraft trajectory design,” *Journal of Guidance, Control, and Dynamics*, vol. 29, no. 6, pp. 1367–1375, 2006.

- [90] P. Maybeck, *Stochastic Models, Estimation, and Control*, ser. Mathematics in Science and Engineering. Elsevier Science, 1982, ISBN: 9780080960036.
- [91] K. Fujimoto and D. J. Scheeres, “Tractable expressions for nonlinearly propagated uncertainties,” *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 6, pp. 1146–1151, 2015.
- [92] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons, 2008, vol. 2.
- [93] A. D. Jaunzemis and M. J. Holzinger, “Evidential reasoning applied to single-object loss-of-custody scenarios for telescope tasking,” in *AAS/AIAA Spaceflight Mechanics Meeting, Napa, CA*, 2016.
- [94] G. Shafer, “The combination of evidence,” *International Journal of Intelligent Systems*, vol. 1, no. 3, pp. 155–179, 1986.
- [95] K. Sentz and S. Ferson, *Combination of evidence in Dempster-Shafer theory*. Cite-seer, vol. 4015.
- [96] L. A. Zadeh and A. Ralescu, “On the combinability of evidence in the dempster-shafer theory,” in *Proceedings of the Second Conference on Uncertainty in Artificial Intelligence*, AUAI Press, 1986, pp. 347–349.
- [97] J. Dezert, P. Wang, and A. Tchamova, “On the validity of dempster-shafer theory,” in *Information Fusion (FUSION), 2012 15th International Conference on*, IEEE, 2012, pp. 655–660.
- [98] R. Munoz-Salinas, R. Medina-Carnicer, F. Madrid-Cuevas, and A. Carmona-Poyato, “Multi-camera people tracking using evidential filters,” *International Journal of Approximate Reasoning*, vol. 50, no. 5, pp. 732–749, 2009.
- [99] P. Smets, “Belief functions: The disjunctive rule of combination and the generalized bayesian theorem,” *International Journal of Approximate Reasoning*, vol. 9, no. 1, pp. 1–35, 1993.
- [100] ———, “Belief functions on real numbers,” *International Journal of Approximate Reasoning*, vol. 40, no. 3, pp. 181–223, 2005.
- [101] Y.-T. Hsia, “Characterizing belief with minimum commitment,” in *Proceedings of the 12th International Joint Conference on Artificial Intelligence - Volume 2*, ser. IJ-CAI’91, Sydney, New South Wales, Australia: Morgan Kaufmann Publishers Inc., 1991, pp. 1184–1189, ISBN: 1-55860-160-0.

- [102] D. Dubois and H. Prade, “The principle of minimum specificity as a basis for evidential reasoning,” in *Uncertainty in Knowledge-Based Systems: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems Paris, France, June 30 – July 4, 1986 Selected and Extended Contributions*, B. Bouchon and R. R. Yager, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1987, pp. 75–84, ISBN: 978-3-540-48020-4.
- [103] D. Koks and S. Challa, “An introduction to bayesian and dempster-shafer data fusion,” Tech. Rep., 2003.
- [104] J. L. W. III and M. J. Holzinger, “Use of uninformative priors to initialize state estimation for dynamical systems,” *Advances in Space Research (submitted)*, 2017.
- [105] T. Reineking, “Particle filtering in the dempster–shafer theory,” *International Journal of Approximate Reasoning*, vol. 52, no. 8, pp. 1124–1135, 2011.
- [106] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.
- [107] A. Naylor and G. Sell, *Linear Operator Theory in Engineering and Science*, ser. Applied Mathematical Sciences. Springer New York, 2000, ISBN: 9780387950013.
- [108] P. C. Mahalanobis, “On the generalized distance in statistics,” *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.
- [109] A. D. Jaunzemis, M. Mathew, and M. J. Holzinger, “Control cost and mahalanobis distance binary hypothesis testing for spacecraft maneuver detection,” *Journal of Guidance, Dynamics, and Control*, vol. 39, no. 9, pp. 2058–2072, 2016.
- [110] J. L. Worthy III and M. J. Holzinger, “Uncued satellite initial orbit determination using signals of opportunity,” in *AAS/AIAA Astrodynamics Specialist Conference*, Vail, CO, 2015.
- [111] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses*. Springer Science & Business Media, 2006.
- [112] J. S. McCabe and K. J. DeMars, “Particle filter methods for space object tracking,” in *AIAA/AAS Astrodynamics Specialist Conference*, 2014, p. 4308.
- [113] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley, 2006, ISBN: 9780470045336.
- [114] H. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, ser. Detection, Estimation, and Modulation Theory. Wiley, 2004, ISBN: 9780471463832.

- [115] H. Cramér, *Mathematical methods of statistics*. Princeton university press, 1945, vol. 9.
- [116] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” in *Selected Papers of Hirotugu Akaike*, Springer, 1998, pp. 199–213.
- [117] J. L. Worthy III and M. J. Holzinger, “Probability density transformations on admissible regions for dynamical systems,” in *AAS/AIAA Astrodynamics Specialist Conference, Vail, CO*, 2015.
- [118] R. Linares, J. Crassidis, M. Jah, and H. Kim, “Astrometric and photometric data fusion for resident space object orbit, attitude, and shape determination via multiple-model adaptive estimation,” in *AIAA Guidance, Navigation, and Control Conference*, p. 8341.