AAS 23-110

DATA-DRIVEN CATEGORIZATION OF SPACECRAFT MOTION WITH UNCERTAINTY IN THE EARTH-MOON SYSTEM

Renee L. Spear* and Natasha Bosanac†

Clustering is used to categorize possible spacecraft trajectories resulting from uncertain state estimates in the Earth-Moon circular restricted three-body problem. Initial conditions are sampled within a region of uncertainty near a state estimate and propagated numerically. The resulting trajectories are described using finite-dimensional feature vectors that capture their spatiotemporal variations. These feature vectors are then grouped using hierarchical and density-based clustering. The result is a summary of the array of distinct geometries exhibited by potential future spacecraft motions. This approach is used to examine motions associated with uncertain state estimates near locations along a variety of periodic orbits.

INTRODUCTION

A challenge that is emerging in space domain awareness and space traffic management is predicting and describing the possible future motions of an object in the presence of uncertainty within cislunar space. In sensitive regions, small variations in the estimated state may result in a variety of possible paths with distinct geometries and itineraries over sufficiently long time intervals. The level of uncertainty associated with a state estimate may also increase when off-nominal conditions are present, the objects are difficult to observe and track over time, and when characterizing an unknown object.[1] Furthermore, the possible motions associated with an initial state estimate and uncertainty can vary in geometry and characteristics in distinct regions of the Earth-Moon system or as model parameters evolve. Recent research in the astrodynamics community to study the impact of uncertainty on trajectories in cislunar space includes, for example, studying downstream uncertainty dispersion and uncertainty propagation[2–5] as well as using Gaussian mixture model estimation filters.[6] Due to the absence of analytical solutions for trajectories in multi-body systems, many works tend to rely on numerically generating a variety of trajectories. In these cases, it may be valuable to extract a clear, digestible summary of the distinct types of numerically-generated trajectories that a spacecraft may potentially follow given an uncertain state estimate.

Large data sets may be complex and time-consuming to manually analyze or require automated analysis when reducing the dependency on a human-in-the-loop. In these cases, data mining techniques may be employed to extract patterns and other information.[7] One technique, clustering, enables the extraction of a set of clusters that group similar data and separate dissimilar data.[7] The result is a data-driven summary of a large data set.

*Graduate Researcher, Colorado Center for Astrodynamics Research, Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, 80303.
†Assistant Professor, Colorado Center for Astrodynamics Research, Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, 80303.

Trajectory clustering, specifically, focuses on discovering patterns or groups of possible motions for a moving object.[8] There are various approaches to clustering trajectories, including summarizing trajectories by a set of defining characteristics, constructing a model that describes the region spanned by the trajectory, or using a time series.[9] A time series representation may offer a high-fidelity representation of a trajectory but suffer from the curse of dimensionality.[8] In this scenario, dimension reduction algorithms can be used with caution to construct a lower-dimensional embedding of the original high-dimensional data prior to clustering.[8, 10]

Clustering has previously been used in astrodynamics. Notable examples include Hadjighasem, Karrasch, Teramoto, and Haller using spectral clustering for detecting Lagrangian vortices within nonlinear dynamical systems;[11] Nakhjiri and Villac using k-means clustering for locating bounded motions near a distant retrograde orbit on a Poincaré map;[12] and Foslien, Guralnik, and Haigh using data mining to autonomously identify anomalous conditions on the International Space Station's gimbal system.[13] More recently, Bosanac as well as Bonasera and Bosanac used distributed clustering, dimension reduction, and classification to extract the geometries of a wide array of trajectories visualized on a Poincaré map in a multi-body system.[14, 15] Smith and Bosanac also used clustering to extract motion primitives from families of periodic orbits and hyperbolic invariant manifolds to supply building blocks for trajectory design in a multi-body system.[16] These examples demonstrate the value of clustering in extracting information from data sets in astrodynamics.

This paper focuses on using clustering to categorize possible spacecraft trajectories resulting from a set of uncertain state estimates by their geometry in the Earth-Moon system. For each state estimate, an array of initial conditions are sampled from within the region of uncertainty defined around a reference state. Each initial condition is then propagated in a desired dynamical model to produce continuous trajectories; in this paper, trajectories are generated in the circular restricted three-body problem (CR3BP). Each trajectory is then discretized into a set of states that are evenly distributed as a function of arclength and described by the position and time at the sampled states. This information is used to form a finite-dimensional feature vector that summarizes each continuous trajectory. Then, Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is used to cluster these feature vectors. Finally, the resulting clusters are refined to ensure trajectories with a similar geometry and itinerary are accurately grouped. The result is a set of clusters that supply a digestible summary of the array of possible future motions for these state estimates when subject to initial state uncertainty. This approach is used to group the trajectories generated from regions of uncertainty defined relative to reference states that are sampled along the following periodic orbits in the Earth-Moon CR3BP: an L_1 Lyapunov orbit, an L_1 northern halo orbit, a distant prograde orbit (DPO), and a 2:1 resonant orbit.

DYNAMICAL MODEL

The circular restricted three-body problem offers an approximate dynamical model for a spacecraft operating in the Earth-Moon system. The CR3BP describes the motion of a spacecraft of assumed negligible mass under the gravitational influence of two massive bodies, labeled primaries.[17] In this paper, the primaries are the Earth and the Moon which are modeled as point masses with constant mass, $M_{\oplus}$ and $M_{☾}$, respectively, and travel on circular orbits about their barycenter.[17] The length, mass, and time quantities of this system are nondimensionalized using their corresponding characteristic quantities l^*, m^*, and t^*: $l^* = 384,400$ km is the assumed constant distance between the Earth and Moon, m^* is the total mass of the primaries, and $t^* = 375,126.416$ seconds produces a nondimensional mean motion of the primary system equal to unity. A rotating reference frame is defined with the origin at the Earth-Moon barycenter and axes $\hat{x}\hat{y}\hat{z}$: $\hat{x}$ points from the Earth to the Moon, $\hat{z}$ is aligned with the orbital angular momentum vector of the primary system, and $\hat{y}$ completes the orthogonal right-handed triad.[17] The nondimensional state of the spacecraft in the rotating frame is then defined as $\bar{x} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$. Using these assumptions and definitions, the nondimensional equations of motion for a spacecraft in the CR3BP are

$$\ddot{x} - 2\dot{y} = \frac{\partial U}{\partial x}, \quad \ddot{y} + 2\dot{x} = \frac{\partial U}{\partial y}, \quad \ddot{z} = \frac{\partial U}{\partial z} \tag{1}$$

where $\mu = M_{☾} / (M_{\oplus} + M_{☾}) \approx 1.21505842 \times 10^{-2}$ is the mass ratio of the system, $U = \left(x^2 + y^2\right)/2 + (1-\mu)/r_1 + \mu/r_2$ is the pseudo-potential function, and the nondimensional distances of the spacecraft with respect to the Earth and Moon are $r_1 = \sqrt{(x+\mu)^2 + y^2 + z^2}$ and $r_2 = \sqrt{(x-1+\mu)^2 + y^2 + z^2}$, respectively. An integral of motion, known as the Jacobi constant, exists in the rotating frame and is equal to

$$C_J = \left(x^2 + y^2\right) + \frac{2(1-\mu)}{r_1} + \frac{2\mu}{r_2} - \dot{x}^2 - \dot{y}^2 - \dot{z}^2 \tag{2}$$

At a single value of this quantity, a wide variety of motions may exist. In particular, the space community commonly leverages five libration points, L_i where $i \in [1, 5]$, and continuous families of periodic orbits in the trajectory design process.[17, 18]

Periodic orbits are paths that are periodic in the rotating frame. Each periodic orbit that exists along a continuous family is completely specified by its orbital period and a non-unique state, $\bar{x}_{PO}$. The stability of a periodic orbit is often assessed using four nontrivial eigenvalues of the monodromy matrix, defined as the state transition matrix propagated for one period along a periodic orbit.[18] Two stability indices, s_i, are then defined using the nontrivial eigenvalues of the monodromy matrix, λ_j;[19] this paper uses the definition $s_1 = \lambda_1 + \lambda_2$ and $s_2 = \lambda_3 + \lambda_4$. When $|s_i| > 2$, the eigenvalues lie on the real axis and correspond to the existence of a pair of unstable and asymptotically stable modes. Larger values of $|s_i|$ indicate that motion that begins nearby will depart (or approach) the periodic orbit faster when exciting these modes. Alternatively, when $|s_i| < 2$, the eigenvalues are complex. If the eigenvalues lie on the unit circle when $|s_i| < 2$, the stability index indicates oscillatory or marginally stable motion.

DATA MINING

This section provides a high-level overview of relevant data mining concepts that are used within this paper. Specifically, this paper employs clustering to group trajectories and nonlinear dimension reduction to produce a lower-dimensional representation during cluster refinement. For more mathematical details on each concept, see the cited works.

Clustering

Clustering is a data mining technique used to group similar data points and separate dissimilar data points.[8] Each point within the data set is described by a feature vector, denoted $\bar{f}$. The elements of this feature vector must be selected to provide a useful summary of the data for the desired application. One method of determining the similarity of two data points is to calculate the distances between feature vectors using a selected distance measure.[7]

A variety of clustering algorithms exist, with the most common approaches classified as partitioning, model-based, hierarchical, or density-based methods.[7] Partitioning algorithms directly divide the data points into k groups whereas hierarchical methods leverage a dendrogram representation of possible clusters before grouping.[7] Density-based methods are often employed to identify clusters as dense regions separated by sparse regions within the data set.[7] Selection of a suitable algorithm is influenced by the characteristics of the data and the application.

Hierarchical Density-Based Spatial Clustering of Applications with Noise

This paper uses HDBSCAN, a clustering algorithm that was developed by Campello, Moulavi, and Sander as a hierarchical extension of the well-known Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm.[20] Common HDBSCAN input parameters used to group data include $N_{minCore}$ and $N_{minClust}$: $N_{minCore}$ governs the calculation of a core distance, d_{core}, used to estimate density across the data set while $N_{minClust}$ specifies the minimum number of members in a single cluster.[20] Specifically, d_{core} is the distance from a point to its ($N_{minCore}$ - 1)-nearest neighbors where any point with $N_{minCore}$ or more closest neighbors is considered a core point. Then, a mutual reachability distance, d_{mreach}, is calculated between points to separate high- from low-density regions: $d_{mreach}(a, b) = max\{d_{core}(a), d_{core}(b), d(a, b)\}$ where a and b are points in the data set and $d(a, b)$ is the distance between the two points.[20] A minimum spanning tree is constructed using the mutual reachability distance. Then, an agglomerative approach is applied to create a hierarchy of groups based on this distance. HDBSCAN uses this hierarchy to construct clusters that contain at least $N_{minClust}$ members and possess sufficient stability. Any members not grouped are classified as noise points and may represent outliers in the data or low-density, insufficiently sampled regions. HDBSCAN has previously been demonstrated to support grouping trajectories by geometry in a multi-body system in the work performed by Bosanac as well as Bonasera and Bosanac.[14,15] In this paper, the *hdbscan* package is used to access HDBSCAN in Python.[21]

Cluster Validation

When the structure of a data set is not known a priori, relative cluster validity indices may be valuable in assessing the quality of a clustering result.[22] The DBCV index is well-suited for cluster validation when HDBSCAN is used because it leverages information about the relative density between objects.[23] Mathematically, the DBCV index is calculated as

$$DBCV = \sum_{i=1}^{k} \left[\frac{|C_i|}{|Q|} V_C(C_i) \right] \tag{3}$$

where Q is the total number of points in the data set (inclusive of noise), k is the number of clusters, C_i is the i^{th} cluster, and $V_C(C_i)$ is the validity index of the i^{th} cluster.[23] This validity index, $V_C(C_i)$, is computed using the density sparseness of a cluster (DSC) and density separation of a pair of clusters ($DSPC$) as

$$V_C(C_i) = \frac{\min\limits_{1\leq j\leq k, j\neq i} (DSPC(C_i, C_j)) - DSC(C_i)}{\max\left(\min\limits_{1\leq j\leq k, j\neq i} (DSPC(C_i, C_j)), DSC(C_i) \right)} \tag{4}$$

The DBCV index possesses values between -1 and 1; when applied to a single data set, relatively higher values indicate a better clustering result with tighter and better separated clusters.[23] Accordingly, this quantity is also valuable in selecting HDBSCAN's input parameters.

Uniform Manifold Approximation and Projection

Uniform Manifold Approximation and Projection (UMAP) is a nonlinear dimension reduction algorithm developed by McInnes, Healy, and Melville.[10] First, UMAP uses Riemannian geometry and algebraic topology to construct a fuzzy topological representation of a higher-dimensional data set. Next, spectral embedding techniques are used to initialize a low-dimensional embedding of the data. This embedding is then optimized to possess a fuzzy topological structure resembling the higher-dimensional data as closely as possible. The result is a lower-dimensional embedding of the high-dimensional data that may be useful for data visualization, dimension reduction prior to clustering, or feature extraction. As an example, Bonasera and Bosanac have demonstrated its use in transforming high-dimensional trajectory data prior to clustering.[15]

UMAP is predominantly governed by three parameters. The first input parameter, m_{dist}, controls the layout by balancing the projection densities. The second input parameter, n_{dim}, sets the dimension of the lower-dimensional embedding. The final input, n_{neigh}, controls the neighborhood size and balances the representation of the local versus global structure of the data set. For a focus on local structure, lower values of n_{neigh} are used.

TECHNICAL APPROACH

This section presents a clustering-based approach for categorizing the trajectories generated from a set of uncertain state estimates in the vicinity of a reference state. The L_1 Lyapunov orbit family in the Earth-Moon system has previously been studied for cislunar space surveillance.[3] Therefore, each step of the approach is demonstrated for planar motion generated from the vicinity of a reference state along an L_1 Lyapunov orbit in the Earth-Moon CR3BP.

Step 1: Sample Initial Conditions

The first step in the categorization process is to sufficiently sample the state space within a region of uncertainty near a reference state. There are a variety of available sampling schemes including stochastic, deterministic, geometric, or hybrid sampling.[24,25] For a uniform, dense distribution of samples, suitable approaches may include Halton point, spiral, or grid-based methods. Halton points are a low-discrepancy point set guaranteed to be well-distributed over all dimensions but may leave areas under-sampled in higher dimensions.[26] Spiral sampling with the Golden ratio ensures a near-uniform, space-filling pattern with a faster computation time than low-discrepancy point sets; yet, it is challenging to uniformly distribute samples within a multi-dimensional object.[27,28] Geometric, uniform grid-based sampling fills the desired space by sampling at fixed intervals uniformly distributed in every dimension but suffers from sparse coverage in higher dimensions.[24]

In this paper, a uniform, pseudo-random grid-based method is used to generate six-dimensional initial conditions within the region of uncertainty. A uniform, densely-sampled region is desired to ensure distinct trajectory geometries are captured, if they exist. While six-dimensional uniform, grid-based sampling was explored, it did not densely fill the six-dimensional state space with a reasonable number of points. Therefore, position and velocity vectors are independently sampled in this paper using a square grid-based approach that supplies dense and uniform sampling in their associated three-dimensional subspaces. Specifically, samples from a unit grid are mapped to spheres defined by the 3σ uncertainties in either the position or velocity components. A ratio of volumes (or areas for planar motion) is used to determine the number of samples, q_{grid}, that are distributed within the grid to obtain the desired number of samples, q_{des}, within the sphere:

$$q_{grid} \approx \begin{cases} q_{des}\frac{A_{sq}}{A_{circ}} & \text{if } z = 0 \\ q_{des}\frac{V_{cube}}{V_{sphere}} & \text{if } z \neq 0 \end{cases} \tag{5}$$

where A_{grid} is the area of a square unit grid encompassing a unit circle, A_{circ} is the area of a unit circle, V_{cube} is the volume of an n-dimensional cube and V_{sphere} is the volume of an n-dimensional sphere. After the unit grid is generated, any sample that lies outside the unit sphere (or circle for planar motion) is discarded. The remaining samples are then scaled to produce position and velocity vectors using the associated quantities in the 3σ state uncertainties. The full six-dimensional state is then constructed by randomly pairing one sample from each of the position and velocity subspaces.

Based on state-of-the-art uncertainty levels for cislunar navigation using optical (angles only) measurements as presented by Bradley et. al.,[29] 3σ uncertainties of 10.5 km in position components and 10.5 m/s in velocity components are used in this paper. Channing Chow II et. al.,[6] Fedeler et. al.,[30] and Williams et. al.[31] present similar uncertainty values in their work for cislunar detection and tracking with uncertainty levels potentially varying for different periodic orbits. However, this paper focuses on the categorization of trajectories by geometry using a simple uncertainty model and leaves robust uncertainty modeling to other work.

To demonstrate this approach, consider a region of uncertainty around a single reference state along an L_1 Lyapunov orbit with a period of 12.269 days and Jacobi constant $C_J = 3.154$ in the Earth-Moon CR3BP. This nondimensional reference state is specified in the rotating frame as $\bar{x}_{PO} = [0.81698, 0, 0, 0, 0.19575, 0]^T$, occurring at the leftmost crossing of the x-axis as depicted in Figure 1. Perturbed, planar states are then sampled within the bounding circle defined by the 3σ state uncertainty using the presented grid-based sampling method. Figure 2 displays the final set of 1,006 initial conditions in position and velocity space for this example with bounding circles shown in black and the reference state denoted by a red dot.

Step 2: Propagate Trajectories

The second step of this process involves propagating the initial conditions in a selected dynamical model. In this paper, all initial conditions are propagated forward in time for at least 17.3 days in

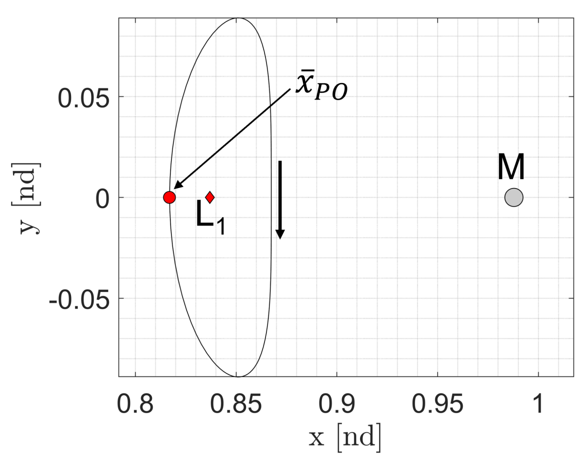

Figure 1: The 12.269-day L_1 Lyapunov orbit in the Earth-Moon CR3BP with reference state $\bar{x}_{PO}$.

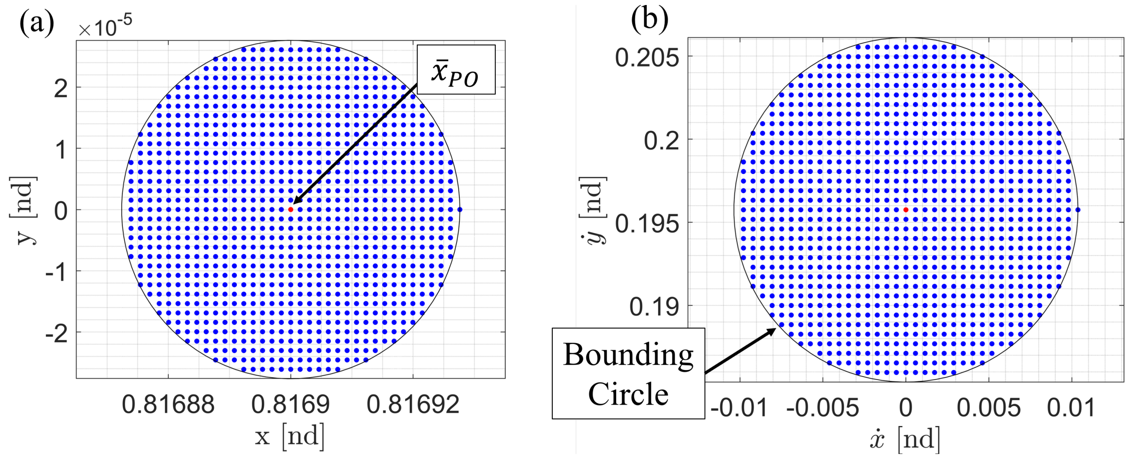

Figure 2: (a) Position and (b) velocity components of 1,006 planar initial conditions sampled from a unit grid given a reference state with uncertainty along the selected L_1 Lyapunov orbit.

the Earth-Moon CR3BP to exceed the periods of the selected orbits and allow distinct geometries to emerge. However, propagation is terminated early upon impact with a spherical approximation of either of the primaries. Following this procedure, Figure 3 displays a 200-member subset of the continuous trajectories generated from the 1,006 initial conditions defined in Step 1. These initial conditions occur near the blue circles with the Moon and Earth appearing as scaled gray circles, and red diamonds locating L_1 and L_2. Motions generated from a region of uncertainty around the selected reference state exhibit a variety of geometrically distinct paths that remain in the vicinity of the Moon or depart through one of the L_1 or L_2 gateways.

Step 3: Summarize Trajectories

Once propagated, each trajectory is discretized into a sequence of p states at equal intervals in arclength. When numerically integrating a state from t_1 to t_2, the arclength d_{arcL} is defined as the

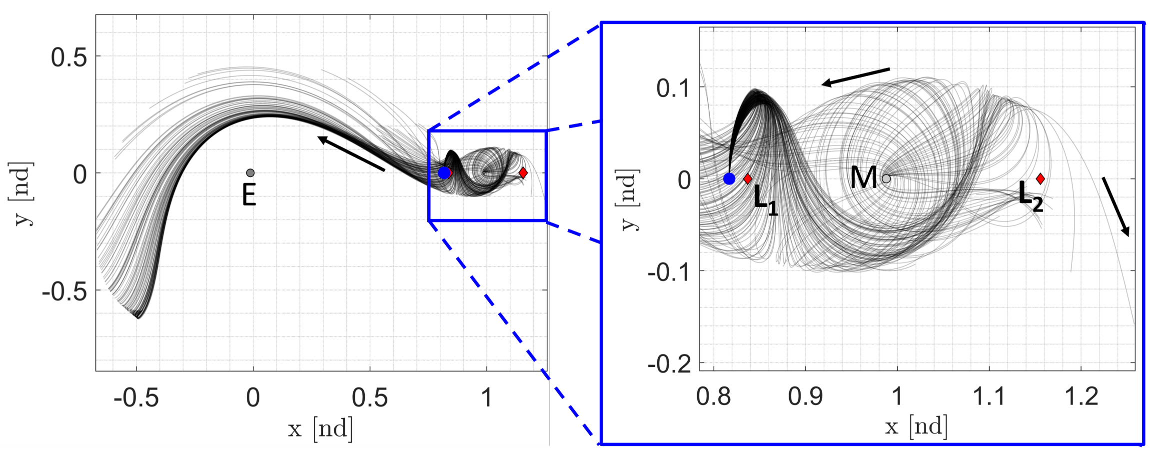

Figure 3: A 200-member subset of trajectories propagated for 17.3 days in the Earth-Moon CR3BP. The initial conditions are marked with blue circles.

distance traversed along the trajectory. In this paper, this quantity in the rotating frame is given as

$$d_{arcL} = \int_{t_1}^{t_2} |\bar{v}|\, dt \tag{6}$$

where $\bar{v} = [\dot{x}, \dot{y}, \dot{z}]^T$ is the velocity in nondimensional units.[32] Once the arclength is calculated for the entire trajectory, p states, including the final state, are sampled at intervals of $d_{arcL}/(p-1)$ along the arclength of the trajectory.

The number of states, p, used to discretize the trajectory is selected using a curve-based approach. Specifically, $p = 2(p_{max}+1)$ where p_{max} is the largest number of local maxima in the curvature along any trajectory in the set. The curvature, $\kappa(\bar{x})$, reflects the deviation from a straight line at the state $\bar{x}$ along a trajectory. Accordingly, local maxima in the curvature tend to occur near turning points in the configuration space, including near periapses and apoapses measured relative to primaries or equilibrium points. However, unlike periapses and apoapses, maxima in the curvature do not require specification of a reference point as the trajectory exhibits turning points in distinct regions of the Earth-Moon system. At a single state, the curvature is mathematically calculated as

$$\kappa(\bar{x}) = \frac{\sqrt{(\ddot{z}\dot{y} - \ddot{y}\dot{z})^2 + (\ddot{x}\dot{z} - \ddot{z}\dot{x})^2 + (\ddot{y}\dot{x} - \ddot{x}\dot{y})^2}}{(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{3/2}} \tag{7}$$

By using the largest number of local maxima in the curvature along any trajectory in a data set to define p, the most complex trajectory in the set tends to be sufficiently represented.

To demonstrate each step of this discretization process, consider a single, planar trajectory generated in Step 2. Figure 4(a) displays this specific trajectory along with blue stars at the locations of maximum curvature and a black star at the final state. Repeating this calculation for each of the 1,006 trajectories generated for 17.3 days in Step 2 produces $p_{max} = 7$. The trajectory from Figure 4(a) is then discretized into $p = 16$ states that are spaced equally in arclength; the resulting discretization is shown in Figure 4(b). This example demonstrates that the trajectory is uniformly discretized as a function of the distance traversed and is not biased by variations in the speed. However, the absence of states near perilune and apolune motivate ongoing investigation of alternative geometry-based discretization schemes.

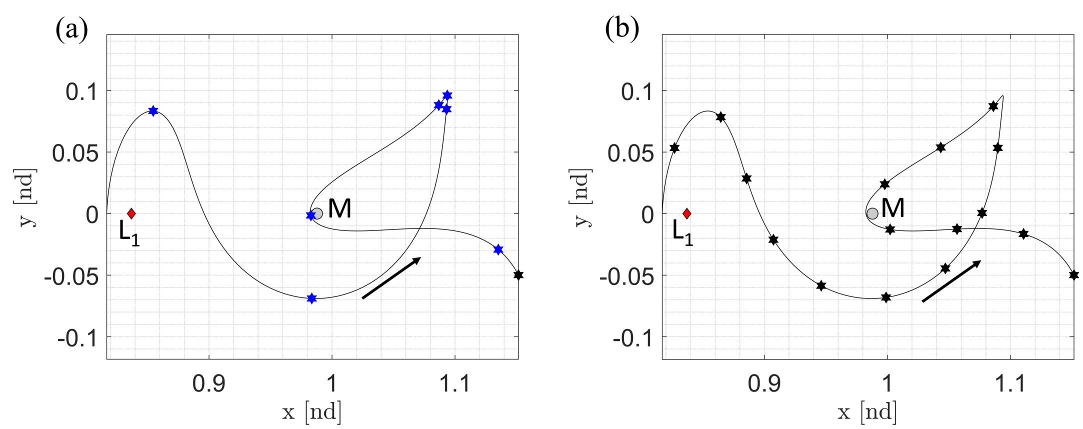

Figure 4: (a) Locations of local maximum curvature along a trajectory. (b) The final discretization of p states spaced equally in arclength along a trajectory.

The p sampled states are used to form finite-dimensional feature vectors $\bar{f}$ that describe the spatiotemporal variation of each trajectory. These features capture the shape of the trajectory, given enough states are sampled, by supplying position information and the elapsed time. In this paper, $\bar{f}$ for a trajectory is defined as

$$\bar{f} = \left[\frac{x_1 - (x_{prim})}{d_{L_1,L_2}}, \frac{y_1}{d_{L_1,L_2}}, \frac{z_1}{d_{L_1,L_2}}, \frac{\tau_1}{\tau_p}, \cdots, \frac{x_p - (x_{prim})}{d_{L_1,L_2}}, \frac{y_p}{d_{L_1,L_2}}, \frac{z_p}{d_{L_1,L_2}}, \frac{\tau_p}{\tau_p} \right] \tag{8}$$

where x_i, y_i, z_i are the nondimensional components of the position vector at state i in the rotating frame, x_{prim} is the nondimensional x-component of the closest primary body in the rotating frame to the initial condition, τ_i is the nondimensional time measured from the initial condition at state i and d_{L_1,L_2} is the nondimensional distance between L_1 and L_2. For planar trajectories, the z-component of position at each state is removed from this feature vector. The values τ_p and d_{L_1,L_2} are used to scale $\tau_i \in [0,1]$, $x_i, y_i, z_i \in [-0.5, 0.5]$ when the state is in the lunar vicinity. Of course, when trajectories depart the lunar vicinity, the normalized values of x_i, y_i, z_i can exceed these values. Nevertheless, this normalization approach limits the potential for one feature to dominate over others.

Step 4: Cluster Trajectories

To govern the clustering process used by HDBSCAN, the $N_{minCore}$ and $N_{minClust}$ hyperparameters are selected using a grid search. Specifically, the clustering process is applied to a single set of trajectories for various combinations of these two quantities, assuming a Euclidean distance metric for assessing similarity between feature vectors. Then, using a similar approach to Bosanac, the resulting number of clusters (k), noise points (O), and values of the DBCV index are examined.[14] The combination of values of $N_{minCore}$ and $N_{minClust}$ that produces a high DBCV index, relatively low number of noise points, and a reasonable number of clusters in the selected example serves as an initial guess prior to manual tuning when clustering each new trajectory set.

As an example, consider a hyperparameter search applied to the trajectories generated from within the region of uncertainty defined around the reference state at the leftmost x-axis crossing along the selected L_1 Lyapunov orbit. Figure 5 displays the corresponding values of (a) k, (b) O, and (c) $DBCV$ calculated from the clustering results produced by applying HDBSCAN with each combination of $N_{minCore}$ and $N_{minClust}$ within the range $[0, 20]$. Based on these results, the values of $N_{minCore}$ and $N_{minClust}$ are selected as 8 and 17, respectively, to produce $k = 7$, $O = 26$, and $DBCV = 0.492$. These values are marked in red in Figure 5.

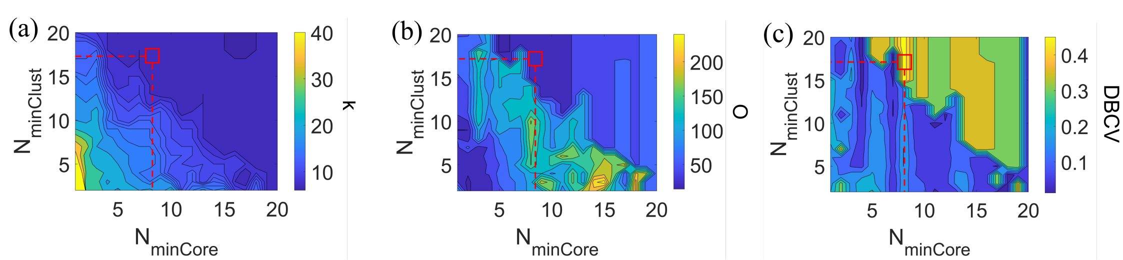

Figure 5: Hyperparameter search as a function of: (a) the number of clusters, k; (b) noise points, O; (c) DBCV index. $N_{minCore}$ and $N_{minClust}$ are selected based on the corresponding maximum DBCV index, boxed in red in all three plots.

The trajectories generated from a region of uncertainty around a reference state are grouped by clustering their feature vectors using HDBSCAN and the selected hyperparameters. Throughout this paper, the similarity between two feature vectors is assessed using the Euclidean distance metric. Although this distance metric can suffer from the curse of dimensionality when clustering high-dimensional feature vectors, it supports fast clustering when the feature vector has a fixed length.[15] Then, with $N_{minCore} = 8$ and $N_{minClust} = 17$, HDBSCAN is used to cluster the 1,006 trajectories generated from a region of uncertainty around the selected reference state along the 12.269-day L_1 Lyapunov orbit in the Earth-Moon CR3BP. Figure 6 displays the resulting clusters that capture the variety of geometries of these trajectories. Each subfigure displays a single group of trajectories in the Earth-Moon rotating frame. In addition, the final state for each trajectory is denoted with a blue dot and color indicates the elapsed time along each trajectory. Given the variety of geometry within this trajectory set, this summary of motion captures distinct sets of geometries; however, 26 of the 1,006 trajectories are categorized as noise points and plotted in Figure 7. These trajectories typically lie near the boundaries of clusters shown in Figure 6 or have a slightly different geometry evolution, such as not impacting the Moon. Additionally, not all trajectories are grouped correctly and may be labeled as outliers, e.g., in the second subfigure in the first row of Figure 6. Cluster refinement in the next step of the data-driven categorization process addresses noise points and outliers.

Step 5: Refine Clusters

The final step of this process aims to refine clusters for more accurate boundary approximation. Such refinement may be useful when leveraging the clusters for collision avoidance or constructing a mapping from initial states to each type of motion. In this paper, this process is composed of the following four steps: (a) inter-cluster sampling, (b) unlabeled trajectory grouping, (c) local clustering, and (d) aggregation.

Inter-cluster Sampling The goal of inter-cluster sampling is to increase the accuracy of cluster boundary approximation. To contextualize this process, note that the variety of geometry between the trajectories in distinct clusters from Step 4 tends to be correlated to variations in the velocity vec-

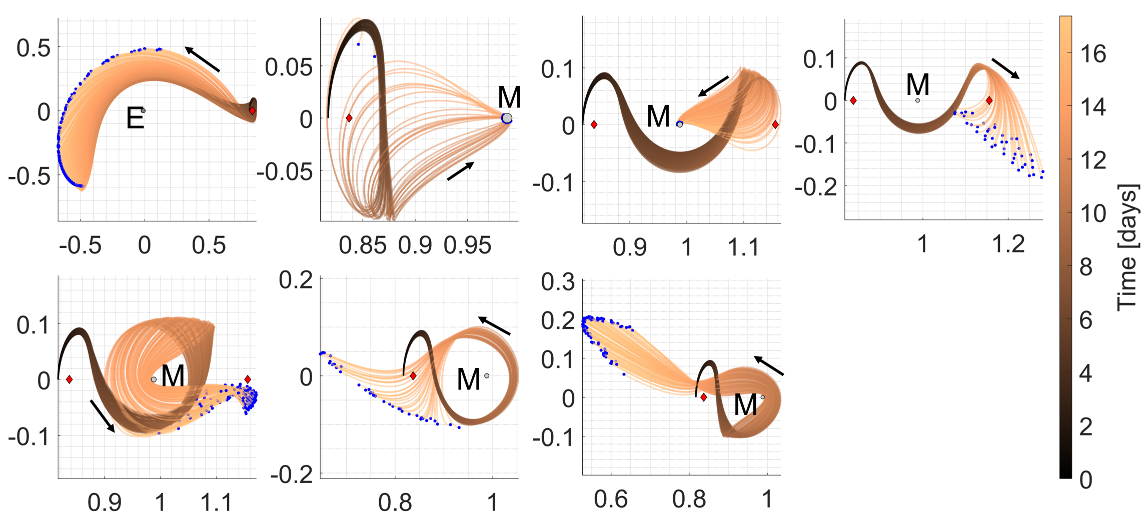

Figure 6: Initial grouping of trajectories generated from an uncertain state estimate near a single state along an L_1 Lyapunov orbit at a Jacobi constant of 3.154 in the Earth-Moon CR3BP.

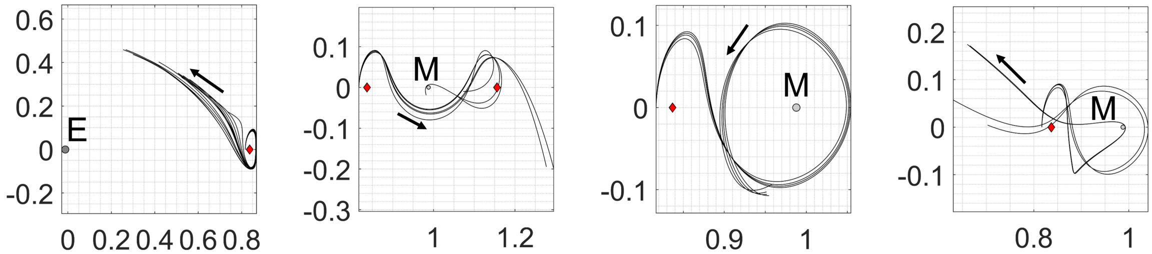

Figure 7: Twenty-six trajectories, displayed across four plots for clarity, designated as noise by HDBSCAN. The horizontal and vertical axes represent the x- and y position coordinates, respectively, in the Earth-Moon rotating frame in nondimensional units.

tor. Figure 8(a) displays the velocity components of each initial condition colored by the assigned cluster with white diamonds locating points designated as noise by HDBSCAN. In this example, the majority of noise points lie between cluster boundaries; therefore, expanding the trajectory set with additional samples between these boundaries may aid in grouping noise points along with increasing the accuracy of cluster boundaries.

The inter-cluster sampling scheme is based on a Voronoi diagram which divides an n-dimensional space into cells such that each cell consists of one point.[33] In this paper, every sample in velocity space is treated as if it lies in its own cell. Then, the cluster labels for adjacent cells are compared to determine if a sample is needed between them. This allows for sampling between clusters without relying on explicit boundary construction.

Adjacent samples are located based on Euclidean distances between initial conditions. In general, for one- to three-dimensional points, the distance, d_{pts}, between any point, in this case $\bar{v}_{pt,1}$, and its adjacent neighbors, $\bar{v}_{pt,i}$, in a square unit grid is $d_{pts} = |\bar{v}_{pt1} - \bar{v}_{pt,i}| < 2u_{grid}$ where u_{grid} is the grid size. The neighbors adjacent to each initial condition are identified by calculating the distance between the current initial condition, referred to as the central sample, and all other samples. If $d_{pts} < 2u_{grid}$, the point is adjacent to the central sample.

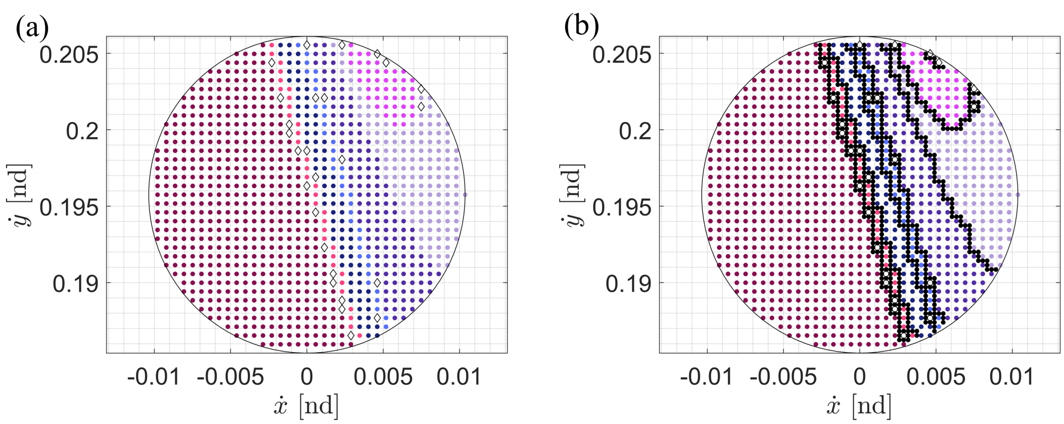

Figure 8: Each color represents a different cluster while white diamonds represent noise points. (a) Velocity space prior to inter-cluster sampling; (b) Five-hundred and twenty-eight samples from inter-cluster sampling are shown with black dots in velocity space.

The next step is to compare cluster labels of adjacent samples: if the central sample's neighbors lie in a different cluster, then one sample is placed linearly between the initial conditions in velocity space. Prior to sampling, the initial condition set is checked to determine if there is a pre-existing sample at that point in velocity space. To construct a six-dimensional state, the position components for the new velocity sample are recovered in a similar manner. The position components corresponding to the two original velocity samples are located with one sample placed linearly between them. Inter-cluster sampling for the L_1 Lyapunov orbit scenario used throughout this section produces an additional 528 planar initial conditions which are shown black in Figure 8(b). These samples fall between cluster boundaries without using explicit boundary construction.

Unlabeled Trajectory Grouping The goal of this step is to assign trajectories corresponding to inter-cluster samples and noise points to the previously defined clusters. First, all trajectories are summarized using position-based feature vectors. Removing time after an initial clustering step, such as in Step 4, increases placement accuracy of unlabeled trajectories in this application since they are predominately associated with a different geometry. This feature vector definition begins with the feature vector presented in Equation 8 and removes the time elements. Then, UMAP is used to construct a lower-dimensional representation to address the curse of dimensionality that makes it challenging to associate unlabeled trajectories with existing clusters. While UMAP does not preserve density, it does preserve relative distances over the global or local scale governed by n_{neigh}.[10] In this paper, the UMAP hyperparameters are selected as $n_{neigh} = 20$, $m_{dist} = 0$, and $n_{dim} = 2$ based on visual inspection; the selected value of $n_{neigh} = 20$ prioritizes capturing the local structure of the dataset in the lower-dimensional embedding. Each unlabeled point is assigned to the cluster belonging to its closest point, assessed using the Euclidean distance. Although this approach assigns every unlabeled point to an existing cluster, the boundaries of each cluster become more dense and outliers tend to be successfully removed during subsequent refinement steps.

In the L_1 Lyapunov orbit scenario, there are 26 noise points from the clustering in Step 4 while the inter-cluster sampling in Step 5a added 528 initial conditions; therefore, there are 554 unlabeled trajectories. Figure 9(a) displays the clustered trajectories in a two-dimensional UMAP representation while Figure 9(b) overlays the 554 unlabeled trajectories in cyan. Figure 10 shows the clusters in the higher-dimensional space after grouping these unlabeled trajectories. In all cases, the clusters are denser than in Figure 6. However, there may be outliers or clusters that should be split into

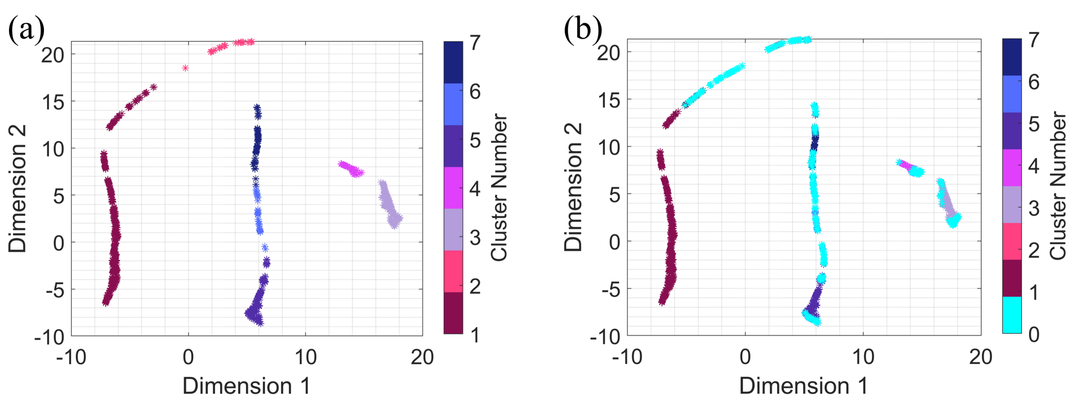

Figure 9: (a) Clustered trajectories represented by cluster color in a two-dimensional space using UMAP; (b) Clustered trajectories augmented with unlabeled data in cyan.

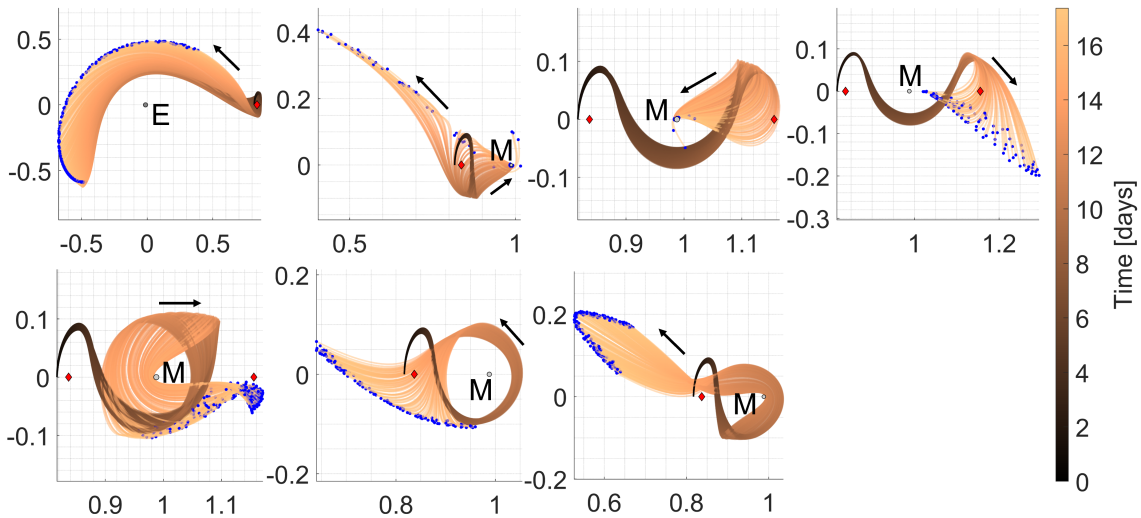

Figure 10: High-dimensional representation of trajectories grouped with UMAP during Step 5b of cluster refinement for an uncertain state estimate along an L_1 Lyapunov orbit.

multiple sub-clusters. The next step of the cluster refinement addresses this issue.

Local Clustering The goal of local clustering is to identify and remove outliers in the existing clusters as well as split clusters, if appropriate. All trajectories are represented with the position-based feature vectors constructed in Step 5b to focus solely on trajectory position in each cluster to help identify outliers in each group. Then, since local clustering is performed on subsets of the expanded dataset, hyperparameter searches must be conducted for $N_{minCore}$ and $N_{minClust}$ for each subset of data prior to local clustering with HDBSCAN. If the local clustering produces noise points, a similarity matrix, $\mathbf{S}$, is constructed using the normalized Euclidean distance between all trajectories within a cluster and a specified noise point, $\mathbf{d}_{traj,\bar{N}_p}$, such that $\mathbf{S} = 1 - \mathbf{d}_{traj,\bar{N}_p}$. Values of $\mathbf{S}$ then provide an intuitive comparison of the grouped trajectories and noise point. For any trajectory and noise point pair, if the largest value in the similarity matrix is greater than or equal to 0.90, indicating a similarity of 90% or greater, then the noise point is merged with its corresponding cluster. This is repeated for all noise points generated from local clustering for a single cluster. If the original cluster was split into multiple sub-clusters, as shown in Figure 11, then a similar process is repeated for comparing trajectories in different sub-clusters: sub-clusters are merged if a 90% or greater similarity is found between any pair of trajectories. This process is repeated for all clusters

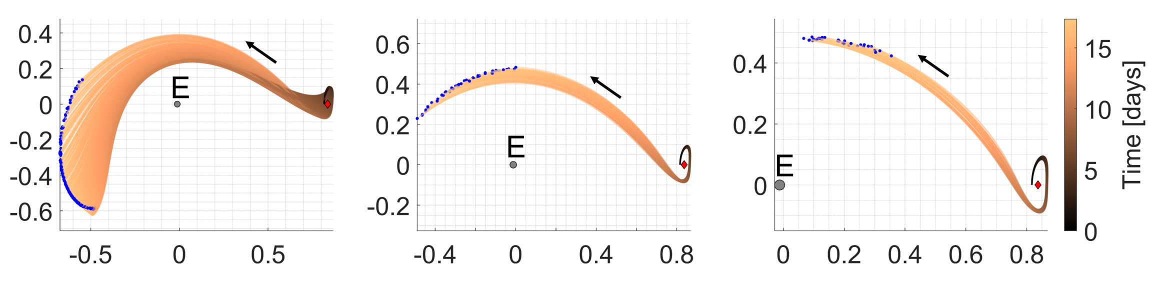

Figure 11: An example of an undesirably split cluster from the L_1 Lyapunov orbit scenario as a result of local clustering.

provided from Step 5b.

Aggregation The goal of the final step of the cluster refinement process, aggregation, is to determine if clusters from Step 5c should be merged to produce a condensed summary of motion. A similar process to merging sub-clusters is used here except the merging criteria is stricter since groups are now well-defined: a similarity of 94% or greater must be met along with a dynamic threshold pertaining to cluster size. The dynamic threshold is based on the size of all clusters: only the smallest clusters are merged with other clusters. The purpose of this threshold is to avoid aggregating clusters with a sufficiently large number of trajectories as the distinctive geometry may be difficult to analyze visually for an exceedingly large cluster with evolving geometry.

Applying this final step to the 1,534 trajectories generated from the vicinity of the state that lies along the L_1 Lyapunov orbit categorizes 1,514 trajectories with 20 remaining noise points. Figure 12 shows the final eight clusters, each with a distinct geometry. The in-plane stability index of this orbit, $s_1 = 1,955.774$, is large in magnitude. Accordingly, a spacecraft will naturally depart this orbit quickly, consistent with the variety of geometries extracted through this clustering approach.

Limitations of current approach A limitation of the cluster refinement method used in this paper is that not all noise points are guaranteed to be grouped. Noise points are a trade-off when using HDBSCAN and not unexpected when clustering in a higher-dimensional solution space. The benefits to using HDBSCAN, despite the noise points, include that it accommodates clusters of arbitrary shape and density located at various relative distances and does not require the number of clusters to be known a priori. However, noise point inclusion is important for understanding all potential trajectories that may be generated from an uncertain state estimate. Accordingly, reducing or eliminating the number of trajectories categorized as noise is an ongoing avenue of work.

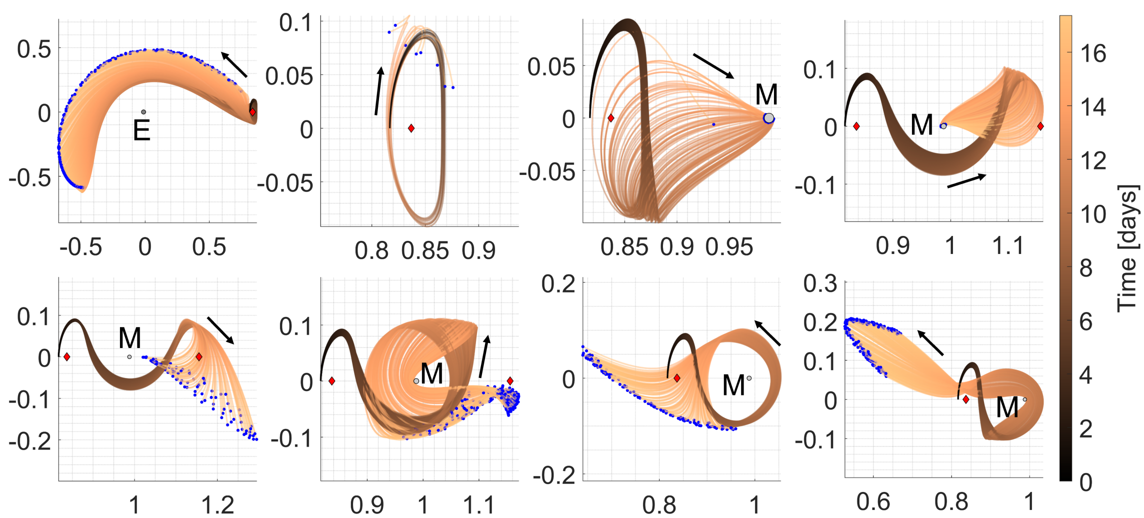

Figure 12: Eight final clusters of trajectories with distinct geometries generated from an uncertain state estimate along an L_1 Lyapunov orbit.

RESULTS

The data-driven categorization process described in the previous section is used to summarize sets of trajectories generated from uncertain state estimates near additional periodic orbits in the

CR3BP. Table 1 lists the stability, Jacobi constant, and periods of the orbits selected for assessment whereas Figure 13 plots each orbit along with the locations of the reference states with uncertainty.

Orbit	**Jacobi Constant**	**Orbit Period [days]**	s_1	s_2
L_1 northern halo orbit	3.148	11.993	1488.827	1.763
Distant prograde orbit	3.169	10.914	-2.091	-1.840
2:1 resonant orbit	2.754	27.246	1.557	2.000

Table 1: Properties of the selected periodic orbits.

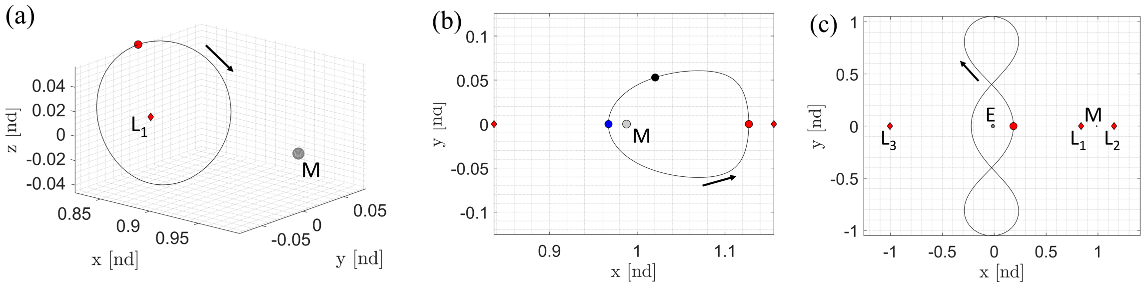

Figure 13: The selected (a) L_1 northern halo, (b) distant prograde, (c) 2:1 resonant orbits in the CR3BP are shown in black with the nominal initial conditions denoted with filled circles.

Summarized Motions Near an L_1 Northern Halo Orbit

A set of trajectories that are generated from within a region of uncertainty about a single state along an L_1 northern halo orbit in the Earth-Moon CR3BP are categorized. Specifically, a reference state at $\bar{x}_{PO} = [0.82412, 0, 0.05669, 0, 0.16712, 0]^T$ nondimensional units is selected along the orbit displayed in Figure 13(a). The data-driven categorization process identifies seven clusters composed of 1,650 trajectories integrated for 17.3 days, each described by $p = 12$ states equally spaced in arclength. Seven trajectories are still considered noise points and, therefore, do not appear in the clusters. The resulting groups of trajectories are displayed in Figure 14 using the same configuration as Figure 12. Within each subfigure, a distinct three-dimensional view is provided for clarity. Similar to the L_1 Lyapunov orbit scenario, there are trajectories that remain within the Moon vicinity and those that depart through the L_1 gateway to revolve about the Earth. It is not as apparent if motion near the L_2 gateway departs to the exterior region; a longer integration time would be needed to understand the future path of these trajectories.

Summarized Motions Near a Distant Prograde Orbit

This subsection presents categorized spacecraft motion with uncertainty near three states along a distant prograde orbit at a Jacobi constant of 3.169 in the Earth-Moon CR3BP. The orbit, along with the three selected reference states, is plotted in Figure 13(b). Once the set of initial conditions are sampled within the region of uncertainty about each reference state, the associated trajectories are generated for 17.3 days.

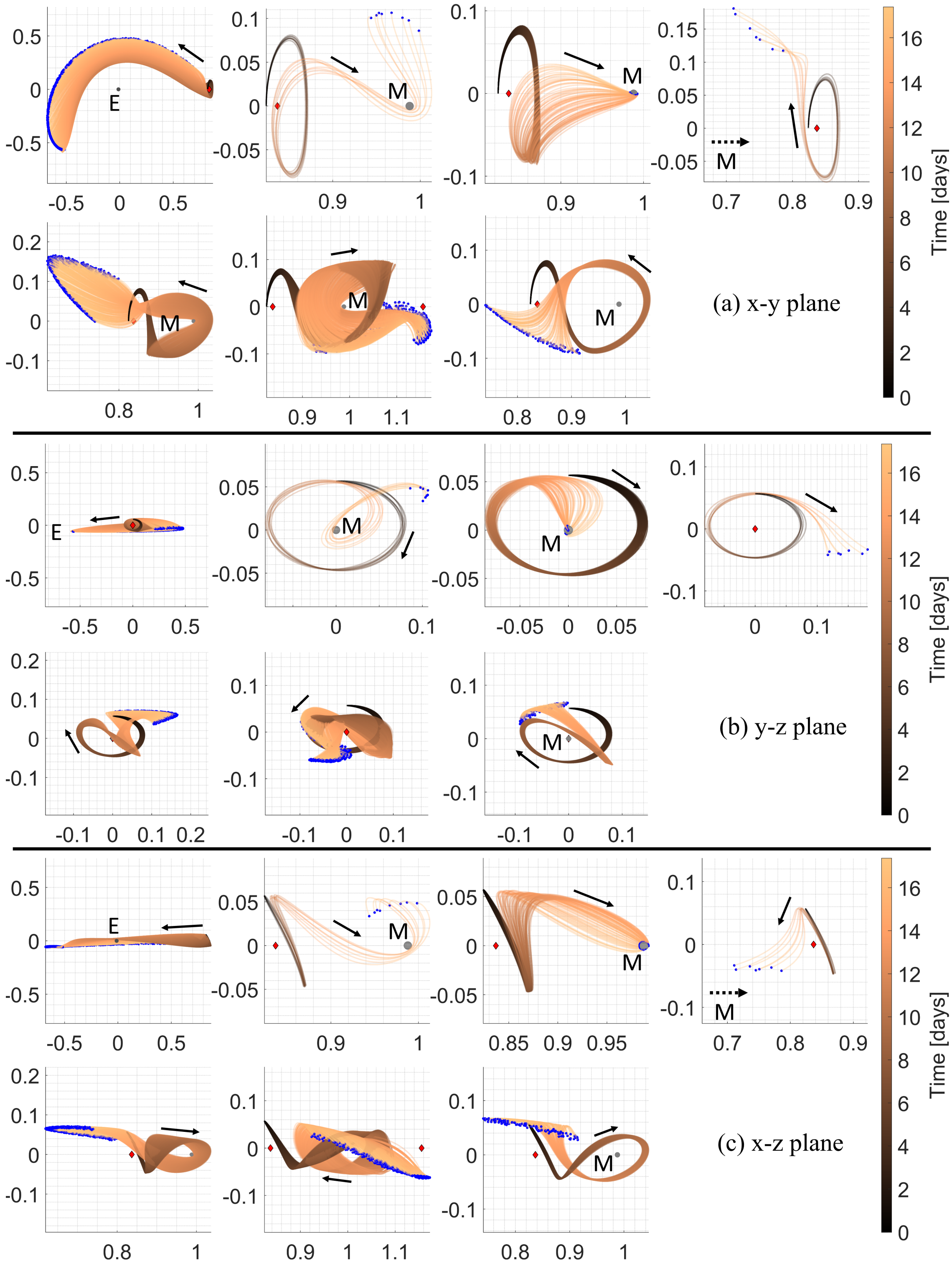

Figure 14: Seven distinct types of trajectories generated from an uncertain state estimate near an L_1 northern halo orbit at a Jacobi constant of 3.148 in the Earth-Moon CR3BP.

The reference state at perilune, shown in blue in Figure 13(b), corresponds to $\bar{x}_{PO} = [0.96732, 0, 0, 0, -1.02023, 0]^T$ nondimensional units. The trajectories generated from the region of uncertainty in its vicinity are discretized into $p = 6$ states equally spaced in arclength. These 1,051 trajectories are grouped into a single cluster with no noise points, as displayed in Figure 15(a). This single type of motion closely follows the DPO for less than one day before departing through the L_2 gateway for the exterior region of the system.

The next reference state, shown in red in Figure 13(b), lies at apolune at $\bar{x}_{PO} = [1.12719, 0, 0, 0, 0.09553, 0]^T$ nondimensional units. All trajectories from the initial condition sampling process are discretized into $p = 14$ states spaced equally in arclength. A single type of motion is generated from uncertain state estimates in its vicinity as displayed in Figure 15(b) for 1,103 trajectories with no noise points. The geometry continuously evolves from departing through the L_2 gateway to revolving about the Moon.

The final reference state, represented by a black circle in Figure 13(b), between perilune and apolune is located at $\bar{x}_{PO} = [1.02045, 0.05293, 0, -0.39930, -0.12813, 0]^T$ nondimensional units. A greater variety of potential motion geometries exist at this state, as seen in Figure 16. This dataset consists of 1,378 trajectories, each described by $p = 18$ states spaced equally in arclength, with only three trajectories not categorized. Once cluster captures motion departing through the L_2 gateway for the exterior region of the system. Two clusters correspond to motion that remains within the Moon vicinity for 17.3 days, but with a different geometry compared to the trajectories emanating from the uncertainty region around the second reference state. In one cluster, trajectories exhibit a significant apsidal rotation with each revolution around the Moon. In the other cluster, the speed reduces near apoapsis to temporarily produce a change in the direction of motion followed by a lower perilune.

A stability analysis of this DPO reveals unstable and stable modes. The in-plane stability index is near the critical value of $+2$, as seen in Table 1. Therefore, planar motions departing from the local neighborhood of a state along this periodic orbit may potentially produce paths with similar geometry to the DPO. However, the region of uncertainty used in this paper is not necessarily contained within a local neighborhood of each state where the linearized system sufficiently predicts the behavior in the nonlinear system. In this analysis, trajectories generated from the region of uncertainty

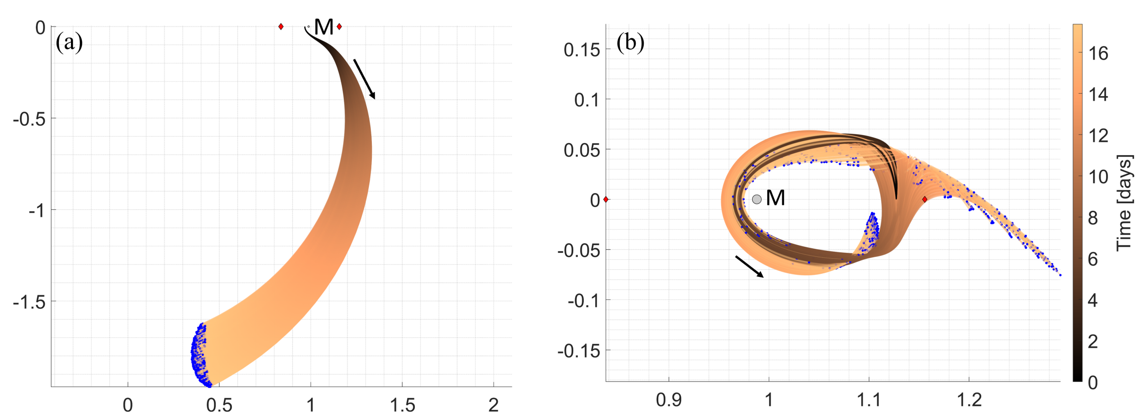

Figure 15: Distinct types of trajectory geometry generated near a distant prograde orbit at a Jacobi constant of 3.169 in the Earth-Moon CR3BP from states near (a) perilune and (b) apolune.

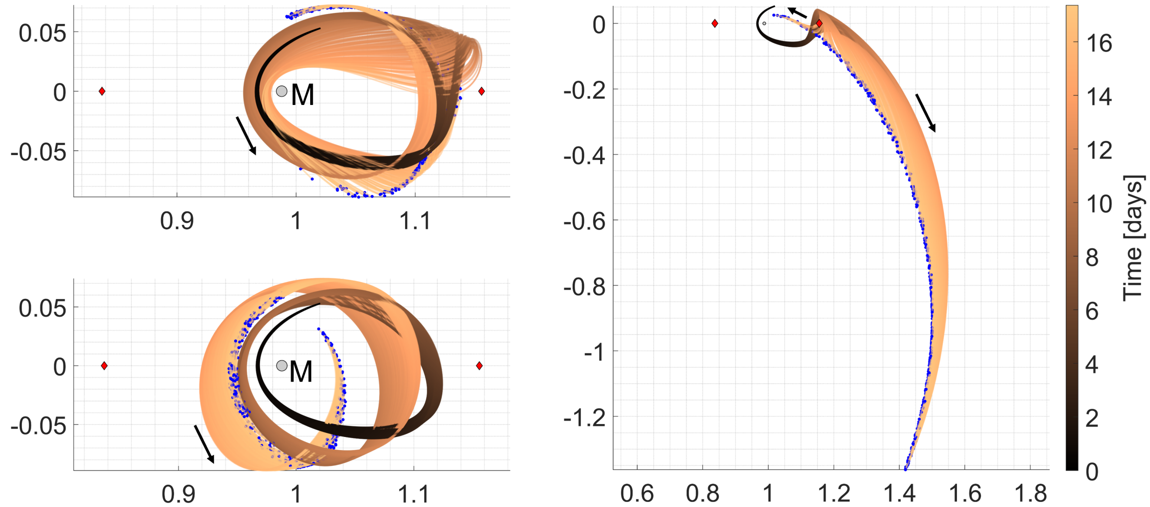

Figure 16: Distinct types of trajectories generated from an uncertain state estimate between perilune and apolune along a distant prograde orbit at a Jacobi constant of 3.169 in the Earth-Moon CR3BP.

defined around apolune tend to resemble the periodic orbit for the majority of the integration time. As the reference state moves closer to perilune, a region of increased sensitivity, new and distinct geometries emerge.

Summarized Motions Near a 2:1 Resonant Orbit

Trajectories generated from a region of uncertainty around a state along a 2:1 resonant orbit are also categorized. The 2:1 resonant orbit family has previously been studied for space surveillance.[4] This example summarizes motion near an orbit from this family in the Earth-Moon CR3BP with a Jacobi constant of 2.754. The selected reference state at $\bar{x}_{PO} = [0.18553, 0, 0, 0, 2.70274, 0]^T$ nondimensional units can be seen in Figure 13(c) along with the 2:1 resonant orbit. All 1,050 trajectories are described by $p = 18$ states equally spaced in arclength. The categorized motion resulting from uncertainty in this state is presented in Figure 17. Unlike previous scenarios, the

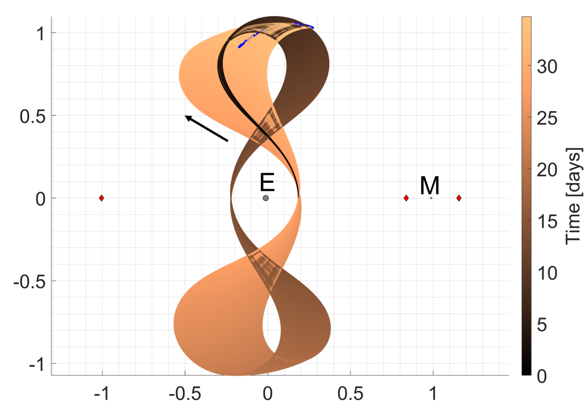

Figure 17: One type of trajectory geometry generated from an uncertain state estimate along a 2:1 resonant orbit at a Jacobi constant of 2.754 in the Earth-Moon CR3BP.

motion is integrated for 34.7 days to exceed the orbital period. A single cluster, with no noise points, summarizes the resulting geometry that resembles the 2:1 resonant orbit with no distinct departure from this geometry. From Table 1, this orbit admits stable in-plane modes. Although the region of uncertainty may exceed the reference state's local neighborhood where the linearized system accurately predicts the nonlinear system behavior, the discovery of a single cluster is consistent with this stability assessment.

CONCLUSION

Clustering is used to extract a digestible summary of the types of possible motions generated from an uncertain state estimate. In this paper, a data-driven categorization method is presented. First, a reference state with uncertainty is selected near motion along a periodic orbit. Then, additional states are sampled within a sphere described by the state uncertainty and centered on the reference state. These initial states are propagated forward in time in the Earth-Moon CR3BP to generate continuous trajectories. Each trajectory is discretized by evenly distributed states along its arclength and then summarized using a finite-dimensional feature vector. Once the feature vectors are constructed, HDBSCAN is used to cluster the trajectories into fundamental path types. Clusters are then refined to produce more compact, smooth groups. A detailed example of this method is presented for trajectories generated near a state that lies along an L_1 Lyapunov orbit in the Earth-Moon CR3BP. Additional results for trajectories near an L_1 northern halo orbit, distant prograde orbit, and 2:1 resonant orbit in the Earth-Moon CR3BP are also supplied. These examples demonstrate that clustering can be used to extract a digestible summary of the types of trajectories that may be generated from an uncertain spacecraft state estimate in a multi-body system.

ACKNOWLEDGMENTS

This paper was supported in part by a fellowship award under contract FA9550-21-F-0003 through the National Defense Science and Engineering Graduate (NDSEG) Fellowship Program, sponsored by the Air Force Research Laboratory (AFRL), the Office of Naval Research (ONR) and the Army Research Office (ARO). Travel to this conference was supported in part by the Bahls Travel Award.

REFERENCES

[1] *Continuing Kepler's Quest: Assessing Air Force Space Command's Astrodynamics Standards*. The National Academies Press, 2012.

[2] M. R. Thompson, M. Bolliger, and N. P. Ré, "Cislunar Orbit Determination and Tracking via Simulated Space-Based Measurements," *Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS)*, Maui, HI, 2021.

[3] C. Frueh, K. Howell, K. J. DeMars, and S. Bhadauria, "Cislunar Space Situational Awareness," *Proceedings of the AAS/AIAA Space Flight Mechanics Meeting*, No. AAS 21-290, Virtual, 2021.

[4] C. Frueh, K. Howell, K. J. DeMars, S. Bhadauria, and M. Gupta, "Cislunar Traffic Management: Surveillance Through Earth-Moon Resonance Orbits," *Proceedings of the 8th European Conference on Space Debris (virtual)*, Darmstadt, Germany, ESA Space Debris Office, 2021.

[5] T. Wolf, E. Zucchelli, and B. A. Jones, "Multi-Fidelity Uncertainty Propagation for Objects in Cislunar Space," *AIAA SciTech Forum*, San Diego, CA, 2022.

[6] C. C. C. II, C. J. Wetterer, J. Baldwin, M. Dilley, K. Hill, P. Billings, and J. Frith, "Cislunar Orbit Determination Behavior: Processing Observations of Periodic Orbits with Gaussian Mixture Model Estimation Filters," *Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS)*, Maui, HI, 2021.

[7] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. Waltham, MA: Morgan Kaufmann Publishers, 3 ed., 2012.

[8] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. CRC Press, 2014.

[9] T. W. Liao, "Clustering of time series data - A survey," *Pattern Recognition*, Vol. 38, No. 11, 2005, pp. 1857–1874.

[10] L. McInnes, J. Healy, and J. Melville, ""UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," arXiv:1802.03426, 2020.

[11] A. Hadjighasem, D. Karrasch, H. Teramoto, and G. Haller, "A Spectral Clustering Approach to Lagrangian Vortex Detection," *Physics Review E*, Vol. 93, No. 6, 2016, pp. 63–83.

[12] N. Nakhjiri and B. Villac, "Automated stable region generation, detection, and representation for applications to mission design," 2015.

[13] K. Z. Haigh, W. Foslien, and V. Guralnik, "Data Mining for Space Applications," *Space OPS 2004 Conference*, Montreal, Quebec, Canada, 2004.

[14] N. Bosanac, "Data Mining Approach to Poincaré Maps in Multi-Body Trajectory Design," *Journal of Guidance, Control, and Dynamics*, Vol. 43, No. 6, 2020, pp. 1190–1200.

[15] S. Bonasera and N. Bosanac, "Applying data mining techniques to higher-dimensional Poincaré maps in the circular restricted three-body problem," *Journal of Celestial Mechanics and Dynamical Astronomy*, Vol. 133, No. 51, 2021.

[16] T. Smith and N. Bosanac, "Constructing Motion Primitive Sets to Summarize Periodic Orbit Families and Hyperbolic Invariant Manifolds in a Multi-Body System," 2022.

[17] V. Szebehely, *Theory of Orbits: The Restricted Problem of Three Bodies*. Academic Press, 1967.

[18] W. S. Koon, M. W. Lo, J. E. Marsden, and S. D. Ross, *Dynamical Systems, the Three Body Problem and Space Mission Design*. 2006.

[19] K. C. Howell, "Three-Dimensional Periodic Halo Orbits," *Celestial Mechanics*, Vol. 32, No. 1, 1984, pp. 53–71.

[20] R. J. Campello, D. Moulavi, and J. Sander, *Density-Based Clustering Based on Hierarchical Density Estimates*, p. 160–172. Springer, Berlin, Heidelberg, 2013.

[21] L. McInnes, J. Healy, and S. Astels, "The hdbscan Clustering Library," `https://hdbscan.readthedocs.io/en/latest/index.html`, 2016.

[22] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M.Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, Vol. 46, 2013, pp. 243–256.

[23] D. Moulavi, P. A. Jaskowiak, R. J. G. B. Campello, A. Zimek, and J. Sander, "Density-Based Clustering Validation," *Proceedings of the 14th SIAM International Conference on Data Mining (SDM)*, Philadelphia, PA, 2014.

[24] D. G. L. R, M. Pedergnana, and S. G. García, "Smart sampling and incremental function learning for very large high dimensional data," *Neural Networks*, 2015.

[25] D. P. Hardin, T. Michaels, and E. B. Saff, "A Comparison of Popular Point Configurations on S^2," *Dolomites Research Notes on Approximation*, Vol. 9, 2016, pp. 16–49.

[26] L. Kocis and W. J. Whiten, "Computational Investigations of Low-Discrepancy Sequences," *ACM Transactions on Mathematical Software*, Vol. 23, No. 2, 1997, pp. 266–294.

[27] C. Schretter, L. Kobbelt, and P.-O. Dehaye, "Golden Ratio Sequences for Low-Discrepancy Sampling," *ACM Journal of Graphics Tools*, Vol. 16, No. 2, 2012, pp. 95–104.

[28] González, "Measurement of Areas on a Sphere Using Fibonacci and Latitude–Longitude Lattices," *Mathematical Geosciences*, Vol. 42, 2010, pp. 49–64.

[29] N. Bradley, Z. Olikara, S. Bhaskaran, and B. Young, "Cislunar Navigation Accuracy Using Optical Observations of Natural and Artificial Targets," *Journal of Spacecraft and Rockets*, Vol. 57, No. 4, 2020, pp. 777–792.

[30] S. Fedeler, M. Holzinger, and W. Whitacre, "Sensor tasking in the cislunar regime using Monte Carlo Tree Search," *Advances in Space Research*, Vol. 70, 2022, pp. 792–811.

[31] J. Williams, D. E. Lee, R. J. Whitley, K. A. Bokelmann, D. C. Davis, and C. F. Berry, "Targeting Cislunar Near Rectilinear Halo Orbits for Human Space Exploration," *Proceedings of the 27th AAS/AIAA Space Flight Mechanics Meeting*, No. AAS 17-267, San Antonio, TX, 2017.

[32] T. Shifrin, *Differential Geometry: A First Course in Curves and Surfaces*. University of Georgia, 2023.

[33] F. Aurenhammer, "Voronoi Diagrams - A Survey of a Fundamental Geometric Data Structure," *ACM Computing Surveys*, Vol. 23, No. 3, 1991, pp. 345–405.