(Preprint) AAS 25-515

CONSTRUCTING AN ATLAS OF NATURAL SPACECRAFT TRAJECTORIES IN AN EPHEMERIS MODEL OF CISLUNAR SPACE

Natasha Bosanac*

This paper presents a foundation for constructing an atlas of natural spacecraft motion that begins near the Moon in a point mass ephemeris model of the Earth-Moon system. These trajectories span up to 21 days subject to the point mass gravitational influences of the Earth, Moon, and Sun. Groups of geometrically distinct trajectories are automatically generated across an array of energy levels and epochs using distributed clustering. These groups of trajectories are then hierarchically visualized in a digestible manner using a dendrogram. An atlas may eventually support trajectory analysis, design, and prediction in cislunar space.

INTRODUCTION

With the increased utilization of cislunar space, a broad summary of the solution space in a high fidelity model is valuable. Understanding these possible motions supports designing complex trajectories with various itineraries or predicting potential future paths for an observed object. In low-fidelity models, researchers have leveraged traditional dynamical systems techniques to study families of periodic orbits, homoclinic or heteroclinic connections between them, and general transfers. For instance, Broucke presented and examined a diverse array of periodic orbits in the Earth-Moon circular restricted three-body problem (CR3BP) [1]. In another example, Leiva and Briocco studied families of periodic orbits that travel between the Earth and Moon vicinities in the Earth-Moon CR3BP [2]. Previous researchers have also studied general trajectories using chaos indicators and Lagrangian coherent structures in models of various fidelity. For instance, Onozaki, Yoshimura, and Ross used Lagrangian coherent structures to identify families of transit and nontransit orbits in the Earth-Moon-Sun system [3]. Building upon these and many more contributions, it is valuable to systematically summarize a broader array of trajectories, including 1) spatial trajectories that do not follow fundamental solutions, or their finite-time equivalents, and 2) trajectories that exist in high-fidelity models or models that depend on physical parameters.

Similar goals in summarizing complex datasets or dynamical mechanisms appear across a wide variety of fields. For instance, in disciplines such as biology, the diverse array of groups of cells or tissues have been used to construct broad summaries often labeled as atlases; each atlas can be useful in supporting knowledge discovery and development of new medical technologies or theories. Examples of these atlases include the Tabula Sapiens [4], mouse cell atlases [5], and brain atlases

*Associate Professor, Colorado Center for Astrodynamics Research, Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, 3775 Discovery Drive, Boulder, CO 80303.

[6]. In these examples, biological data is either manually grouped or clustered in an unsupervised manner to identify detailed groups. Then, techniques such as hierarchical clustering and dimension reduction have been used to identify broader classes of cell types or hierarchical taxonomies. Atlases like these, which exist across a broader array of disciplines, inspire the work in this paper.

This paper leverages a clustering-based framework that has recently been introduced and refined by Bosanac to automatically summarize a complex solution space in a chaotic multi-body gravitational system [7, 8]. The current procedure, presented by Bosanac [8, 9] and used within this paper, begins with two definitions: 1) an approach for discretely summarizing a continuous trajectory using geometry information, and 2) a two-step density-based clustering process. Using these definitions, the summarization process consists of three key steps: 1) sampling trajectories across the phase space in a geometry-aware manner, 2) summarizing smaller partitions of trajectories, and 3) aggregating local clusters from distinct partitions to form a global cluster summary. The existing clustering-based approach presented by Bosanac [8, 9] has been demonstrated by summarizing 21-day trajectories that begin near the Moon across an array of energy levels in the circular restricted three-body problem and a single energy level in a point mass ephemeris model of cislunar space. These works have demonstrated the capability to automatically summarize a geometrically-diverse array of spacecraft trajectories and associate them across models of distinct fidelity.

This paper focuses on creating a foundation for constructing a preliminary atlas of natural spacecraft motion that begins in the vicinity of the Moon in a point mass ephemeris model of the Earth-Moon system. These trajectories are generated to span up to 21 days in an ephemeris model that incorporates the point mass gravitational influences of the Earth, Moon, and Sun. Groups of geometrically distinct trajectories are generated automatically using clustering across an array of energy levels and epochs, selected to span the period of the Moon's orbit relative to the Earth. These groups of trajectories are then organized hierarchically in a dendrogram to support visualization and analysis, similar to cell and animal taxonomies. Selected groups within the atlas are also analyzed in this paper. The goal of generating an atlas is to support further knowledge discovery within our technical community in spacecraft trajectory design, analysis, and prediction.

BACKGROUND

Reference Frames

A Moon-centered inertial frame is used during trajectory generation and sampling. The origin of this frame is located at the center of the Moon whereas the axes $\hat{X}\hat{Y}\hat{Z}$ are equal to the axes of most recent realization of the International Celestial Reference Frame (ICRF) [10, 11]. These axes are also used to access ephemerides of selected celestial bodies in the DE440 kernel [12]. In this frame, the state of the spacecraft is expressed as $\bar{X} = [X, Y, Z, X', Y', Z']^T$, where the notation $()'$ indicates an observer fixed in the inertial frame.

A pulsating Earth-Moon rotating frame is used for trajectory description and visualization. The origin of this frame is located at the Earth-Moon barycenter. The axes are defined consistent with their traditional use in the lower-fidelity circular restricted three-body problem as follows: $\hat{x}$ is directed from the center of the Earth to the center of the Moon, $\hat{z}$ is parallel to the orbital angular momentum vector of the Moon's orbit relative to the Earth, and $\hat{y}$ completes the right-handed triad. These axes are scaled to maintain a distance of unity between the centers of the Earth and Moon. In this frame, the nondimensional state of the spacecraft is expressed as $\bar{x} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$, where the notation $\dot{(\_)}$ indicates an observer fixed in this pulsating, rotating frame.

To support the framework presented in this paper, states are transformed between the two frames. The transformation matrices rely on the position vector $\bar{R}_{E,L}(t)$ of the Moon relative to the Earth at a specified epoch t in the axes of the ICRF and its time derivatives; in this definition, the subscripts L and E identify the Moon and Earth, respectively. The axes of the rotating frame are then expressed in terms of $\bar{R}_{E,L}(t)$ as the following three column vectors:

$$\hat{x} = \frac{\bar{R}_{E,L}}{||\bar{R}_{E,L}||} \qquad \hat{z} = \frac{\bar{R}_{E,L} \times \bar{V}_{E,L}}{||\bar{R}_{E,L} \times \bar{V}_{E,L}||} \qquad \hat{y} = \hat{z} \times \hat{x} \tag{1}$$

The first time derivatives of these axes are calculated as

$$\frac{d\hat{x}(t)}{dt} = \frac{\bar{R}'_{E,L}}{||\bar{R}_{E,L}||} - \hat{x}\frac{\hat{x} \cdot \bar{R}'_{E,L}}{||\bar{R}_{E,L}||} \qquad \frac{d\hat{z}(t)}{dt} \approx 0 \qquad \frac{d\hat{y}(t)}{dt} = \hat{z} \times \frac{d\hat{x}(t)}{dt} \tag{2}$$

where the second approximation is also used in GMAT [13]. The second time derivatives of these axes are calculated as

$$\begin{aligned}\frac{d^2\hat{x}(t)}{dt^2} = \frac{\bar{R}''_{E,L}}{||\bar{R}_{E,L}||} - \frac{\bar{R}'_{E,L}(\bar{R}'_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} - \frac{d\hat{x}(t)}{dt}\frac{(\bar{R}'_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} \\ - \hat{x}\left(\frac{\bar{R}''_{E,L} \cdot \hat{x}}{||\bar{R}_{E,L}||} + \frac{\bar{R}'_{E,L}}{||\bar{R}_{E,L}||} \cdot \frac{d\hat{x}(t)}{dt} - \frac{(\bar{R}'_{E,L}(t) \cdot \hat{x})^2}{||\bar{R}_{E,L}(t)||^2}\right)\end{aligned} \tag{3}$$

$$\frac{d^2\hat{y}(t)}{dt^2} = \hat{z} \times \frac{d^2\hat{x}(t)}{dt^2} \qquad \frac{d^2\hat{z}(t)}{dt^2} \approx 0 \tag{4}$$

Finally, the third time derivatives of these axes are equal to

$$\begin{aligned}&\frac{d^3\hat{x}(t)}{dt^3} = \frac{\bar{R}'''_{E,L}}{||\bar{R}_{E,L}||} - 2\frac{\bar{R}''_{E,L}(\bar{R}'_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} - \frac{\bar{R}'_{E,L}}{||\bar{R}_{E,L}||^2}\left(\bar{R}'_{E,L} \cdot \frac{d\hat{x}(t)}{dt}\right) - \frac{\bar{R}'_{E,L}(\bar{R}''_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} \\ &+ 2\frac{\bar{R}'_{E,L}(\bar{R}'_{E,L} \cdot \hat{x})^2}{||\bar{R}_{E,L}||^3} - \frac{d^2\hat{x}(t)}{dt^2}\frac{(\bar{R}'_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} - 2\frac{d\hat{x}(t)}{dt}\left(\frac{\bar{R}'_{E,L}}{||\bar{R}_{E,L}||} \cdot \frac{d\hat{x}(t)}{dt}\right) - \frac{d\hat{x}(t)}{dt}\frac{(\bar{R}''_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} \\ &+ \frac{d\hat{x}(t)}{dt}\frac{(\bar{R}'_{E,L} \cdot \hat{x})^2}{||\bar{R}_{E,L}||^2} - \hat{x}\frac{(\bar{R}'_{E,L})}{||\bar{R}_{E,L}||} \cdot \frac{d^2\hat{x}(t)}{dt^2} - 2\frac{d^2\hat{x}(t)}{dt^2}\frac{\bar{R}''_{E,L}}{||\bar{R}_{E,L}||} \cdot \frac{d\hat{x}(t)}{dt} - 3\frac{\hat{x}(\bar{R}'_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2}\left(\bar{R}'_{E,L} \cdot \frac{d\hat{x}(t)}{dt}\right) \\ &- \frac{d\hat{x}(t)}{dt}\frac{(\bar{R}''_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} - \frac{\hat{x}(\bar{R}'''_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} - 3\frac{\hat{x}(\bar{R}'_{E,L} \cdot \hat{x})(\bar{R}''_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} + \frac{d\hat{x}(t)}{dt}\frac{(\bar{R}'_{E,L} \cdot \hat{x})^2}{||\bar{R}_{E,L}||^2} - 2\frac{\hat{x}(\bar{R}'_{E,L} \cdot \hat{x})^3}{||\bar{R}_{E,L}||^3}\end{aligned} \tag{5}$$

$$\frac{d^3\hat{y}(t)}{dt^3} = \hat{z} \times \dddot{\hat{x}} \qquad \frac{d^3\hat{z}(t)}{dt^3} \approx 0 \tag{6}$$

In these expression, first-order, forward finite differences of the velocity vector are used to approximate the second and third time derivatives of the Moon relative to the Earth in the inertial frame.

The expressions for the axes of the rotating frame and their time derivatives are used to transform the state vector of the spacecraft between the rotating and inertial frames. An inertial position vector for the spacecraft at an epoch t is transformed to the rotating frame as

$$\bar{r}_{B,sc}(t) = \left[{}^{R}C(t)^{I}\right] \bar{R}_{L,sc}(t) + \bar{r}_{B,L} \tag{7}$$

where the rotation matrix $[^R C(t)^I]$ equals

$$\left[^R C(t)^I\right] = \begin{bmatrix} \hat{x}^T(t) \\ \hat{y}^T(t) \\ \hat{z}^T(t) \end{bmatrix} \tag{8}$$

and the subscript B indicates the Earth-Moon barycenter. The first, second, and third time derivatives of the position vector are also expressed in the rotating frame as

$$\bar{v}_{B,sc}(t) = \dot{\bar{r}}_{B,sc}(t) = \left[^R \dot{C}(t)^I\right] \bar{R}_{L,sc}(t) + \left[^R C(t)^I\right] \bar{V}_{L,sc}(t) \tag{9}$$

$$\bar{a}_{B,sc}(t) = \ddot{\bar{r}}_{B,sc}(t) = \left[^R \ddot{C}(t)^I\right] \bar{R}_{L,sc}(t) + 2\left[^R \dot{C}(t)^I\right] \bar{V}_{L,sc}(t) + \left[^R C(t)^I\right] \bar{A}_{L,sc} \tag{10}$$

$$\dddot{\bar{r}}_{B,sc}(t) = \left[^R \dddot{C}(t)^I\right] \bar{R}_{L,sc}(t) + 3\left[^R \ddot{C}(t)^I\right] \bar{V}_{L,sc}(t) + 3\left[^R \dot{C}(t)^I\right] \ddot{\bar{R}}_{L,sc}(t) + \left[^R C(t)^I\right] \dddot{\bar{R}}_{L,sc} \tag{11}$$

In these expressions, the time derivatives of the rotation matrix are calculated using row vectors equal to the transpose of the corresponding time derivatives of $\hat{x}(t), \hat{y}^T(t), \hat{z}(t)$ as defined in the previous paragraph.

Ephemeris Model

The motion of the spacecraft within cislunar space is modeled using a point mass ephemeris model of the Moon, Earth, and Sun. In this model, the spacecraft is assumed to possess a negligible mass in comparison to these three celestial bodies. Furthermore, the state vectors of the Moon, Earth, and Sun are accessed in the axes of the ICRF using the DE440 lunar and planetary ephemerides along with the associated Spacecraft, Planet, Instrument, C-matrix, Events (SPICE) toolkit provided by NASA's Navigation and Ancilliary Information Facility [12, 14].

The equations of motion are expressed in the Moon-centered inertial axes. The second-order, vector differential equation is written as

$$\ddot{\bar{R}}_{L,sc} = -GM_L\left(\frac{\bar{R}_{L,sc}}{R_{L,sc}^3}\right) + G\sum_{i=E,S} M_i\left(\frac{\bar{R}_{sc,i}}{R_{sc,i}^3} - \frac{\bar{R}_{L,i}}{R_{L,i}^3}\right) \tag{12}$$

where the subscript S identifies the Sun, G is the universal gravitational constant, and M_i is the mass of body i [15]. The quantities in this equation are nondimensionalized using the instantaneous values of the characteristic quantities from the lower-fidelity Earth-Moon circular restricted three-body problem [16]: length quantities are nondimensionalized to produce an instantaneous distance of unity between the centers of the Earth and Moon, time quantities are normalized to produce an instantaneous mean motion of the Earth-Moon system that is equal to unity, and mass quantities are nondimensionalized to ensure that the sum of the masses of the Earth and Moon are equal to unity.

Curvature

Differential geometry supplies foundational concepts that are useful for shape interrogation of nonlinear trajectories [17]. At an instant of time t, the spacecraft possesses a position vector $\bar{r}(t) = [x(t), y(t), z(t)]^T$, velocity vector $\dot{\bar{r}}(t) = [\dot{x}(t), \dot{y}(t), \dot{z}(t)]^T$, and acceleration vector $\ddot{\bar{r}}(t) =$

$[\ddot{x}(t), \ddot{y}(t), \ddot{z}(t)]^T$ in the pulsating, rotating frame. Over an interval $t \in [t_0, t_f]$, the arclength s describes the distance traversed along a trajectory and is equal to [18]

$$s = \int_{t_0}^{t_f} ds = \int_{t_0}^{t_f} ||\dot{\bar{r}}(t)||dt \tag{13}$$

In addition, the curvature $\kappa(t)$ at a single state along the trajectory captures the instantaneous deviation from a straight line and, therefore, the rate of change of the orientation of the tangent vector as a function of arc length [17]. For a spatial trajectory, the unsigned curvature is calculated as [17]

$$\kappa(t) = \frac{||\dot{\bar{r}}(t) \times \ddot{\bar{r}}(t)||}{||\dot{\bar{r}}(t)||^3} \tag{14}$$

but possesses a singularity when the speed is exactly equal to zero. Integrating this quantity as a function of arclength produces the total absolute curvature [17], equal to

$$\kappa_{tot}(t_0, t_f) = \int_{t_0}^{t_f} \kappa(s)ds = \int_{t_0}^{t_f} \kappa(s)\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}dt \tag{15}$$

This quantity reflects the angle swept out by the trajectory within its evolving osculating plane, monotonically increasing by 2π with each revolution.

Clustering

The goal of clustering algorithms is to automatically group similar objects [19]. This similarity is assessed using a set of characteristics that are encoded into a finite-dimensional feature vector $\bar{f}$ and a specified distance measure for comparing the feature vectors of two objects. Although a variety of clustering algorithms exist, this paper follows the procedure presented by Bosanac [8, 9] to use the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm developed by Ester et al. [20] and the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm developed by Campello, Moulavi, and Sander [21] in a two-step process. These two density-based clustering algorithms are used because they select clusters based on sufficiently dense groupings of data within the feature vector space and, therefore, do not require a priori knowledge of the expected number of clusters.

DBSCAN DBSCAN organizes members of a dataset into a cluster if their neighborhoods are density-connected [20]. To define this clustering process, DBSCAN relies on categorizing members into one of the following three classes [20]:

- Core points possess at least m_{pts} neighbors that lie within a radius of ϵ in the feature vector space; this region is labeled its m_{pts}-neighborhood. Through this definition, core points lie in sufficiently dense regions of the feature vector space.
- Border points lie in the m_{pts}-neighborhood of a core point but do not lie in sufficiently dense regions of the feature vector space.
- Noise points do not lie in the m_{pts}-neighborhood of any core points and lie in insufficiently dense regions of the feature vector space.

Using these definitions, two core points are density-connected if they exist within a distance of ϵ in the feature vector space [20]. A single cluster is first defined by the chain of core points that are density-connected to each other. Then, the border points of these core points are added to the cluster. This process labels each member of a dataset as either belonging to a specific cluster or as noise

[20]. These cluster or noise assignments are influenced by the values of the governing parameters m_{pts} and ϵ. Accordingly, DBSCAN is useful for clustering when both parameters can be specified or calculated heuristically to be constant across an entire dataset.

HDBSCAN HDBSCAN was developed as a hierarchical extension of DBSCAN, eliminating the specification of a single value of ϵ prior to clustering [21]. First, the core distance $d_{core}(\bar{f}_i)$ of the ith member of a dataset is calculated as the distance to its m_{pts}-nearest neighbor [21]. Then, the mutual reachability distance is defined between the i-th and j-th members as

$$d_{reach}(\bar{f}_i, \bar{f}_j) = \max(d_{core}(\bar{f}_i), d_{core}(\bar{f}_j), d(\bar{f}_i, \bar{f}_j)) \tag{16}$$

where $d(\bar{f}_i, \bar{f}_j)$ is the distance between the feature vectors of the two members [21]. This transformation to a mutual reachability distance further separates members that also lie in lower-density regions. HDBSCAN then constructs a graph with nodes corresponding to members of the dataset and edges weighted by their mutual reachability distance [21]. This graph is summarized by a minimum spanning tree that is used to construct a cluster hierarchy by gradually removing the edges with the highest weights and identifying the connected components [21]. From this hierarchy of all possible clustering assignments as a function of mutual reachability distance, the selected clusters possess at least m_{clmin} members and maximize stability using an excess of mass definition [21]. As presented by Malzer and Baum [22], a scalar quantity ϵ_{merge} can also be used to constrain the minimum value of the mutual reachability distance between two members that are assigned to distinct clusters. Following this procedure, HDBSCAN assigns each member of a dataset to a cluster or as noise. Although HDBSCAN does not leverage the concept of border points, noise points that meet the definition of a border from DBSCAN can be assigned to the cluster of the associated core point in a post-processing step [23]. In contrast to DBSCAN, HDBSCAN is useful when the value of ϵ is not known a priori or may vary across the expected clusters.

TECHNICAL APPROACH

This section presents a brief overview of the technical approach used to generate a clustering-based summary of a geometrically diverse array of trajectories. First, two general definitions are presented for 1) generating a dataset of trajectories across the solution space and 2) the two-step clustering process for generating groups of geometrically similar trajectories. These definitions are then used in the clustering-based framework for automatically summarizing the solution space. This general procedure follows the approach developed by Bosanac [8, 9], with some new modifications to accommodate the use of a high-fidelity ephemeris model. A new step to hierarchically visualize the clusters is also presented.

Generating a Dataset

Generating Trajectories The trajectories of interest are generated to begin near the Moon at a specified initial epoch. Each trajectory is propagated from an initial condition in the point mass ephemeris model for up to 21 days [8]; this duration is selected to enable sufficiently diverse geometries to emerge. However, this propagation terminates early upon impact with the Earth or Moon, each modeled as spheres with a radius equal to their equatorial radii of 6,378.1363 km and 1738.2 km, respectively [13].

Each trajectory is generated numerically along with information about its curvature. The augmented state used during propagation consists of 1) the state vector of the spacecraft in the Moon-

centered inertial frame and 2) the elapsed total curvature in the Moon-centered inertial frame. Although trajectories are described by their state information in the rotating frame prior to clustering, calculating the total curvature in the inertial frame reduces the required computational time as frame transformations and additional SPICE calls to ephemerides are not required at each time step. This augmented state is propagated using the first-order differential equations capturing the equations of motion in Eq. 12 and the integrand of Eq. 15. Integration is performed using the GNU Scientific Library in C++ with an 8/9th order Runge-Kutta method [24]. Events, such as satisfaction of the termination criteria or those used to sample the trajectory, are detected during numerical integration using Brent's root-finding method for computational efficiency [25].

Sampling Trajectories A nonlinear, continuous trajectory is sampled geometrically to produce a sequence of states [8, 9]; these states are directly used to construct a finite-dimensional description for clustering. In a Moon-centered inertial frame, a trajectory with a total curvature of $\kappa_{tot}(t_0, t_f)$ completes $r = \lceil \kappa_{tot}(t_0, t_f)/\pi \rceil$ half-revolutions within its osculating plane. States are sampled along this trajectory as follows [8, 9]:

- The first sample is placed at the initial condition
- Along the first $r-1$ half-revolutions, N_a states are sampled at every integer multiple of pi/N_a in the total curvature
- Along the final half-revolution or less, N_a states are evenly distributed in the elapsed total curvature along this half-revolution, including the final state

Sampling in the total absolute curvature produces a sequence of $N_s = N_a r + 1$ states that captures the geometry of the trajectory in the inertial frame. States are located closer together at segments of the trajectory where the shape is changing most significantly in the inertial frame, but more evenly spaced when the trajectory resembles a circle, for example. The value of N_a must be selected to balance increasing the fidelity of capturing trajectory geometry via larger values of N_a with reducing the computational requirements via smaller values of N_a. In this paper, $N_a = 3$ is selected to balance both objectives, consistent with the approach previously presented by Bosanac [8, 9]. Although this approach does not sample states using the geometry of the trajectory in the rotating frame, it supports fast computation and tends to distribute fewer states near lower speed segments in the rotating frame where the trajectory is changing direction.

Describing Trajectories The states sampled along each trajectory are used to generate two finite-dimensional vectors that are used during clustering [8, 9]. First, a shape-based feature vector $\bar{f}_s$ is defined using unit vectors $\hat{T}_i$ that are tangent to the path in the Earth-Moon pulsating rotating frame at the ith sample. This feature vector is calculated as

$$\bar{f}_s = \left[\hat{T}_1, \hat{T}_2, ..., \hat{T}_{N_f-1}, \hat{T}_{N_f}\right] \tag{17}$$

This vector possesses a dimension of $3N_f$ and possesses a singularity when the speed is exactly zero. Using the tangent or velocity unit vector instead of the velocity vector limits disparities in the order of magnitude of the elements as trajectories pass through regions of distinct sensitivity. Next, a position-based feature vector $\bar{f}_p$ is defined using the position vectors $\bar{r}_i$ at the ith sample in the Earth-Moon pulsating rotating frame. This feature vector is calculated as

$$\bar{f}_p = \left[\bar{r}_1, \bar{r}_2, ..., \bar{r}_{N_f-1}, \bar{r}_{N_f}\right] \tag{18}$$

to produce a $3N_f$-dimensional vector.

Clustering a Set of Trajectories

The two-step clustering procedure developed by Bosanac is used to group trajectories by their geometric similarity [8, 9]. The first step involves coarsely generating initial groups based on the shape of the trajectories for their entire duration. Then, each coarse group is refined to produce clusters of trajectories with a consistently similar path and shape through the configuration space.

Coarse Shape-Based Clustering Initial, coarse clustering is performed in the shape-based feature vector space using HDBSCAN [8, 9]. First, each trajectory is described by its N_s-dimensional shape-based feature vector, $\bar{f}_s$, calculated for its entire duration. Then, these feature vectors of trajectories that complete the same number of half-revolutions are input to HDBSCAN for clustering. This particular algorithm and approach is useful when the number of expected clusters is not known a priori and those clusters may possess distinct densities across the dataset. The result is a set of coarse groupings $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_n\}$ and a set of unlabeled, noise points $\mathcal{N}_c$. The trajectories that are designated as noise are discarded from further analysis.

When using HDBSCAN to coarsely cluster the trajectories, the governing parameters are selected as predefined constant values [8, 9]. The minimum cluster size is selected as m_{clmin}=5 to prioritize discovering localized variations between trajectories but with at least a few neighbors. Then, the neighborhood sizes are selected using $m_{pts} = m_{clmin} - 1 = 4$ to span the minimum cluster membership. In addition, the minimum threshold for separating clusters in $\bar{f}_s$ is selected using the sum of distances between a sequence of N_s unit vectors that are separated by an angle of 5 degrees, i.e., $\epsilon_{merge} = 2\sqrt{N_s}\sin(5/2)$. Finally, the distance metric used through clustering is selected as the Euclidean distance for increased computational speed and to only compare states at similar locations within their sequences.

Because the shape-based feature vectors are higher-dimensional, the same distance between two trajectories could correspond to geometric differences at the start or end of the trajectories, for example. Although those trajectories might be geometrically different, the clustering algorithm only assigns groups based on the scalar distance between the trajectories for their entire duration. Alternatively, the same distance between two trajectories may result in substantial differences in geometry depending on the region of the system. Accordingly, the coarse groupings may contain trajectories with noticeable geometric differences. Furthermore, two trajectories with a similar shape but distinct path through the configuration space may also be contained within the same coarse group. Thus, cluster refinement is a valuable part of the trajectory clustering process to ensure high-quality groupings as the trajectories pass through various regions of the system.

Cluster Refinement A refined set of clusters is constructed to ensure that each group consists only of trajectories that possess a similar path and shape through the configuration space for their entire duration. This process, previously developed by Bosanac [8, 9], is modeled after a convoy detection scheme that groups moving objects together if their sampled configurations are density connected throughout the entire trajectory. First, consider a set of N trajectories in a single coarse grouping from the first step of the clustering process. The ith state along each of these trajectories is then used to form N three-dimensional position-based feature vectors. These N feature vectors are clustered using DBSCAN to identify groups of trajectories that are density-connected at their ith sample in $\bar{f}_p$. A graph is then formed to capture this grouping decision. In this graph, each trajectory is represented by a node. Then, trajectories in each group that are clustered together and possess position-based feature vectors at the ith sample that lie within the ϵ- neighborhood of another trajectory define an undirected, unweighted edge between their corresponding nodes. This

process is repeated in the shape-based feature vector space at the ith sample. Next, this process is repeated for all samples along the trajectories to produce $2N_s$ clustering results. Groups of at least $n_{minclust}$ trajectories that are consistently connected in all $2N_s$ graphs form a refined cluster. Any trajectories that are not assigned to a cluster are then discarded from further consideration.

During the cluster refinement process, DBSCAN's governing parameters are selected adaptively using heuristics [8, 9]. The size of the neighborhood used to capture density is selected to match the associated parameter in HDBSCAN, i.e., $m_{pts} = 4$. Then, the neighborhood radius used to determine when two samples are density-connected in the selected feature vector space is selected using the following heuristic: $\epsilon = m_{pts} \max(e, \epsilon_{threshold})$ where e is the median value of the distance between the sample along a trajectory and its nearest neighbor. This heuristic for e adapts to the data from each coarse grouping, without being biased by the presence of outliers. Furthermore, when this value is recalculated at each sampled state along a trajectory, e adapts to a trajectory that passes through regions of distinct sensitivity in a system. In the heuristic for ϵ, $\epsilon_{threshold}$ is a minimum threshold for separating trajectories, selected as 4×10^{-3} in the position-based feature vector space to be consistent with the grid size and $2\sin(5/2)$ in the shape-based feature vector space. Similar to the previous step, the Euclidean distance is used to assess similarity between the three-dimensional feature vectors.

Cluster Representative A refined cluster of trajectories is summarized by a representative member [8, 9]. This member is extracted as the medoid of the cluster [7], i.e., the member that is most similar to all other members in a selected feature vector space [19]. In this paper, this calculation is performed using the position-based feature vector evaluated along the entire duration of each trajectory. For a cluster C_k that is composed of P_k members, the medoid is located as

$$\mathcal{T}_{med,k} = \text{argmin}_{\mathcal{T}_i \in \mathcal{C}_k} \left[\sum_{j=1, i \neq j}^{P_k} d(\bar{f}_i, \bar{f}_j) \right] \tag{19}$$

where $d(\bar{f}_i, \bar{f}_j)$ is the Euclidean distance between the feature vectors of the i-th and j-th trajectories and $\mathcal{T}_i$ is the i-th trajectory.

Data-Driven Summarization Process

Step 1: Sample the Solution Space The initial position vectors of all trajectories to be sampled are defined using a uniform grid in the Earth-Moon pulsating, rotating frame [8, 9]. This grid is defined with the following nondimensional position components: $x \in [-0.834, 1.156]$, $y \in [-0.2, 0.2]$, $z \in [-0.2, 0.2]$ with a step size of $\Delta_{pos} = 0.004$. Accordingly, these initial position vectors all lie in the vicinity of the Moon and extend just beyond the L_1 and L_2 gateways.

Within a single dataset, the initial conditions are constrained to produce a specified value of the Jacobi constant $C_{J,des}$ in the Earth-Moon CR3BP and initial epoch. First, the speed in the rotating frame v_{rot} is calculated as [**szbehely**]

$$v_{rot} = \sqrt{C_{J,des} - 2U^*} \tag{20}$$

Although this quantity is not a constant of motion in an ephemeris model, the use of the Jacobi constant to calculate the speed simply supports consistently discretizing and sampling the phase space. If a position vector produces an imaginary speed at the selected value of the Jacobi constant, it is discarded. Next, each dataset is also generated at a single initial epoch. The following initial epochs are evenly spaced across one period of the Moon's orbit, i.e., 27.32 days.

The velocity vector of each initial condition is selected to produce a maximum in the curvature in the Moon-centered inertial frame at a single, specified epoch. Accordingly, the initial condition is a geometrically meaningful initial state that plays an analogous role to an apse relative to a specified reference point. This maximum in the curvature is calculated in the inertial frame to ensure consistency with the sampling procedure and computational speed. To select these initial conditions, the families of velocity vectors that produce maxima in the curvature at a single position vector are numerically and discretely approximated. Then, their associated trajectories are generated and coarsely clustered to produce a set of representative velocity vectors that summarizes the broader array of paths.

The velocity vector is parameterized by two angles in the rotating frame: 1) an in-plane angle $\theta_{xy} \in [0, 360^\circ]$ measured from the x-axis and 2) an out-of-plane angle $\theta_z \in [-90^\circ, 90^\circ]$ measured from the xy-plane. The velocity vector is expressed in terms of these angles as

$$\bar{v} = \dot{x}\hat{x} + \dot{y}\hat{y} + \dot{z}\hat{z} = v_{rot}\cos(\theta_{xy})\cos(\theta_z)\hat{x} + v_{rot}\sin(\theta_{xy})\cos(\theta_z)\hat{y} + v_{rot}\sin(\theta_z)\hat{z} \tag{21}$$

Note that these velocity vectors are parameterized in the rotating frame for convenience and consistency but are used to calculate maxima in the curvature along a trajectory in the inertial frame.

Next, initial guesses for the velocity vectors that produce maxima in the curvature in the Moon-centered inertial frame are calculated. At each value of θ_z, sampled in 1 degree increments, θ_{xy} is sampled in 1 degree increments and the time derivative of the curvature $\dot{\kappa}(t_0)$ is calculated in the Moon-centered inertial frame. If two neighboring values of θ_{xy} produce a values of $\dot{\kappa}(t_0)$ with an opposite sign, their average is stored as an initial guess for a velocity vector that may produce a maximum in the curvature at the current value of θ_z.

These initial guesses are then used to calculate the velocity vectors with the desired speed in the rotating frame and produce maxima in the curvature in the inertial frame. The values of $\dot{x}$ and $\dot{y}$ must satisfy the following conditions:

$$\frac{d\kappa(\bar{X})}{dt} = 0, \qquad \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{v^2} - 1 = 0 \tag{22}$$

The value of $\dot{z}$ does not change, ensuring that the out-of-plane angle is equal to θ_z in the rotating frame. The first condition requires that the state vector be transformed into the inertial frame. The two variables, $\dot{x}$ and $\dot{y}$ are iteratively updated from their initial guess using the fsolve toolbox in Matlab. If a solution is generated, the value of $\ddot{\kappa}$ is calculated. If this quantity is negative, a maximum in curvature has been calculated. Repeating this process for all initial guesses at all values of θ_z produces one or more one-parameter families of velocity vectors in the rotating frame.

Because state vectors and other quantities are calculated in the inertial and rotating frames, the implementation must be efficient to support reasonable computational times. To achieve this goal, the initial epoch is fixed when generating segments of the dataset. Thus, the transformation matrix between the rotating and inertial frames and its time derivatives only need to be calculated once before generating a large number of velocity vectors for initial guesses that produce curvature maxima at a fixed initial position. Furthermore, the position vector can be transformed only once between the inertial and rotating frames. This efficiency is valuable as the transformation between rotating and inertial frames requires the use of SPICE functions to access ephemerides of the Moon relative to the Earth and matrix computations as described in the background.

Trajectories are generated from each velocity vector that produces a maximum in the curvature in the inertial frame at the discretely sampled values of $\dot{z}$. Once these trajectories are generated

numerically as described earlier, they are sampled based on their total curvature in the inertial frame and described by the position vectors and velocity unit vectors at those sampled states in the rotating frame. The shape-based feature vectors along trajectories completing the same number of half-revolutions in the inertial frame are then input to HDBSCAN for coarse clustering. At this step, $m_{clmin} = 3$, $m_{pts} = 2$, and $\epsilon_{merge} = 0$ are selected as the governing parameters due to the localized nature of the dataset. This clustering step generates groups of trajectories that emanate from the same position vector at the same initial epoch. Each cluster is then summarized by a representative trajectory, equal to its medoid, as described earlier. The initial velocity vectors of the representative trajectories of all clusters form the velocity vectors sampled at this position vector. Furthermore, the associated trajectories are added to the dataset for clustering throughout the rest of the procedure.

Step 2: Cluster Individual Partitions of Trajectories To support computational efficiency, the trajectories sampled from across the solution space in Step 1 are assigned to smaller partitions. First, trajectories that complete the same number of half-revolutions in the Moon-centered inertial frame and, as a result, are sampled by the same number of state vectors, are isolated. These trajectories are sorted by their initial value of x in the Earth-Moon pulsating rotating frame and then segmented into evenly sized partitions composed of no more than 10,000 members. This upper bound on the number of trajectories per partition is observed to produce fast clustering and analysis. Furthermore, sorting the trajectories ensures that trajectories that begin in a similar region of the solution space appear in the same partition.

Each partition is clustered independently to produce a set of local clusters. Following the two-step clustering procedure outlined earlier, the trajectories are grouped by their geometry. The ith partition is summarized by c_i local clusters, labeled as $\mathcal{L}^i = \{\mathcal{L}^i_1, \mathcal{L}^i_2, \ldots, \mathcal{L}^i_{c_i}\}$, and noise. Recall that any trajectories that are designated as noise are discarded from further consideration.

Step 3: Aggregate Clusters to Form Global Summary Local clusters that contain geometrically similar trajectories but exist across distinct partitions are merged. This aggregation procedure, which is typically performed in distributed clustering, is implemented using pairwise decisions for computational ease and accuracy. First, candidate pairs of local clusters for merging are coarsely identified using the feature vectors of their representative trajectories. Then, trajectories within these candidate pairs are clustered in a single step to determine whether the local clusters should be merged.

Due to the large number of trajectories sampled across the solution space and local clusters across all partitions, pairs of local clusters that may serve as candidates for merging are identified coarsely. First, the representative trajectory of the ith local cluster is compared to the representative trajectories of all other local clusters that complete the same number of half revolutions in the inertial frame. The k-nearest neighboring local clusters in each of the position-based and shape-based feature vector spaces then define the pairs of candidates for merging. In this paper, $k = 2$ to reduce the computational effort associated with cluster aggregation.

Each candidate pair of local clusters is assessed to determine whether they should be merged. Trajectories from each of the two local clusters are clustered together using the refinement procedure, modeled after a convoy detection scheme, as outlined earlier. If any trajectories from two distinct local clusters are grouped together, they are deemed geometrically similar. Accordingly, this pair of local clusters is marked for merging. For large clusters, particularly those composed of trajectories that complete a large number of revolutions around the Moon and remain tightly bounded, this process is implemented in a piecewise manner for computational efficiency. If a cluster possesses

more than 2,000 members, it is segmented into evenly sized subsets with fewer than 2,000 members. Then, each subset from the first local cluster is combined with each subset from the second local cluster. If trajectories in any combination of these smaller subsets are merged together, the entire clusters are merged.

The pairwise cluster aggregation decisions are used to generate global clusters across the entire dataset. This step is performed using a graph representation. Each node is defined as a local cluster within any partition. Then, undirected and unweighted edges are added between two nodes if their local clusters are marked for merging. A connected component in this graph defines the entire set of local clusters that are merged to form a single, global cluster. This cluster aggregation procedure results in a set of P global clusters $\mathcal{G} = [\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_P]$. At this step, a global cluster could be formed from one or more local clusters.

In this paper, aggregation is performed in two steps. First, local clusters are aggregated across a dataset generated with a single combination of $C_{J,des}$ and initial epoch. The aggregated clusters from each dataset are then further aggregated across distinct datasets to produce a single set of global clusters as one or both parameters vary.

To support visualization and analysis, a dendrogram representation of each cluster of geometrically distinct motion types is useful. This dendrogram is constructed using the cluster representatives of each global cluster that captures trajectories completing a specified number of half-revolutions in the inertial frame and described in the position-based feature vector space. At the base of each dendrogram or tree, the leaves correspond to each global cluster representative. Moving from the bottom of the dendrogram to the top, the branch associated with each leaf connects to the branch of its nearest neighboring global cluster representative at a height that is equal to their distance in the position-based feature vector space. This branching or merging process continues towards the top of the dendrogram until all global clusters are joined at a single tree root. Across the tree, longer branches indicate that the corresponding global cluster representatives possess a larger distance in the position-based feature vector space. Furthermore, groups of similar global clusters should be located in similar regions of the dendrogram, consistent with possessing a relatively small distance between their representatives. Through this graphical representation, each dendrogram coarsely summarizes the relationships between every global cluster corresponding to trajectories that complete the same number of half-revolutions in the inertial frame.

RESULTS

Ten datasets are generated with various combinations of the Jacobi constant used to define the initial state and initial epochs. In this paper, the Jacobi constants lie within the range $C_{J,des} \in [3.16, 3.19]$ and the initial epochs are selected as one of the following four values: January 9, 2025, 00:00.000 UTC, when the Moon is close to perigee; January 15, 2025, 19:55:12.0000 UTC; January 22, 2025, 15:50:24.0000 UTC, when the Moon is close to apogee; and January 29, 2025, 11:45:36.0000 UTC. The exact combinations of these values that are used to construct each dataset are listed in Table 1 along with the number of trajectories that appear in the final global cluster summary and the number of global clusters.

To support visualization and analysis, a dendrogram representation of the global clusters of geometrically distinct motion types is constructed. This dendrogram is constructed using the cluster representatives of each global cluster described in the position-based feature vector space across the first seven datasets listed in Table 1, composed of trajectories that begin from initial conditions with

Table 1. Generated trajectories and summary across various energy levels and epochs in the point mass ephemeris model

$C_{J,des}$	Initial epoch (UTC)	Number of trajectories in clusters	Number of global clusters
3.16	01/09/2025 00:00:00	1,757,699	1,239
3.165	01/09/2025 00:00:00	1,510,277	2,044
3.17	01/09/2025 00:00:00	1,311,936	1,718
3.175	01/09/2025 00:00:00	1,154,180	1,151
3.18	01/09/2025 00:00:00	1,030,481	930
3.185	01/09/2025 00:00:00	935,345	690
3.19	01/09/2025 00:00:00	841,367	719
3.16	01/15/25 19:55:12	1,668,494	3,141
3.16	01/22/2025 15:50:24	1,604,275	3,203
3.16	01/29/2025 11:45:36	1,675,383	2,799

Jacobi constants in the range $C_{J,des} \in [3.16, 3.19]$ and a single initial epoch. Figures 1-7 display the dendrograms for trajectories that complete one, two, three, four, six, eight, and ten half-revolutions in the inertial frame. In each figure, the dendrogram appears in the central part of the figure. Then, selected regions of the dendrogram are highlighted. From left to right, selected members of a global cluster that lies within these regions are plotted below the dendrogram in the rotating frame. In each figure, trajectories are colored on a blue to red scale by the Jacobi constant of their initial conditions: blue indicating $C_{J,des} = 3.19$ and red indicating $C_{J,des} = 3.16$. The large circle marker locates the initial condition whereas the Moon is plotted with a gray circle and the L_1 and L_2 equilibrium points from the circular restricted three-body problem drawn with red diamonds. The dendrograms in Figures 1-7 reflect the subtle and obvious geometric differences between trajectories in distinct global clusters.

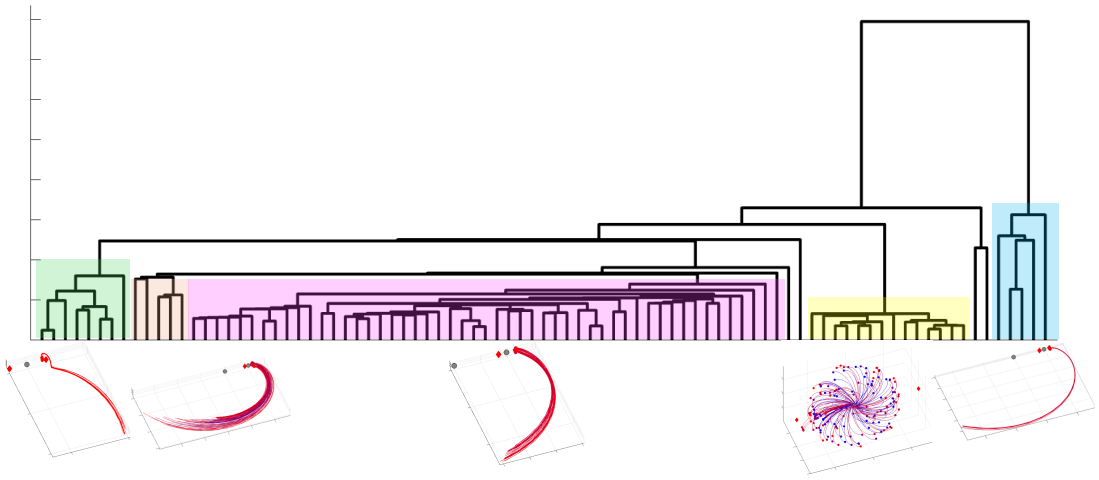

Figure 1. Dendrogram of global cluster representatives that complete up to half a revolution in the inertial frame.

Figure 1, for example, displays trajectories the complete up to one half-revolution in the inertial

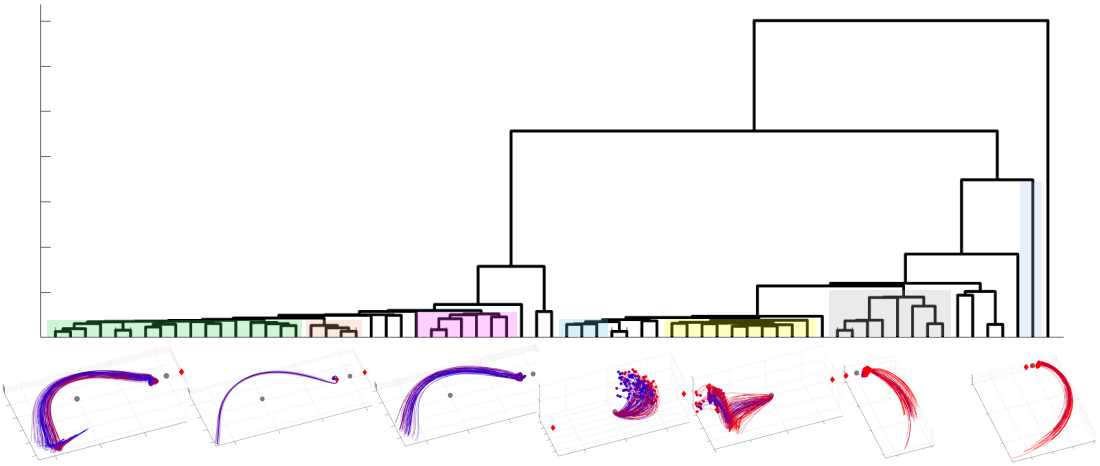

Figure 2. Dendrogram of global cluster representatives that complete up to two half-revolutions in the inertial frame.

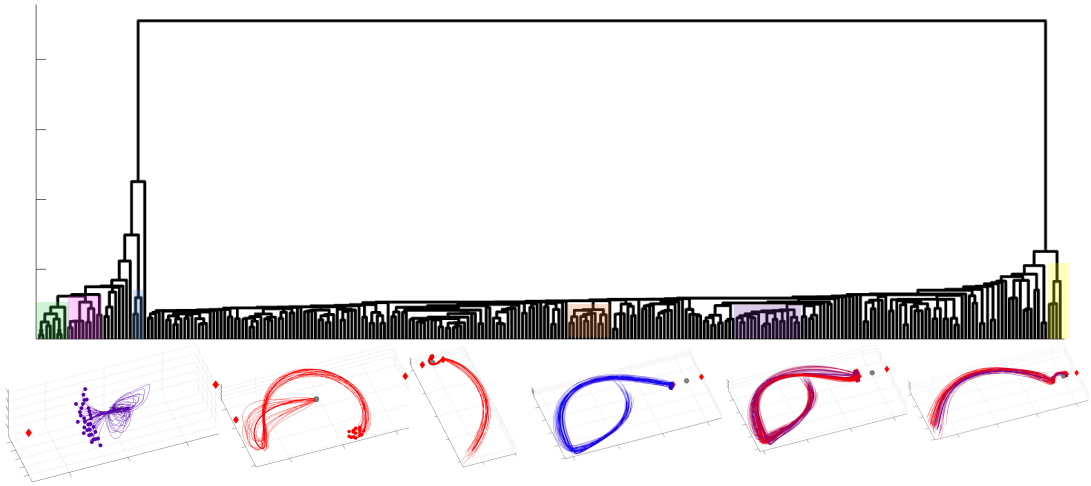

Figure 3. Dendrogram of global cluster representatives that complete up to three half-revolutions in the inertial frame.

frame. In the leftmost branches that are highlighted in green, trajectories revolve around L_2 before departing to the exterior region. Trajectories in the orange and purple highlighted regions directly depart the L_2 gateway with a similar geometry, consistent with their connecting branches possessing a shorter length. The trajectories in the yellow highlighted region impact the Moon quickly and possess a longer branch before connecting to the branches derived from other global clusters.

Similarly, Figure 3 displays trajectories the complete up to three half-revolutions in the inertial frame. In this example, the trajectories in global clusters within the green and magenta highlighted regions of the dendrogram revolve around the Moon prior to impact. In the middle blue highlighted region, trajectories revolve around the Moon before departing through the L_2 gateway. However, in the majority of the dendrogram, trajectories pass through the L_1 gateway before revolving around the Moon. These global clusters are connected by smaller branches, consistent with their high similarity. However, their connection to the global clusters of trajectories that revolve around the

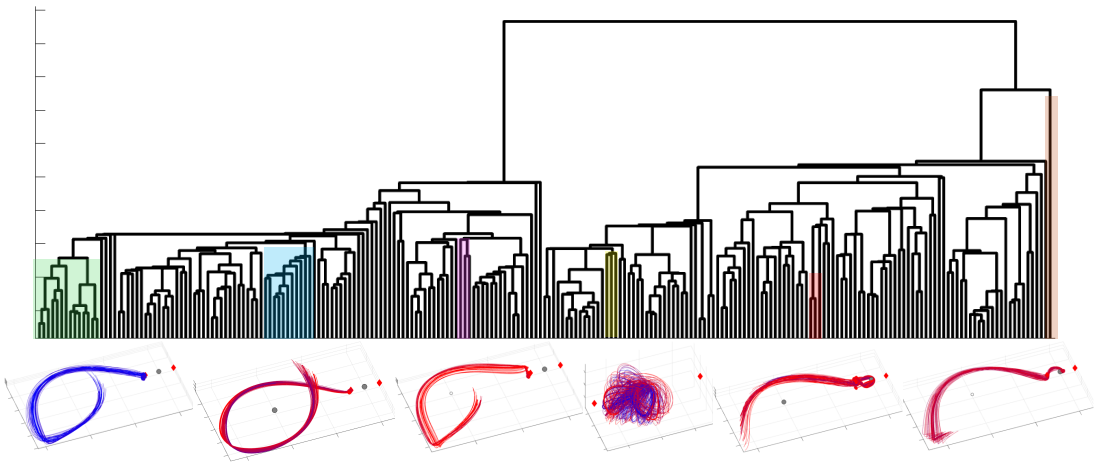

Figure 4. Dendrogram of global cluster representatives that complete up to four half-revolutions in the inertial frame.

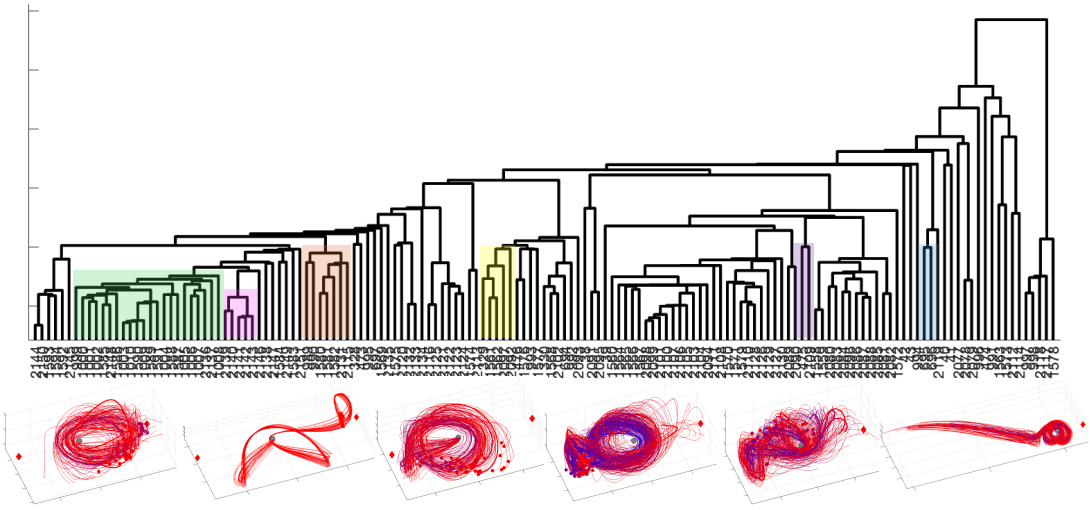

Figure 5. Dendrogram of global cluster representatives that complete up to six half-revolutions in the inertial frame.

Moon before impact or pass through the L_2 gateway are much longer, consistent with the substantial difference in their itinerary.

As the number of half-revolutions completed by trajectories in a global cluster increases, the dendrogram possesses shorter branch lengths. For instance, in Figures 6 and 7, the trajectories all revolve around the Moon. In this case, their itineraries are similar. However, the paths followed by geometrically distinct trajectories possess more subtle variations in geometry. In this case, classes or groupings of branches in the dendrogram are still visible.

CONCLUSIONS

This paper uses a distributed clustering framework to summarize natural trajectories for a spacecraft in a point mass ephemeris model of the Earth-Moon-Sun system. Each trajectory is generated

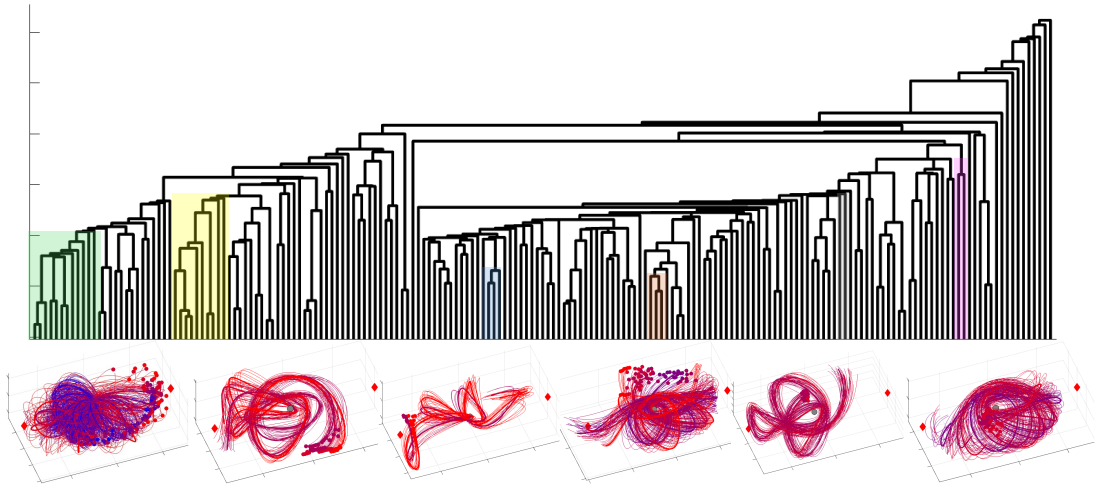

Figure 6. Dendrogram of global cluster representatives that complete up to eight half-revolutions in the inertial frame.

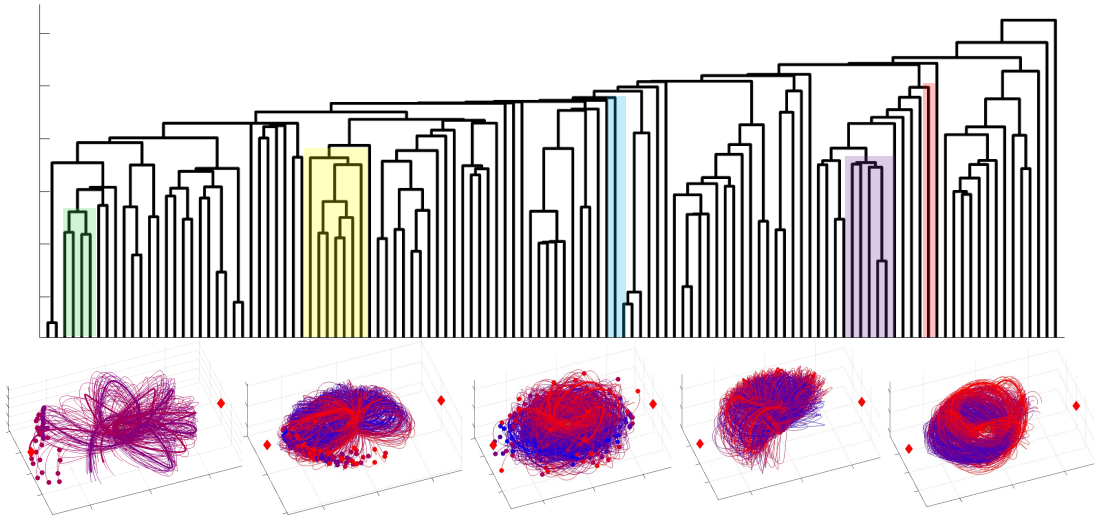

Figure 7. Dendrogram of global cluster representatives that complete up to ten half-revolutions in the inertial frame.

for up to 21 days, or until impact with a primary. Then, each trajectory is sampled according to its geometry in the inertial frame. Next, the trajectory is summarized by its shape and path in the configuration space, evaluated at these samples, but described in the rotating frame. This summary captured trajectories across a range of energy levels and initial epochs; at each combination of these values, the trajectories are summarized by 690-3,203 global clusters, each containing trajectories with a distinct geometry in the rotating frame. To visualize and analyze these global clusters, a dendrogram representation is employed. In these dendrograms, classes of trajectories with similar itineraries, but smaller variations in geometry, emerge through groups of branches.

ACKNOWLEDGMENT

This work was performed at the University of Colorado Boulder and supported by funding from the Air Force Research Laboratory through a cooperative agreement titled "Multi-Domain Awareness, Decision, and Exploitation".

REFERENCES

[1] R. Broucke, *Periodic Orbits in the Restricted Three-Body Problem With Earth-Moon Masses*, NASA Technical Report 82-1168, 1968.

[2] A. Leiva and C. Briozzo, *The Earth–Moon CR3BP: a full Atlas of low-energy fast periodic transfer orbits*, arXiv.org astro-ph/0612386, 2006.

[3] K. Onozaki, H. Yoshimura, and S. D. Ross, "Tube dynamics and low energy Earth–Moon transfers in the 4-body system," *Advances in Space Research*, vol. 60, no. 10, pp. 2117–2132, 2017.

[4] The Tabula Sapiens Consortium, "The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans," *Science*, vol. 376, no. 6594, 2022.

[5] The Tabula Muris Consortium, "A single-cell transcriptomic atlas characterizes ageing tissues in the mouse," *Nature*, vol. 583, pp. 590–595, 2020.

[6] Elaine H. Shen and Caroline C. Overly and Allan R. Jones, "The Allen Human Brain Atlas," *Trends in Neurosciences*, vol. 35, no. 12, pp. 711–714, 2012.

[7] N. Bosanac, "Data-Mining Approach to Poincaré Maps in Multi-Body Trajectory Design," *Journal of Guidance, Control, and Dynamics*, vol. 43, no. 6, 2020.

[8] N. Bosanac, "Data-Driven Summary of Motion in an Ephemeris Model of Cislunar Space," in *AAS/AIAA Space Flight Mechanics Meeting*, Kaua'i, HI, Jan. 2025.

[9] N. Bosanac, "Clustering Natural Trajectories in the Earth-Moon Circular Restricted Three-Body Problem," *Under Review*, 2025.

[10] D. Folta, N. Bosanac, I. Elliott, L. Mann, R. Mesarch, and J. Rosales, *Astrodynamics Convention and Modeling Reference for Lunar, Cislunar, and Libration Point Orbits (Version 1.1)*, NASA/TP–20220014814, 2022.

[11] G. Petit and B. Luzum (eds), *IERS Conventions (2010)*, International Earth Rotation and Reference Systems Service, Technical Note 36, 2010.

[12] R. Park, W. Folkner, J. Williams, and D. Boggs, "The JPL Planetary and Lunar Ephemerides DE440 and DE441," *The Astronomical Journal*, vol. 161, no. 3, 2021.

[13] NASA Goddard Space Flight Center, *General Mission Analysis Tool Version R2020a: Mathematical Specifications*, 2020.

[14] C. Acton, "Ancillary Data Services of NASA's Navigation and Ancillary Information Facility," *Planetary and Space Science*, vol. 44, no. 1, pp. 65–70, 1996.

[15] D. Vallado, *Fundamentals of Astrodynamics and Applications: Fourth Edition*. Hawthorne, CA: Microcosm Press, 2013.

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

[17] N. Patrikalakis, T. Maekawa, and W. Cho, *Shape Interrogation for Computer Aided Design and Manufacturing*. 2009.

[18] K. Wardle, *Differential Geometry*. Mineola, NY: Dover Publications, Inc., 2008.

[19] J. Han and M. Kamber, *Data Mining: Concepts and Techniques, 2nd ed.* New York, NY: Proquest EBook Central: Elsevier Science and Technology, 2006, ch. 7.

[20] M. Ester, H. Kriegel, J. Sander, and X. Xu, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996.

[21] R. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Advances in Knowledge Discovery and Data Mining*, J. Pei, V. Tseng, L. Cao, H. Motoda, and G. Xu, Eds., Heidelberg: Springer, Berlin, 2013.

[22] C. Malzer and M. Baum, "A Hybrid Approach To Hierarchical Density-Based Cluster Selection," in *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2020, pp. 223–228.

[23] R. Campello, D. Moulavi, A. Zimek, and S. Sander, "Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection," *ACM Transactions on Knowledge Discovery from Data*, vol. 10, no. 1, pp. 1–51, 2015.

[24] GNU, *GNU Scientific Library (GSL)*, 2025.

[25] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes: The Art of Scientific Computing, Third Edition*. New York, NY: Cambridge University Press, 2007.