Use of Uninformative Priors to Initialize State Estimation for Dynamical Systems

AE8900 MS Special Problems Report
Space Systems Design Lab (SSDL)
Guggenheim School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA

Author:
Johnny L. Worthy III

Advisor:
Marcus J. Holzinger

December 11, 2015

Use of Uninformative Priors to Initialize State Estimation for Dynamical Systems

Johnny L. Worthy III*, Marcus J. Holzinger†
Georgia Institute of Technology, Atlanta, GA, 30332

The admissible region must be expressed probabilistically in order to be used in Bayesian estimation schemes. When treated as a probability density function (PDF), a uniform admissible region can be shown to have non-uniform probability density after a transformation. This paper uses the fundamental multivariate probability transformation theorem to show that regardless of which state space an admissible region is expressed in, the probability density must remain uniform. The admissible region is shown to be a special case of the Jeffreys' prior, an uninformative prior with a probability density that remains constant under reparameterization. This paper introduces requirements on how these uninformative priors may be transformed and used for state estimation.

I. Introduction

Observation systems often act in information deprived environments where a single observation cannot fully determine the state of an observed object. Even though the full state cannot be determined, it is useful to be able to initialize an estimation scheme capable of constructing a distribution over the possible true states of the system. While this problem is applicable to a vast variety of different dynamical systems and observation schemes, in this paper this problem will be address with respect to space situational awareness. For example, short optical or radar measurements of space objects do not provide enough information to uniquely determine the state of the space object. However, with the increasing number of objects in orbit around the Earth, characterizing these space objects is an active area of research. Currently, over 20,000 objects larger than 10 cm are tracked by a group of optical and radar sensors in the Space Surveillance Network (SSN) and it is estimated that catastrophic collisions are likely to occur every 5 to 9 years [1] [2]. Both types of sensors operate in data deprived environments as optical measurements cannot determine the range to the target and radar measurements cannot determine the angular position of the space object. However, over long observation periods or with multiple observations it is possible to use approaches such as Gauss's method or Lambert's method to fully estimate the state of a space object using these measurements [3]. The difficulty with these methods is that over short observation periods (relative to the time scale of the dynamics) there is an unobservable subspace causing traditional methods, such as Gauss's, to fail. For example, this problem readily presents itself in the case of an optical observation of an object generated by a streak captured over a matter of seconds. Traditional methods cannot produce an orbit state estimate from such an observation. Because of this, the development of initial orbit determination approaches based on short-arc optical and radar measurements is a very active area of research.

Several nonlinear initial orbit determination approaches are based on the admissible region method. When an observation is too short to provide enough geometric data for an initial orbit estimate, a continuum of possible solutions exist. The admissible region method uses hypothesized constraints to bound the feasible solutions to a closed and computationally tractable state space volume [4]. First proposed by Milani et. al., the admissible region method constrains the possible solutions for a given too-short arc (TSA) observation using the dynamics of the orbiting bodies and hypothesized constraints [4] [5]. Many have extended the admissible region's applicability to SSA since Milani et. al. For example, methods have been presented that discretize the admissible region and consider solutions at discrete points [6] [7]. Multiple hypothesis filter or particle filter methods can also be initialized from discretized admissible regions [8]. Optimization methods to identify a best fitting orbit solution are presented by Siminski et. al. [9]. A boundary value problem approach is applied to the admissible region by Fujimoto and Alfriend which uses the angle-rate information to eliminate hypotheses [10].

In general, Bayesian estimation techniques are initiated with a prior distribution of the initial state. Thus, admissible regions must be expressed probabilistically when used as prior distributions to initiate Bayesian estimation schemes. Fujimoto et. al. showed that the admissible region possesses a uniform probability density over the constrained unobservable state space volume; every state satisfying the constraint is equally likely to be true [11]. Without the inclusion of measurement, observer, and parameter uncertainty, a uniform PDF results in a probability discontinuity at the admissible region boundary. It is important to include uncertainty effects to remove this discontinuity when generating an admissible region in order to account for states that would otherwise have been assigned zero probability. DeMars and Jah indirectly accomplish this by using a Gaussian mixture model (GMM) to approximate the admissible region [12]. The edges of the admissible region as represented by a GMM are continuous as a result of the sum of the Gaussians, however this approach does not directly incorporate the measurement uncertainty into the admissible region. An approximate analytical expression for the exact probability distribution of an admissible region is presented by Worthy and Holzinger which directly accounts for uncertainty and errors in the measurements and observer state [13]. Hussein et. al. generate a probabilistic admissible region by uniformly sampling from alternative state space, such as semi-major axis and eccentricity, and mapping these uniformly sampled states into the admissible region [14]. The resulting non uniform PDF is approximated with a Gaussian mixture to describe the new admissible region. All of these approaches essentially treat the admissible region as a probability density function (PDF). The motivation of this paper is to determine what the correct statistical representation of the admissible region should be. Again, Fujimoto et. al. show that an admissible region is necessarily a uniform distribution [11], but what does this mean in regard to using the admissible region to instantiate a Bayesian estimator and how can the admissible region probability distribution be transformed to be used in such estimators.

Admissible regions are typically formulated in a specific state space based on the hypothesized constraints. Operational requirements may necessitate estimation schemes which operate in a different state space. Thus, the statistical representation of the admissible region must be mapped into the different state space. In general, probability mappings of PDFs are only restricted by the requirement that the transformation be left invertible [15]. This condition preserves probability across the transformation and is important to ensure that Bayesian estimation schemes are initiated properly. If the statistical representation of the admissible region is not best represented by a PDF then the requirements for probability mappings are no longer valid. Another motivation for this work is to determine the correct way to initiate a Bayesian estimation scheme from a statistical representation of an admissible region and identify the requirements for transforming an admissible region prior.

To address the questions of the correct statistical representation of admissible regions and the proper way to initialize a Bayesian estima-

*Graduate Researcher, School of Aerospace Engineering, and AIAA Student Member.

†Assistant Professor, School of Aerospace Engineering, and AIAA Senior Member.

tion scheme with an admissible region, this paper includes 1) a formal analytical relationship between the observability of a dynamical system and the statistical representation of the admissible region, 2) a formal expression of the constant gradient condition for the transformation of an admissible region prior consistent with Jeffrey's Rule and 3) a novel extension of the approximate analytical probability distribution function for the transformation of an admissible region.

This paper is organized as follows. The general requirements and approach to transform probability distributions are described in §II.A with applicability to admissible regions shown in §II.B. In §III.A the theory is applied to assess the validity of the transformation of probability from topocentric spherical to cartesian coordinates and show the conditions for observability of the system.

II. Probability Transformations for Admissible Regions

The general theory of probability transformations is an exhaustively studied topic in statistics and probability with a wide range of applications [15] [16] [17]. The purpose of this section is to introduce fundamental results regarding general probability mappings and apply them to the admissible region problem. This will show that the statistical representation of the admissible region is consistent with that of a Jeffrey's prior and not of a PDF.

A. General Probability Transformations

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

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}[\mathbf{X} \leq \mathbf{x}] = \int_A f_X(\mathbf{x})\mathrm{d}\mathbf{x} \tag{1}$$

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

$$F_{\widetilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = \mathbb{P}[\widetilde{\mathbf{X}} \leq \tilde{\mathbf{x}}] = \int_{\tilde{A}} f_X(\mathbf{g}^{-1}(\tilde{\mathbf{x}})) \cdot \mathrm{abs}\left(\left|\frac{\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}}\right|\right) \tag{2}$$

where $\tilde{A} = (-\infty, \tilde{x}_1] \times \cdots \times (-\infty, \tilde{x}_n]$ and $\mathrm{abs}\left(\left|\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})/\partial \tilde{\mathbf{x}}\right|\right)$ is the determinant of the Jacobian matrix and the absolute value ensures $f_{\widetilde{\mathbf{X}}}(\tilde{\mathbf{x}})$ is non-negative for all values of $\tilde{\mathbf{x}}$ [15]. The integrand of Eqn. (2) is by definition the PDF of $\widetilde{\mathbf{X}} = \mathbf{g}(\mathbf{X})$ and the following foundational theorem in multivariate statistics gives the PDF of the transformed variable.

Theorem 1 (Transformation theorem for continuous random variables [15])**.** *Given a PDF $f_X(\boldsymbol{x})$ and a left-invertible transformation $\tilde{\boldsymbol{x}} = \boldsymbol{g}(\boldsymbol{x})$ the PDF of the transformed variable $f_{\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{x}}_u)$ is given by*

$$f_{\widetilde{X}}(\tilde{\boldsymbol{x}}) = \begin{cases} f_X(\boldsymbol{g}^{-1}(\tilde{\boldsymbol{x}})) \cdot \mathrm{abs}(\left|\frac{\partial}{\partial \tilde{\boldsymbol{x}}}\boldsymbol{g}^{-1}(\tilde{\boldsymbol{x}})\right|) & \text{for } \tilde{\boldsymbol{x}} \in \mathcal{R}(\boldsymbol{g}(\tilde{\boldsymbol{x}})) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Proof. The proof of Theorem 1 is given in [15]. □

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

$$F_{\widetilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = F_{\mathbf{X}}(\mathbf{g}^{-1}(\tilde{\mathbf{x}})) \tag{4}$$

where $F_{(\cdot)}$ denotes the CDF over $\tilde{\mathbf{x}}$ or $\mathbf{x}$ [18]. 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 $\widetilde{\mathbf{X}}$.

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

Proof. The proof of Corollary 1 follows directly from the derivation and analysis of Eqn. (1). By definition Eqn. (1) is equal to Eqn. (2) and thus Eqn. (4) must hold. □

B. Admissible Region Transformations

The purpose of the following subsections is to outline why, in general, an admissible region prior cannot be transformed. The first subsection shows the application of the derivation of Eqn. (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) may be applied to an a posteriori PDF based on an admissible region without any additional conditions. As this section will prove, a uniform PDF representation is in contradiction to Theorem 1. This contradiction supports defining the statistical representation of the admissible region as an admissible region prior as opposed to a PDF.

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

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{5}$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}; \mathbf{k}, t) \tag{6}$$

where the measurement function is defined as $\mathbf{h} : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R} \to \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^m$ 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 exists to show observability of the general system given by Eqns. (5) and (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 $\mathbf{P} \in \mathbb{S}_+^{n \times n}$ [19] which is given in most general form as

$$\mathbf{P}(t_f, t_0, \mathbf{x}(t)) = \int_{t_0}^{t} \mathbf{\Phi}^T(\tau, t_0) \frac{\partial \mathbf{h}(\mathbf{x}(\tau); \mathbf{k}, \tau)}{\partial \mathbf{x}(\tau)}^T \frac{\partial \mathbf{h}(\mathbf{x}(\tau); \mathbf{k}, \tau)}{\partial \mathbf{x}(\tau)} \mathbf{\Phi}(\tau, t_0)\mathrm{d}\tau \tag{7}$$

where $\mathbf{\Phi} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ 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 $\mathrm{rank}[\mathbf{P}(\mathrm{t_f}, \mathrm{t_0}, \mathbf{x}(\mathrm{t}))] = \mathrm{n}$. If $\mathrm{rank}[\mathbf{P}(\mathrm{t_f}, \mathrm{t_0}, \mathbf{x}(\mathrm{t}))] < \mathrm{n}$ 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 [20]. It is shown in [21] that this condition may be sufficiently satisfied by linearizing Eqns. (5) and (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

$$\mathbf{W} = \{(\mathbf{x}_r(t), \mathbf{x}_r(t)) : \mathbf{x}_{r_{0,u}} \in \mathcal{R}, \mathbf{k}_{r_{0,d}} = \mathbf{h}^{-1}(\mathbf{y}, \mathbf{k}, t_0)\} \tag{8}$$

Linearization of Eqns. (5) and (6) about trajectories in $\mathbf{W}$ yields

$$\delta\dot{\mathbf{x}}(t) = \left.\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}_r(t),\mathbf{k}_r(t)} \delta\mathbf{x}(t) \tag{9}$$

$$\delta\mathbf{y}(t) = \left.\frac{\partial \mathbf{h}(\mathbf{x}, \mathbf{k}, t)}{\partial \mathbf{x}}\right|_{\mathbf{x}_r(t),\mathbf{k}_r(t)} \delta\mathbf{x}(t) \tag{10}$$

Observability of the linearized system may be determined directly by application of Eqn. (7) and checking the rank of the local linearized observability gramian. By Theorem 3.2 in [21], 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 [21] 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

$$\mathcal{O}(\mathbf{x}) = \begin{bmatrix} \mathbf{y} \\ \dot{\mathbf{y}} \\ \ddot{\mathbf{y}} \\ \vdots \end{bmatrix} \tag{11}$$

where the order of the derivatives of the output depend on the system [20]. 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 [21] 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 Keplerian dynamics over a time period* $t \in [t_0, t_f]$ *such that* $\Delta t = t_f - t_0 \ll \|\boldsymbol{r}\|^3/3\mu$ *yields a local linearized observability gramian with rank* $\mathbf{P}(t_f, t_0, (\boldsymbol{x}_r(t), \boldsymbol{k}_r(t))) = d < n\ \forall\ (\boldsymbol{x}_r(t), \boldsymbol{k}_r(t)) \in \boldsymbol{W}$*. Every point* $\boldsymbol{x}_u \in \mathcal{R}$ *is therefore unobservable for this short observation sequence. This Lemma is generalizable to any Newtonian system.*

Proof. Utilizing the linearized system described in Eqns. (9) and (10), the observability gramian is given by Eqn. (7). Consider the Taylor series approximation of the state transition matrix.

$$\boldsymbol{\Phi}(\tau, t_0) = \mathbb{I}_6 + \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\tau - t_0) + \text{H.O.T} \tag{12}$$

Taking the first order terms and assuming two body dynamics,

$$\mathbf{f}(\mathbf{x}, t) = \begin{bmatrix} \dot{\mathbf{r}} \\ -\frac{\mu}{\|\mathbf{r}\|^3}\mathbf{r} \end{bmatrix} \tag{13}$$

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

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0} & \mathbb{I}_3 \\ \mathbf{M} & \mathbf{0} \end{bmatrix} \tag{14}$$

where $\mathbb{I}_v$ is the $v \times v$ identity matrix and

$$\boldsymbol{\Phi}(\tau, t_0) \approx \mathbb{I}_6 + \begin{bmatrix} \mathbf{0} & (\tau - t_0)\mathbb{I}_3 \\ (\tau - t_0)\mathbf{M} & \mathbf{0} \end{bmatrix} \tag{15}$$

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

$$\mathbf{M} = \begin{bmatrix} \frac{3\mu r_x^2}{\|\mathbf{r}\|^5} - \frac{\mu}{\|\mathbf{r}\|^3} & \frac{3\mu r_x r_y}{\|\mathbf{r}\|^5} & \frac{3\mu r_x r_z}{\|\mathbf{r}\|^5} \\ \frac{3\mu r_y r_x}{\|\mathbf{r}\|^5} & \frac{3\mu r_y^2}{\|\mathbf{r}\|^5} - \frac{\mu}{\|\mathbf{r}\|^3} & \frac{3\mu r_y r_z}{\|\mathbf{r}\|^5} \\ \frac{3\mu r_z r_x}{\|\mathbf{r}\|^5} & \frac{3\mu r_z r_y}{\|\mathbf{r}\|^5} & \frac{3\mu r_z^2}{\|\mathbf{r}\|^5} - \frac{\mu}{\|\mathbf{r}\|^3} \end{bmatrix} \tag{16}$$

which can be written as

$$\mathbf{M} = \frac{3\mu}{\|\mathbf{r}\|^3} \begin{bmatrix} \frac{r_x^2}{\|\mathbf{r}\|^2} - \frac{1}{3} & \frac{r_x r_y}{\|\mathbf{r}\|^2} & \frac{r_x r_z}{\|\mathbf{r}\|^2} \\ \frac{r_y r_x}{\|\mathbf{r}\|^2} & \frac{r_y^2}{\|\mathbf{r}\|^2} - \frac{1}{3} & \frac{r_y r_z}{\|\mathbf{r}\|^2} \\ \frac{r_z r_x}{\|\mathbf{r}\|^2} & \frac{r_z r_y}{\|\mathbf{r}\|^2} & \frac{r_z^2}{\|\mathbf{r}\|^2} - \frac{1}{3} \end{bmatrix} \tag{17}$$

It is clear that without the $3\mu/\|\mathbf{r}\|^3$ factor, the $\mathbf{M}$ matrix satisfies $M_{i,j} < 1$ where $M_{i,j}$ is the component of $\mathbf{M}$ in row i and column j. Thus it is possible to define a time interval sufficiently small enough that $\mathbf{M}(\tau - t_0)$ can be approximated to have a negligible contribution. To quantify sufficiently small, if

$$(\tau - t_0) \ll \frac{\|\mathbf{r}\|^3}{3\mu} \tag{18}$$

then each of the terms in $\mathbf{M}(\tau - t_0)$ are very small and considered negligible. Thus, Eqn. (15) may be written as

$$\boldsymbol{\Phi}(\tau, t_0) \approx \mathbb{I}_6 + \begin{bmatrix} \mathbf{0} & (\tau - t_0)\mathbb{I}_3 \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{19}$$

and the dynamics essentially follow a straight line.

Define

$$\mathbf{H}_{\mathbf{r}}(\tau) = \frac{\partial \mathbf{h}(\mathbf{x}; \mathbf{k}, \tau)}{\partial \mathbf{r}(\tau)} \tag{20}$$

$$\mathbf{H}_{\dot{\mathbf{r}}}(\tau) = \frac{\partial \mathbf{h}(\mathbf{x}; \mathbf{k}, \tau)}{\partial \dot{\mathbf{r}}(\tau)} \tag{21}$$

$$\mathbf{H}^T\mathbf{H} = \frac{\partial \mathbf{h}(\mathbf{x}(\tau); \mathbf{k}, \tau)}{\partial \mathbf{x}(\tau)}^T \frac{\partial \mathbf{h}(\mathbf{x}(\tau); \mathbf{k}, \tau)}{\partial \mathbf{x}(\tau)} \tag{22}$$

$$= \begin{bmatrix} \mathbf{H}_{\mathbf{r}}^T\mathbf{H}_{\mathbf{r}} & \mathbf{H}_{\mathbf{r}}^T\mathbf{H}_{\dot{\mathbf{r}}} \\ \mathbf{H}_{\dot{\mathbf{r}}}^T\mathbf{H}_{\mathbf{r}} & \mathbf{H}_{\dot{\mathbf{r}}}^T\mathbf{H}_{\dot{\mathbf{r}}} \end{bmatrix} \tag{23}$$

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 $\boldsymbol{\zeta} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ which maps the state vector into a partitioned state vector containing the observable and unobservable states as follows

$$\mathbf{z} = \boldsymbol{\zeta}(\mathbf{x}) = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \tag{24}$$

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

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \frac{\partial \mathbf{h}}{\partial \mathbf{z}}\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{h}}{\partial \mathbf{z}_1} & \frac{\partial \mathbf{h}}{\partial \mathbf{z}_2} \end{bmatrix} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \tag{25}$$

where $\text{rank}[\frac{\partial \mathbf{z}}{\partial \mathbf{x}}] = n$ since $\boldsymbol{\zeta}$ is a bijective transformation. The unobservable states play no role in the measurements, $\partial\mathbf{h}/\partial\mathbf{z}_2 = \mathbf{0}$ leaving

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{h}}{\partial \mathbf{x}_d} & \mathbf{0} \end{bmatrix} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \tag{26}$$

which implies that $\text{rank}[\mathbf{H}^T\mathbf{H}] = q$, the dimension of the observable states. With the approximation for $\boldsymbol{\Phi}(\tau, t_0)$ 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, and introducing a change of variables $s = \tau - t_0$ then

$$\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t))) = \int_0^{\Delta t} \frac{\partial \mathbf{z}}{\partial \mathbf{x}}^T \begin{bmatrix} \mathbf{H}_\mathbf{r}^T \mathbf{H}_\mathbf{r} & \mathbf{H}_\mathbf{r}^T \mathbf{H}_\mathbf{r} s\mathbb{I}_3 + \mathbf{H}_\mathbf{r}^T \mathbf{H}_{\dot{\mathbf{r}}} \\ s\mathbb{I}_3 \mathbf{H}_\mathbf{r}^T \mathbf{H}_\mathbf{r} + (\mathbf{H}_\mathbf{r}^T \mathbf{H}_{\dot{\mathbf{r}}})^T & s^2 \mathbb{I}_3 \mathbf{H}_\mathbf{r}^T \mathbf{H}_\mathbf{r} + s\mathbb{I}_3 \mathbf{H}_\mathbf{r}^T \mathbf{H}_{\dot{\mathbf{r}}} + (\mathbf{H}_\mathbf{r}^T \mathbf{H}_{\dot{\mathbf{r}}})^T s\mathbb{I}_3 + \mathbf{H}_{\dot{\mathbf{r}}}^T \mathbf{H}_{\dot{\mathbf{r}}} \end{bmatrix} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \mathrm{d}s \tag{27}$$

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

$$\begin{cases} \text{Case 1}: & \mathbf{h}(\mathbf{x}; \mathbf{k}, t_1) \\ \text{Case 2}: & \mathbf{h}(\mathbf{x}; \mathbf{k}, \{t_1, \cdots, t_v\}), \quad (t_v - t_1) \ll \frac{\|\mathbf{r}\|^3}{3\mu} \\ \text{Case 3}: & \mathbf{h}(\mathbf{x}; \mathbf{k}, \{t_1, \cdots, t_v\}), \quad (t_v - t_1) \gg \frac{\|\mathbf{r}\|^3}{3\mu} \end{cases} \tag{28}$$

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. (27) 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 $(t_v - t_1) \gg \|\mathbf{r}\|^3/3\mu$. 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 inverval. Case 3 details a measurement function dependent upon a sufficiently long time period where the assumption of Eqn. (19) is no longer valid.

For both Case 1 and Case 2, since $\Delta t \ll \|\mathbf{r}\|^3/3\mu$ is small, thus it is reasonable to assume that any higher order powers of Δt in $\mathbf{\Phi}(t_v, t_0)$ may be considered negligible. With this in mind, the integration of Eqn. (27) gives a simple result after neglecting higher order terms of Δt.

$$\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t))) \approx \Delta t \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^T & \mathbf{H}_3 \end{bmatrix} \tag{29}$$

The matrix $\mathbf{P}(t_f, t_0, (\mathbf{x}_r(t), \mathbf{k}_r(t)))$ has 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, rank$[\mathbf{P}(t_f, t_0, \mathbf{x}(t))] = d$ as long as $\Delta t \ll \|\mathbf{r}\|^3/3\mu$. 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.

It can be shown that this Lemma is generalizable to any Newtonian system where

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{0} & \mathbb{I}_3 \\ \mathbf{M}_g & \mathbf{0} \end{bmatrix} \tag{30}$$

and $\mathbf{M}_g$ is a general matrix capturing the linearized dynamics of the system. For this general case, there exists some constant C such that $C\mathbf{M}_g \approx \mathbf{0}$. When the time period satisfies $t_f - t_0 \ll C$ then the exact proof shown above for Keplerian dynamics holds and this general Newtonian system is unobservable over that time period. □

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 sections build upon the existence of the admissible region and show its definition and properties.

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

$$\mathbf{y} = \mathbf{h}(\mathbf{x}; \mathbf{k}, t) \tag{31}$$

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 $u + d = n$ [13]. This means that

$$\mathbf{y} = \mathbf{h}(\mathbf{x}_d; \mathbf{k}, t) \tag{32}$$

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 $\kappa_i : \mathbb{R}^u \times \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R} \to \mathbb{R}$. The admissible region for the ith hypothesized constraint $\boldsymbol{\kappa}_i(\cdot)$ is then defined as

$$\mathcal{AR}_i := \{\mathbf{x}_u \in \mathbb{R}^u \mid \kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) \leq 0\} \tag{33}$$

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

$$\mathcal{AR} = \bigcap_{i=1}^{c} \mathcal{AR}_i \tag{34}$$

where $\mathcal{AR}$ must be a compact set [22]. The requirement that $\mathcal{AR}$ be compact ensures the assumed uniform distribution has non-zero probability. Thus, each state $\mathbf{x} \in \mathcal{AR}$ can be assigned a non-zero uniform probability. This is by definition consistent with $\mathcal{AR}$ being statistically represented as an uninformative prior.

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

$$\mathbb{P}[\mathbf{x}_u \in \mathcal{AR}_i] = \mathbb{P}[\kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) \leq 0] \tag{35}$$

Without any additional information, the inequality defining $\mathcal{AR}_i$ in Equation (33) is a binary constraint and $\mathbb{P}[\mathbf{x}_u \in \mathcal{AR}_i] \in \{0, 1\}$ 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

$$m_i(\mathbf{x}_u) = \begin{cases} 1, & \kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) \leq 0 \\ 0, & \kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) > 0 \end{cases} \tag{36}$$

Thus $\mathbb{P}[\mathbf{x}_u \in \mathcal{AR}_i] = m_i(\mathbf{x}_u)$, and the prior distribution for a particular constraint hypothesis can then be defined as [16]

$$f_{i,\mathbf{x}_u}(\mathbf{x}_u) = \frac{m_i(\mathbf{x}_u)}{\int_{\mathcal{AR}_i} \mathrm{d}\mathbf{x}_\mathrm{u}} \tag{37}$$

Eqn. (37) results in a uniform distribution, which is demonstrated in [11]. Applying the chain rule of probabilities, the general joint probability function over all k constraints can be written as

$$\begin{aligned} f_{\mathbf{x}_u}(\mathbf{x}_u) &= \frac{\mathbb{P}[\mathbf{x}_u \in \mathcal{AR}]}{\int_{\mathcal{AR}} \mathrm{d}\mathbf{x}_\mathrm{u}} \\ &= \frac{1}{\int_{\mathcal{AR}} \mathrm{d}\mathbf{x}_\mathrm{u}} \prod_{k=1}^{c} \mathbb{P}\left[\mathbf{x}_u \in \mathcal{AR}_k \,\middle|\, \mathbf{x}_u \in \bigcap_{j=1}^{k-1} \mathcal{AR}_j\right] \end{aligned} \tag{38}$$

where the bracketed term gives the probability that k^{th} constraint is satisfied given that each of the $k-1$ previous constraints are satisfied [17]. If the constraints κ_i are assumed to be independent, then by Bayes' rules the conditional probability terms evaluate to 1 and Eqn. (38) simplifies to

$$f_{\mathbf{x}_u}(\mathbf{x}_u) = \frac{\prod_{k=1}^{c} \mathbb{P}\left[\mathbf{x}_u \in \mathcal{AR}_k\right]}{\int_{\mathcal{AR}} \mathrm{d}\mathbf{x}_\mathrm{u}} \tag{39}$$

$$= \frac{\prod_{k=1}^{c} m_k(\mathbf{x}_u)}{\int_{\mathcal{AR}} \mathrm{d}\mathbf{x}_\mathrm{u}} \tag{40}$$

By this formulation, every $\mathbf{x}_u \in \mathcal{AR}$ 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 $\mathcal{AR}$ 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 [23]. 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 consistent with Jeffreys' prior, not as a PDF. The reason for this distinction will become clear in the next section.

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 spate in which the admissible region constraints are formed. Following the general probability transformation approach, a transformation $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ can be defined. This transformation must also be able to be partitioned into $\mathbf{g}_u : \mathbb{R}^u \to \mathbb{R}^u$ and $\mathbf{g}_d : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\tilde{\mathbf{x}}_u = \mathbf{g}_u(\mathbf{x}_u; \mathbf{y}, \mathbf{k}, t) \tag{41}$$

$$\tilde{\mathbf{x}}_d = \mathbf{g}_d(\mathbf{x}_d; \mathbf{y}, \mathbf{k}, t) \tag{42}$$

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 $\boldsymbol{x}_u$ may be transformed to state space $\tilde{\boldsymbol{x}}_u = \boldsymbol{g}_u(\boldsymbol{x}_u;\cdot)$ only if there exist some $\tilde{\boldsymbol{x}}_d = \bar{\boldsymbol{g}}_d(\boldsymbol{y};\cdot)$, $\bar{\boldsymbol{g}}_d : \mathbb{R}^m \to \mathbb{R}^d$ such that $\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}_d;\boldsymbol{k},t) = \tilde{\boldsymbol{h}}(\tilde{\boldsymbol{x}}_d;\boldsymbol{k},t)$, $\tilde{\boldsymbol{h}} : \mathbb{R}^d \times \mathbb{R}^l \times \mathbb{R} \to \mathbb{R}^m$.*

Proof. The undetermined states $\mathbf{x}_u$ are independent of the determined states $\mathbf{x}_d$ as defined in [13]. 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 $\bar{\mathbf{g}}_d = \mathbf{g}_d \circ \mathbf{h}^{-1} : \mathbb{R}^m \to \mathbb{R}^d$ 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 $\mathbf{x}_u$,

$$\kappa_i(\mathbf{x}_u, \mathbf{y}, \mathbf{k}, t) = \tilde{\kappa}_i(\mathbf{g}_u(\mathbf{x}_u;\cdot), \mathbf{y}, \mathbf{k}, t) \tag{43}$$

$$\tilde{\kappa}_i(\tilde{\mathbf{x}}_u, \mathbf{y}, \mathbf{k}, t) = \kappa_i(\mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u;\cdot), \mathbf{y}, \mathbf{k}, t) \tag{44}$$

Eqns. (43) and (44) then imply that $\mathbb{P}[\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{AR}}_i] = \mathbb{P}[\mathbf{x}_u \in \mathcal{AR}_i]$ and $m_i(\mathbf{x}_u) = m_i(\tilde{\mathbf{x}}_u)$ where,

$$\widetilde{\mathcal{AR}}_i := \{\tilde{\mathbf{x}}_u \in \mathbb{R}^u \mid \tilde{\kappa}_i(\tilde{\mathbf{x}}_u, \mathbf{y}, \mathbf{k}, t) \leq 0\} \tag{45}$$

and

$$\tilde{m}_i(\tilde{\mathbf{x}}_u) = \begin{cases} 1, & \tilde{\kappa}_i(\tilde{\mathbf{x}}_u, \mathbf{y}, \mathbf{k}, t) \leq 0 \\ 0, & \tilde{\kappa}_i(\tilde{\mathbf{x}}_u, \mathbf{y}, \mathbf{k}, t) > 0 \end{cases} \tag{46}$$

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

$$f_{\tilde{\mathbf{X}}_u}(\tilde{\mathbf{x}}_u) = \frac{1}{\int_{\widetilde{\mathcal{AR}}} d\tilde{\mathbf{x}}_u} \prod_{k=1}^{c} \mathbb{P}\left[\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{AR}}_k \;\middle|\; \tilde{\mathbf{x}}_u \in \bigcap_{j=1}^{k-1} \widetilde{\mathcal{AR}}_j\right] \tag{47}$$

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

$$f_{\tilde{\mathbf{X}}_u}(\tilde{\mathbf{x}}_u) = \frac{\prod_{k=1}^{c} \tilde{m}_k(\tilde{\mathbf{x}}_u)}{\int_{\widetilde{\mathcal{AR}}} d\tilde{\mathbf{x}}_u} \tag{48}$$

A general nonlinear transformation of a uniform PDF must yield a non-uniform PDF according to Eqn. (3). The uniform PDF of an admissible region is a statistical representation of the fact that each state $\mathbf{x}_u \in \mathcal{AR}$ 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, the necessary relationship between $f_{\mathbf{X}_u}(\mathbf{x}_u)$ and $f_{\tilde{\mathbf{X}}_u}(\tilde{\mathbf{x}}_u)$ is given by Theorem 2, leading to a contradiction which identifies the issue with classifying the admissible region as a PDF.

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

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 $\mathbf{x} \in \mathcal{AR}$. 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) as follows

$$f_{\tilde{\mathbf{X}}_u}(\tilde{\mathbf{x}}_u) = \frac{\prod_{k=1}^{c} m_k\left(\mathbf{g}^{-1}(\tilde{\mathbf{x}}_u)\right)}{\int_{\mathcal{AR}} d\mathbf{x}_u} \text{abs}\left(\left|\frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u}\right|\right) \tag{49}$$

Each of the terms in Eqn. (49) have been defined thus far except for the Jacobian term $\left|\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}}_u)/\partial \tilde{\mathbf{x}}_u\right|$. Rearranging Eqn. (49), by substituting the $\tilde{\mathbf{x}}_u$ PDF on the left hand side and multiplying by the denominator of the right hand side,

$$\frac{\prod_{k=1}^{c} \tilde{m}_k(\tilde{\mathbf{x}}_u)}{\int_{\widetilde{\mathcal{AR}}} d\tilde{\mathbf{x}}_u} \int_{\mathcal{AR}} d\mathbf{x}_u = \prod_{k=1}^{c} m_k\left(\mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)\right) \text{abs}\left(\left|\frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u}\right|\right) \tag{50}$$

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

$$\frac{\int_{\mathcal{A}} \mathrm{d}\mathbf{x}_u}{\int_{\widetilde{\mathcal{A}}} \mathrm{d}\tilde{\mathbf{x}}_u} = \zeta = \mathrm{abs}\left(\left|\frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u}\right|\right) \tag{51}$$

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

$$\frac{\int_{\mathcal{A}} \mathrm{d}\mathbf{x}_u}{\int_{\widetilde{\mathcal{A}}} \mathrm{d}\tilde{\mathbf{x}}_u} \neq \zeta \tag{52}$$

which then implies $\mathbb{P}[\mathbf{x}_u \in \mathcal{A}_i] \neq \mathbb{P}[\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{A}_i}]$ for Eqn. (50) to hold. But this is a contradiction since, by definition, the admissible region approach gives that $\mathbb{P}[\mathbf{x}_u \in \mathcal{A}_i] = \mathbb{P}[\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{A}_i}]$ regardless of the transformation. □

Theorem 2 imposes a geometric constraint on the transformation **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), but as shown by Theorem 2, under most practical transformations, an admissible region fails to satisfy Eqn. (3).

The result of Theorem 2 is directly related to Jeffreys' prior [24]. A Jeffreys' prior is a non-informative prior which satisfies

$$f(\mathbf{x}) \propto \sqrt{\det \mathcal{I}(\mathbf{x})} \tag{53}$$

where $f()$ denotes the prior and $\mathcal{I}(\mathbf{x})$ is the Fisher information matrix [25]. If $\mathcal{I}(\mathbf{x})$ is singular then Jeffreys' prior does not exist [26]. 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. (53) gives that a Jeffreys prior is invariant to a reparamaterization 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

$$f(\tilde{\mathbf{x}}) = f(\mathbf{x}) \left|\frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}}\right| \tag{54}$$

It can likewise be shown that

$$\sqrt{\det \mathcal{I}(\tilde{\mathbf{x}})} = \sqrt{\det \mathcal{I}(\mathbf{x})} \left|\frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}}\right| \tag{55}$$

thus for Jeffrey's prior to hold, Eqn. (53) may be rewritten as

$$f(\tilde{\mathbf{x}}) \left|\frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}}\right|^{-1} \propto \sqrt{\det \mathcal{I}(\tilde{\mathbf{x}})} \left|\frac{\partial \mathbf{x}}{\partial \tilde{\mathbf{x}}}\right|^{-1} \tag{56}$$

The proportionality of Eqn. (53) requires that $|\partial \mathbf{x}/\partial \tilde{\mathbf{x}}|^{-1} = \left|\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)/\partial \tilde{\mathbf{x}}_u\right|$ 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

$$f_{\tilde{\mathbf{X}}_u}(\tilde{\mathbf{x}}_u) \left|\frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u}\right| \propto \sqrt{\det \mathcal{I}(\tilde{\mathbf{x}}_u)} \left|\frac{\partial \mathbf{g}_u^{-1}(\tilde{\mathbf{x}}_u)}{\partial \tilde{\mathbf{x}}_u}\right| \tag{57}$$

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 **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

$$f_{\tilde{\mathbf{X}}_u}(\tilde{\mathbf{x}}_u) = \zeta \frac{\prod_{k=1}^{c} m_k(\mathbf{x}_u)}{\int_{\mathcal{A}} \mathrm{d}\mathbf{x}_u} \tag{58}$$

Eqn. (58) 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 [11]. 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 an admissible region prior misrepresenting the true distribution.

5. *Priors Using Uncertain Admissible Region Constraints*

While Eqn. (58) 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 $\mathcal{A}$ is shown in [13]. The approximate analytical probability for a given admissible region is given by

$$\mathbb{P}[(\mathbf{x}_u \in \mathcal{A}_i)] = m_i(\mathbf{x}_u) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{||\mathbf{x}_u - \mathbf{x}_{u,\mathcal{B}_\perp,i}||}{\sqrt{2\mathrm{tr}\mathbf{P}_{\mathbf{x}_{u,\mathcal{B}_\perp,i}}}}\right)\right] \tag{59}$$

which updates the piecewise membership function given by Eqn. (36) to a continuous membership function by including uncertainty effects. These uncertainties are quantified as the covariance matrix $\mathbf{P}_\mathbf{z}$ 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 $\mathcal{A}_i$ 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. (59) into Eqn. (40) then gives the non-uniform PDF.

Corollary 2 (Systematic Uncertainty in Admissible Regions)**.** *If the combined measurements, parameters, and time covariance matrix $\boldsymbol{P}_z$ are known then transformation of the non-uniform admissible region probability is given by*

$$\mathbb{P}[(\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{A}_i})] = \tilde{m}_i(\tilde{\mathbf{x}}_u) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{||\boldsymbol{g}_u(\boldsymbol{x}_u;\cdot) - \boldsymbol{g}_u(\boldsymbol{x}_{u,\mathcal{B}_\perp,i};\cdot)||}{\sqrt{2\mathrm{tr}\boldsymbol{P}_{\tilde{\boldsymbol{x}}_{u,\mathcal{B}_\perp,i}}}}\right)\right] \tag{60}$$

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

Proof. Given the previous transformation $\tilde{\mathbf{x}}_u = \mathbf{g}_u(\mathbf{x}_u;\cdot)$, Eqn. (59) can be derived for $\tilde{\mathbf{x}}_u$. The simplified Taylor series expansion from Eqn. (17) in [13] now becomes

$$-\frac{\partial \kappa_i}{\partial \tilde{\mathbf{x}}_u}\frac{\partial \tilde{\mathbf{x}}_u}{\partial \mathbf{x}_u}\delta \mathbf{X}_u = \frac{\partial \kappa_i}{\partial \mathbf{z}}\delta \mathbf{Z} \tag{61}$$

Carrying the notation defined in [13], a new perpendicular vector $\tilde{\mathbf{p}}$ is defined as

$$\tilde{\mathbf{p}} = \frac{\partial \kappa_i}{\partial \tilde{\mathbf{x}}_u}\frac{\partial \mathbf{g}_u^{-1}(\mathbf{x}_u;\cdot)}{\partial \mathbf{x}_u}\bigg|_{\mathbf{x}_u} \tag{62}$$

The rest of the derivation can be carried out as specified in [13] by replacing $\mathbf{p}$ with $\tilde{\mathbf{p}}$ resulting in

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

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

$$\mathbf{P}_{\tilde{\mathbf{x}}_u} = \widetilde{\mathbf{M}}\mathbf{P}_{\mathbf{z}}\widetilde{\mathbf{M}}^T \tag{63}$$

Eqn. (60) is obtained by substituting $\mathbf{P}_{\tilde{\mathbf{x}}_u}$ and $\mathbf{g}(\mathbf{x}_u)$ into Eqn. (59). □

Eqn. (60) defines the approximate analytical probability distribution function for an admissible region $\mathcal{AR}_i$ in the $\tilde{\mathbf{x}}_u$ space. Alternatively, from Eqns. (43) and (44)

$$\mathbb{P}[(\mathbf{x}_u \in \mathcal{AR}_i)] = \mathbb{P}[(\tilde{\mathbf{x}}_u \in \widetilde{\mathcal{AR}_i})] \tag{64}$$

$$= \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\|\mathbf{g}_{\mathrm{u}}(\mathbf{x}_{\mathrm{u}};\cdot) - \mathbf{g}_{\mathrm{u}}(\mathbf{x}_{\mathrm{u}};\cdot)_{\mathrm{u},\mathcal{B}_\perp}\|}{\sqrt{2\mathrm{tr}\mathbf{P}_{\widetilde{\mathbf{x}}_{\mathrm{u},\mathcal{B}_\perp}}}}\right)\right] \tag{65}$$

$$\approx \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\|\mathbf{x}_{\mathrm{u}} - \mathbf{x}_{\mathrm{u},\mathcal{B}_\perp}\|}{\sqrt{2\mathrm{tr}\mathbf{P}_{\mathbf{x}_{\mathrm{u},\mathcal{B}_\perp}}}}\right)\right] \tag{66}$$

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 $\mathbb{P}[\mathbf{x} \in \mathcal{AR}_i]$ or vice-versa.

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 $\mathcal{AR}$. However, if an additional measurement can be taken at a time t such that each state $\mathbf{x}_u \in \mathcal{AR}$ 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) to transform probabilities between state spaces. As such, it is of interest to determine when the states $\mathbf{x}_u \in \mathcal{AR}$ become locally observable.

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

Proof. The admissible region $\mathcal{AR}$ is, as defined, a subset of the unobservable state space where each state $\mathbf{x}_u \in \mathcal{AR}$ 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 $\mathbf{x}_u \in \mathcal{AR}$ 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 $\mathbf{h}_o : \mathbb{R}^n \to \mathbb{R}^m$ 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 $\tilde{\mathbf{h}}_o : \mathbb{R}^n \to \mathbb{R}^m$ 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). □

The main result of Corollary 3 is that the PDF associated with a given $\mathcal{AR}$ may not generally be transformed until it is observable. 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 $\mathcal{AR}$ becomes locally observable. Furthermore, by Lemma 1, if every $\mathbf{x}_u \in \mathcal{AR}$ is locally observable, then the region is necessarily not an admissible region.

C. Additional Transformations

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

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 $\mathbf{T}_u \in \mathbb{R}^{n\times n}, \mathrm{rank}\,\mathbf{T}_\mathrm{u} = \mathrm{n}$ that can be defined $\forall\,\mathbf{x}_u \in \mathcal{AR}$ 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.

2. *Sigma Point Transformations*

An additional application of the general probability transformation comes from sigma point transformations and filters [27]. 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 [28]. 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

$$\mathbf{x} + \delta\mathbf{x} = \mathbf{g}^{-1}(\tilde{\mathbf{x}}) + \frac{\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}}\delta\tilde{\mathbf{x}} \tag{67}$$

$$\mathbf{x} + \delta\mathbf{x} = \mathbf{x} + \frac{\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}}\delta\tilde{\mathbf{x}} \tag{68}$$

$$\delta\mathbf{x} = \frac{\partial \mathbf{g}^{-1}(\tilde{\mathbf{x}})}{\partial \tilde{\mathbf{x}}}\delta\tilde{\mathbf{x}} \tag{69}$$

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. *Transformations Over Time*

General probability transformations also apply to transformations through time as shown by Park and Scheeres [29]. 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

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{70}$$

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

$$\mathbf{x}(t) = \boldsymbol{\phi}(t; \mathbf{x}_0, t_0) \tag{71}$$

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

$$\frac{\mathrm{d}\boldsymbol{\phi}}{\mathrm{dt}} = \mathbf{f}(\boldsymbol{\phi}(t; \mathbf{x}_0, t_0), t) \tag{72}$$

$$\boldsymbol{\phi}(t_0; \mathbf{x}_0, t_0) = \mathbf{x}_0 \tag{73}$$

In the case of time transformations, the function $\boldsymbol{\phi}$ 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 [30]

$$\frac{\partial f_{\mathbf{x}}(\mathbf{x},t)}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial \mathbf{x}_i} \left(f_{\mathbf{x}}(\mathbf{x},t)\mathbf{f}_i(\mathbf{x},t) \right) \tag{74}$$

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 [29] [31].

$$f(\boldsymbol{\phi}(t;\mathbf{x}_0,t_0),t) = f(x_0,t_0)\left|\frac{\partial \mathbf{x}}{\partial \mathbf{x}_0}\right|^{-1} \tag{75}$$

which is the exact form given in Eqn. (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 [29]. For a Hamiltonian system Eqn. (75) 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 $\zeta f(\mathbf{x}_0,t_0) = f(\mathbf{x}_0,t_0)$.

D. Discussion

The results presented in this paper show that in general, transformations of admissible region probabilities are only possible under strict conditions outlined by Theorem 2. 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. 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 posteori PDF with appropriate $\tilde{\mathbf{x}} = \mathbf{g}(\mathbf{x})$ as desired. As such, for any filter to be properly instantiated, it must remain in the state space of the original admissible region prior formulation unless either Theorem 2 or Corollary 3 is satisfied.

III. 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 [13], the measurement vector is given by,

$$\mathbf{y} = \begin{bmatrix}\alpha & \delta & \dot{\alpha} & \dot{\delta}\end{bmatrix}^T \tag{76}$$

with the object state vector,

$$\mathbf{x} = \begin{bmatrix}\mathbf{r} & \mathbf{v}\end{bmatrix} \tag{77}$$

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,

$$\tilde{\mathbf{x}} = \begin{bmatrix}\alpha & \delta & \dot{\alpha} & \dot{\delta} & \rho & \dot{\rho}\end{bmatrix}^T \tag{78}$$

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

$$\mathbf{r} = \begin{bmatrix}-0.9281\\-0.0489\\0.6167\end{bmatrix}\text{DU} \qquad \mathbf{v} = \begin{bmatrix}-0.5171\\0.1292\\-0.7662\end{bmatrix}\text{DU/TU} \tag{79}$$

where 1 DU = 6378 km and 1 DU/TU = 7.90538 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

$$\mathbf{x}_d = \begin{bmatrix}-3.0337 \text{ rad} & -0.0538 \text{ rad} & -0.1003 \text{ rad/TU} & -0.4482 \text{ rad/TU}\end{bmatrix} \tag{80}$$

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 km $\le a \le$ 50000 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 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 1.

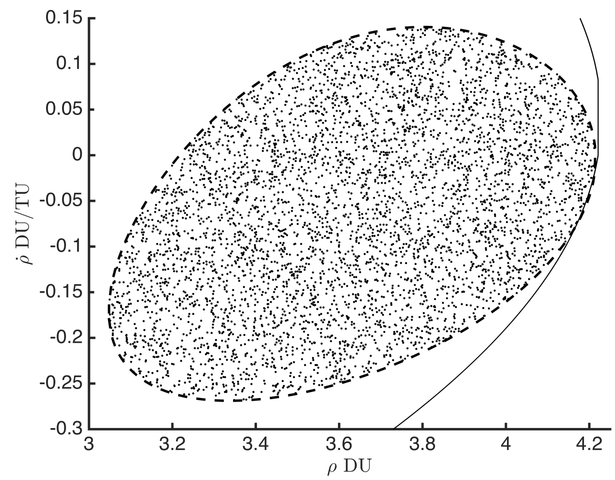

Figure 1. A set of 5000 points sampled uniformly from the admissible region.

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) must be applied. The transformation from $\tilde{\mathbf{x}}$ to $\mathbf{x}$ is given by,

$$\mathbf{r} = \mathbf{o} + \rho\hat{\mathbf{l}} \tag{81}$$

$$\mathbf{v} = \dot{\mathbf{o}} + \dot{\rho}\hat{\mathbf{l}} + \rho\dot{\alpha}\hat{\mathbf{l}}_\alpha + \rho\dot{\delta}\hat{\mathbf{l}}_\delta \tag{82}$$

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, $\dot{\rho}$, 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 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.

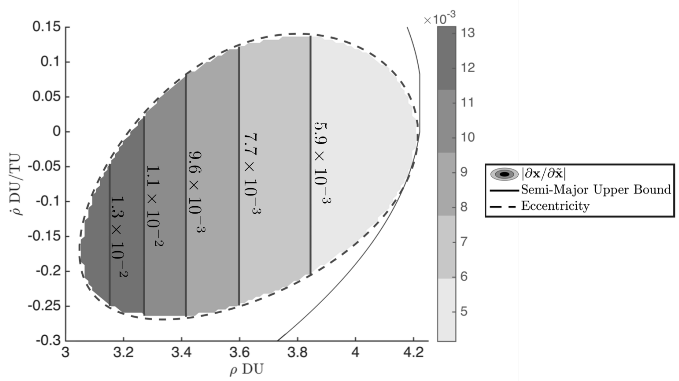

Figure 2. Values of $|\partial \mathbf{x}/\partial \tilde{\mathbf{x}}|$ evaluated for each particle $\mathbf{x}(t)$

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). 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 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 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). Figure 5 shows the updated PDF after the second observation is made and can equivalently be expressed in cartesian coordinates by Eqn. (3).

To demonstrate the importance of Theorem 2 and Corollary 3, consider the process shown in Figure 3 by which the cartesian PDF for these observations can be determined. The original admissible region in ρ and $\dot{\rho}$ is represented by $\widetilde{\mathcal{AR}}_{t_0}$ and after the second observation is made the PDF over the particles is given by $f_{\tilde{\mathbf{X}}_u}(\tilde{\mathbf{x}}_u)$. The admissible region given by $\mathcal{AR}$ represents the transformation of $\widetilde{\mathcal{AR}}_{t_0}$ 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 $f_{\tilde{\mathbf{X}}_u}(\tilde{\mathbf{x}}_u)$ 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 6 shows the resulting PDF for the unobservable and observable transformations outlined in Figure 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 noticable bias in the particle filter. Figures 7 and 8 illustrate this dependence on time. If the second observation is one hour after the initial observation, as seen in Figure 8, there is little difference between the PDFs because the particle filter eliminates the bias introduced by the incorrect initial weighting of the transformed particles. However, if the second observation is only 10 minutes after the initial observation, as seen in Figure 7, there is a considerable bias in the PDF for the unobservable transformation. 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. The general exception to this are linear transformations which always satisfy the requirements of Theorem 2.

IV. Conclusions

The general theory of probability transformations is presented and applied directly to the admissible region problem. It is found that general probability transformations are invalid for admissible regions, 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 shows that the statistical representation of the admissible region is consistent with Jeffreys' prior and inconsistent with a PDF. Furthermore it is shown once the system becomes observable with new measurements the admissible region prior becomes a true PDF. It is also shown that probability transformations of admissible regions can also take into account measurement, parameter, and observer uncertainties.

V. Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1148903. The authors would also like to acknowledge Islam Hussein, Christopher Roscoe, Matthew Wilkins, and Paul Schumacher for their discussions regarding this work.

References

[1] S. A. Hildreth and A. Arnold, "Threats to US National Security Interests in Space: Orbital Debris Mitigation and Removal," DTIC Document, 2014.

[2] J.-C. Liou, "Modeling the Large and Small Orbital Debris Populations for Environment Remediation," tech. rep., NASA Orbital Debris Program Office, June 2014.

[3] L. G. Taff, "On initial orbit determination," *Astronomical Journal*, Vol. 89, Sept. 1984, pp. 1426–1428, doi:10.1086/113644.

[4] 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, 2004, pp. 57–85, doi:10.1007/s10569-004-6593-5.

[5] 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, 2005, pp. 350 – 374, doi:10.1016/j.icarus.2005.07.004.

[6] 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, 2009, pp. 194–209, doi:10.2514/1.36398.

[7] 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, 2010, pp. 157–182, doi:10.1007/s10569-009-9245-y.

[8] 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, 2012, pp. 2628–2637, doi:10.1109/TAES.2012.6237613.

[9] J. Siminski, O. Montenbruck, H. Fiedler, and T. Schildknecht, "Short-arc tracklet association for geostationary objects," *Advances in Space Research*, Vol. 53, No. 8, 2014, pp. 1184 – 1194, doi:10.1016/j.asr.2014.01.017.

[10] K. Fujimoto and K. T. Alfriend, "Optical Short-Arc Association Hypothesis Gating via Angle-Rate Information," *Journal of Guidance, Control, and Dynamics*, 2015/06/08 2015, pp. 1–12, 10.2514/1.G000927.

[11] 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, 2014, pp. 295 – 308, doi:10.1016/j.asr.2013.11.021.

[12] K. DeMars and M. K. Jah, "Probabilistic Initial Orbit Determination Using Gaussian Mixture Models," *Journal of Guidance Control, and Dynamics*, Vol. 36, No. 5, 2013, pp. 1324–1335, doi:10.2514/1.59844.

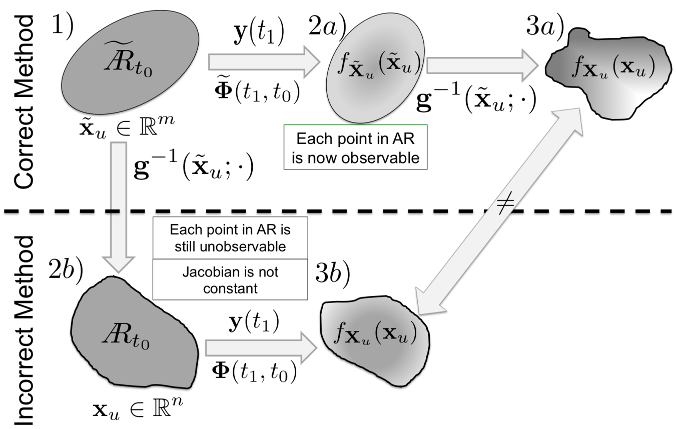

Figure 3. Outline of the two approaches for generating the PDF in cartesian coordinates

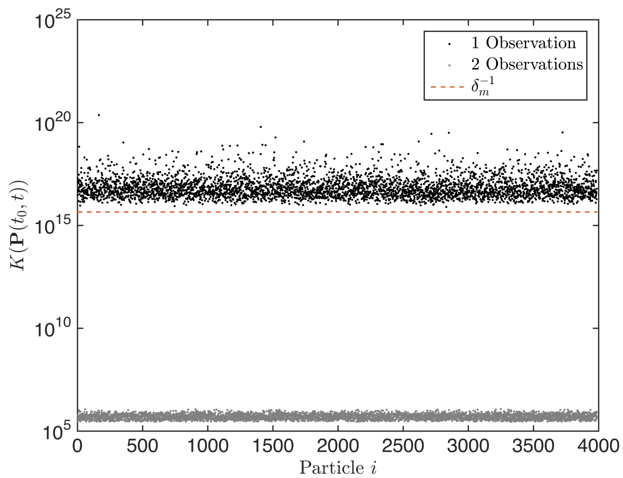

Figure 4. Condition number of $\mathbf{P}(t, t_0, \mathbf{x}(t))$ computed for each particle $\mathbf{x}_u(t)$

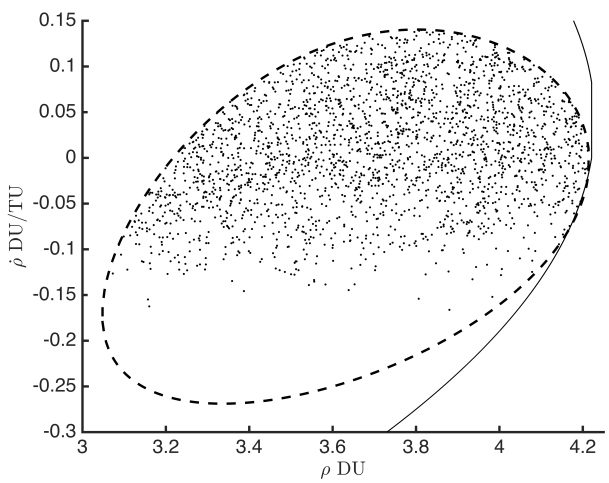

Figure 5. Admissible Region prior expressed in ρ and $\dot{\rho}$ 30 minutes after the initial observation

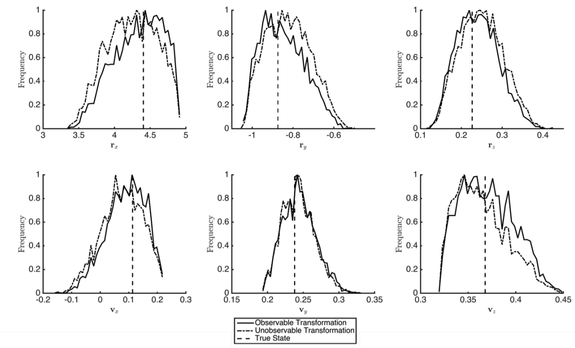

Figure 6. Difference between the cartesian PDFs, second observation is 30 minutes after the initial observation.

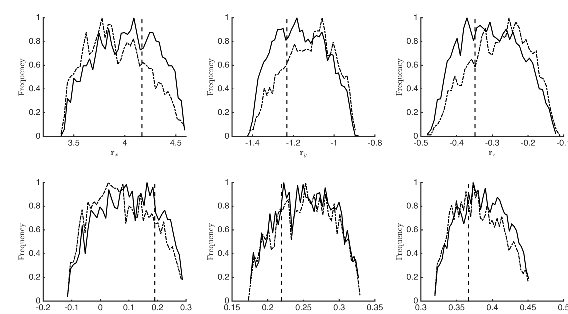

Figure 7. Difference between the cartesian PDFs, second observation is 10 minutes after the initial observation.

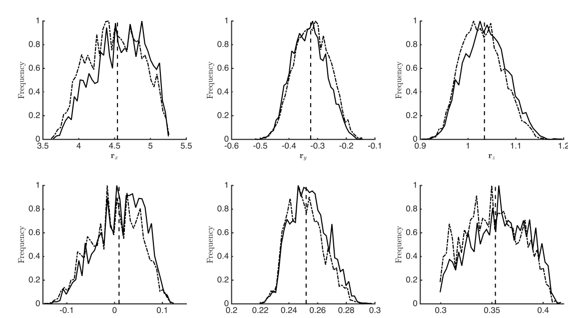

Figure 8. Difference between the cartesian PDFs, second observation is 60 minutes after the initial observation.

[13] J. L. Worthy and M. J. Holzinger, "Incorporating Uncertainty in Admissible Regions for Uncorrelated Detections," *Journal of Guidance, Control, and Dynamics*, Vol. 38, 2015/11/24 2015, pp. 1673–1689, 10.2514/1.G000890.

[14] I. Hussein, C. Roscoe, M. Wilkins, and P. Schumacher, "Probabilistic Admissible Region for Short-Arc Angles-Only Observations," *Advanced Maui Optical and Space Surveillance Technologies Conference*, Vol. 1, 2014, p. 76.

[15] B. Flury, *A First Course in Multivariate Statistics*. Springer Texts in Statistics, Springer, 1997.

[16] D. Montgomery and G. Runger, *Applied Statistics and Probability for Engineers*. John Wiley & Sons, 5th ed., 2010.

[17] A. Chorin and O. Hald, *Stochastic Tools in Mathematics And Science*. Stochastic Tools in Mathematics and Science, Springer, 2006.

[18] R. Hogg, J. McKean, and A. Craig, *Introduction to Mathematical Statistics*. Pearson, 2013.

[19] W. Brogan, "Modern Control Theory," 1991, p. 382.

[20] R. Hermann and A. J. Krener, "Nonlinear controllability and observability," *Automatic Control, IEEE Transactions on*, Vol. 22, Oct 1977, pp. 728–740, 10.1109/TAC.1977.1101601.

[21] M. Hwang and J. Seinfeld, "Observability of nonlinear systems," *Journal of Optimization Theory and Applications*, Vol. 10, No. 2, 1972, pp. 67–77, 10.1007/BF00934972.

[22] G. Tommei, A. Milani, and A. Rossi, "Orbit determination of space debris: admissible regions," *Celestial Mechanics and Dynamical Astronomy*, Vol. 97, No. 4, 2007, pp. 289–304, doi:10.1007/s10569-007-9065-x.

[23] 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, 2012, pp. 208–221, doi: 10.2514/1.53126.

[24] 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, 1946, pp. pp. 453–461.

[25] G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*, Vol. 40. John Wiley & Sons, 2011.

[26] C. Jauffret, "Observability and fisher information matrix in nonlinear regression," *Aerospace and Electronic Systems, IEEE Transactions on*, Vol. 43, April 2007, pp. 756–759, 10.1109/TAES.2007.4285368.

[27] D.-J. Lee and K. T. Alfriend, "Sigma Point Filtering for Sequential Orbit Estimation and Prediction," *Journal of Spacecraft and Rockets*, Vol. 44, 2015/05/11 2007, pp. 388–398, 10.2514/1.20702.

[28] S. Julier, "The scaled unscented transformation," *American Control Conference, 2002. Proceedings of the 2002*, Vol. 6, 2002, pp. 4555–4559 vol.6, 10.1109/ACC.2002.1025369.

[29] 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, 2015/03/31 2006, pp. 1367–1375, 10.2514/1.20177.

[30] P. Maybeck, *Stochastic Models, Estimation, and Control*. Mathematics in Science and Engineering, Elsevier Science, 1982.

[31] K. Fujimoto and D. J. Scheeres, "Tractable Expressions for Nonlinearly Propagated Uncertainties," *Journal of Guidance, Control, and Dynamics*, Vol. 38, 2015/05/28 2015, pp. 1146–1151, 10.2514/1.G000795.