

REPRESENTATION OF PROBABILITY DENSITY FUNCTIONS FROM ORBIT DETERMINATION USING THE PARTICLE FILTER

Alinda K. Mashiku⁽¹⁾, James Garrison⁽²⁾, and J. Russell Carpenter⁽³⁾

⁽¹⁾⁽²⁾Purdue University, 701 W. Stadium Ave, West Lafayette, IN 47906, (765-496-2396, 765-496-7482), (aaligawe, jgarriso@purdue.edu)

⁽³⁾NASA Goddard Space Flight Center, Greenbelt, MD 20771, 301-286-7526, james.r.carpenter@nasa.gov

Abstract: *Statistical orbit determination enables us to obtain estimates of the state and the statistical information of its region of uncertainty. In order to obtain an accurate representation of the probability density function (PDF) that incorporates higher order statistical information, we propose the use of nonlinear estimation methods such as the Particle Filter. The Particle Filter (PF) is capable of providing a PDF representation of the state estimates whose accuracy is dependent on the number of particles or samples used. For this method to be applicable to real case scenarios, we need a way of accurately representing the PDF in a compressed manner with little information loss. Hence we propose using the Independent Component Analysis (ICA) as a non-Gaussian dimensional reduction method that is capable of maintaining higher order statistical information obtained using the PF. Methods such as the Principal Component Analysis (PCA) are based on utilizing up to second order statistics, hence will not suffice in maintaining maximum information content. Both the PCA and the ICA are applied to two scenarios that involve a highly eccentric orbit with a lower a priori uncertainty covariance and a less eccentric orbit with a higher a priori uncertainty covariance, to illustrate the capability of the ICA in relation to the PCA.*

Keywords: *Orbit Determination, Particle Filter, Non-Gaussian, Data Compression, Nonlinear Estimation.*

1. Introduction

The Statistical Orbit Determination (OD) problem involves the estimation of the states of a space object based on noisy measurements. In recent years, there has been a dramatic increase of space objects (assets and space debris) particularly in Low Earth Orbit (LEO) [?]. This increase of space objects pose a threat to the space assets based on potential collisions as well as the increase in operational costs during maneuvers required to avoid collisions. For instance, in 2009, Celestrak predicted that an Iridium Satellite and the defunct Russian Satellite Cosmos, would have a close approach of 584 meters [?, ?], nevertheless they collided with one another. Since 1998, the ISS has performed 13 maneuvers to avoid collisions with space debris [?]. The last maneuver was in 2009, performed to avoid the debris from the Iridium and Cosmos satellite collisions. This required 30 hours to plan and execute [?] as well as the cost of propellant for the delta-V maneuver implemented. The classical methods of statistical OD, such as the Extended Kalman Filter (EKF) [?] and the Unscented Kalman Filter (UKF) [?], may not always be the best choice of an estimator for all problems because they only consider the second moments of the state. As long as the observation error and process noise can be accurately assumed to have a Gaussian distribution, these second moments are sufficient to infer all other statistics as well. One example in which the non-Gaussian errors could arise is when observations are based upon short arcs of tracking data when tracking

orbit debris. Hence, an accurate prediction of the states over longer durations of time is imperative for such tracking and collision avoidance applications.

A good prediction of not only the mean state, but also the shape of the state distribution is important for the orbit debris problem, since even events with very low probabilities are of concern. To accurately predict the likelihood of these low-probability events, we must accurately predict the tails of the distribution. Given that the orbit dynamics are non-linear, even if the a priori state vector has a Gaussian distribution, propagation with non-linear dynamics would evolve this distribution into a non-Gaussian one. In addition to predicting the state distribution, there also needs to be a method of numerically representing this distribution in order to generate a catalog of all orbiting objects for predicting the likelihood of possible collisions. At present, orbit catalogs use some definition of a common ephemerides, for example the NORAD Two-line elements ephemerides, which only represent the mean orbit state, providing no information on its probability distribution. Development of Nonlinear filtering techniques capable of predicting the state distribution (a full Probability Density Function (PDF)), are necessary to address this type of problem. The Particle Filter algorithm (PF) is used as a nonlinear filtering technique, based on a sequential Monte Carlo approach representing the required state PDF by a set of random samples or particles. The accuracy of the prediction and estimation of the states increases with the increase in the number of particles.

Having developed a method for predicting the state PDF, it is then necessary to derive a compact form for storing and distributing the distribution. Representing the state PDF of an orbit with a large number of particles is not a practical method for storing or distributing orbit state data. The objective of this work, therefore, is to develop new methods for representing the full PDF of the orbit state, in a compact data record which could be distributed much in the same way as ephemerides are used today. We will approach this problem by investigating methods to decorrelate the the state into independent uni- or multi-variate components, whose PDFs could be more compactly represented. For example, some components of the decorrelated PDF might be accurately represented by many fewer particles than the original state PDF, or might be easier to approximate with continuous functions using methods such as the wavelet transform, characteristic functions, and/or kernel densities. Conventional methods that perform orthogonal transformations to obtain linearly uncorrelated variables, such as the Principal Component Analysis (PCA), would lose information present in the higher moments if the distribution is not Gaussian. In this paper, we will apply a method known as the Independent Component Analysis (ICA) [?, ?, ?, ?] method, which is capable of decorrelating non-Gaussian data by finding the local extrema of the kurtosis (fourth order moment) of a linear combination of the states and thus estimating the non-Gaussian independent components.

Methods such as PCA and ICA may also offer additional compression from dimensional reduction in some cases. For example, the OD process may require the augmentation of the state with additional components that are not strongly correlated with the minimal subset of components required for accurate state prediction. PCA and ICA would be expected to assign relatively lower eigenvalues to components arising from such states, identifying them as good candidates for elimination. Furthermore, the OD process may use a state vector whose coordinate representation is well-suited to OD, but less suitable for long-term propagation. The transformation of coordinates inherent in PCA and ICA may help to identify alternative representations, analogous to orbital elements, which

have a simpler PDF structure, and hence be more easily predicted.

To demonstrate our approach, we apply ICA and PCA to both decorrelate and dimensionally reduce Cartesian particle representations of the PDF of a two-body orbit. Since out-of-plane motion in such orbits is known to be decoupled from in-plane motion, we anticipate a dimensional reduction from 6 states to 4 may reasonably approximate the most strongly non-Gaussian components of the PDF. We will show that the resulting compression of the original 6-dimensional Cartesian set of 1000 particles can be used to accurately reconstruct the original PDF, at the epoch at which we perform the translation. The ability of the compressed representation to form the basis for an accurate prediction is the topic of ongoing work, and will not be addressed in this paper.

This paper is organized into eight sections. Section 1 covers the introduction, section 2 describes the Particle Filter algorithm. In sections 3 and 4, the methods of dimensional reduction and the problem description are presented. The results are presented in section 5 and the conclusion and future work are presented in section 6. Section 7 will constitute the acknowledgements and the references are listed in section 8.

2. Particle Filter

In order to incorporate the higher order moments of the state and measurement errors, we need to be able to study the evolution of the full state PDF in the OD process. The Particle Filter (PF) is a simulation-based filter based on the sequential Monte Carlo approach, effective in estimating the full state PDF of nonlinear models or non-Gaussian PDFs [?, ?, ?]. The central idea of a PF is to represent the required probability density function (PDF) by a set of $N \gg 1$ random samples known as particles $\{\mathbf{X}_k^{(i)}\}_{i=1}^N$, with associated weights $\{w_k^{(i)}\}_{i=1}^N$, such that

$$p(\mathbf{X}_k|\mathbf{Y}_k) \approx \sum_{i=1}^N w_k^{(i)} \delta(\mathbf{X}_k - \mathbf{X}_k^{(i)}) \quad (1)$$

where the particles are identically and independently distributed (i.i.d) samples drawn from the initial PDF with an initial Gaussian distribution assumption. The weights are the probability values that are the approximations of the relative densities of the particles such that their sum equals one [?, ?, ?]. The state estimate is given as a conditional density called the posterior density $p(\mathbf{X}_k|\mathbf{Y}_k)$. Estimators based on this posterior density are constructed from Bayes' theorem, which is defined as

$$p(\mathbf{X}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})}{p(\mathbf{Y})} \quad (2)$$

where $p(\mathbf{X})$ is called the prior density (before measurement), $p(\mathbf{Y}|\mathbf{X})$ is called the likelihood and $p(\mathbf{Y})$ is called the evidence (normalizes the posterior to assure the integral sums up to unity) [?, ?]. In the execution of the general PF algorithm, a common effect ensues where the variance of the weights increases over time. This phenomenon is known as sampling degeneracy, in that there is an insufficient number of effective particles required to fully capture the PDF. A way of overcoming the sampling degeneracy is to derive the importance distribution that minimizes the variance of the distribution of the weights. The corresponding weights are given by:

$$w_k = w_{k-1}p(\mathbf{Y}_k|\mathbf{X}_{k-1}) \quad (3)$$

$$w_k = w_{k-1} \int p(\mathbf{Y}_k|\mathbf{X}_k)p(\mathbf{X}_k|\mathbf{X}_{k-1})d\mathbf{X}_k \quad (4)$$

This means that the weights can be calculated before the particles are propagated to time k . From the equations above, we see that there is a problem in the sense that the evaluation of the integral generally does not have an analytic form and that the particles' weight updates are based on the knowledge of the state at the past time step and the current measurement only i.e. $p(\mathbf{X}_k|\mathbf{X}_{k-1}, \mathbf{Y}_k)$. So, a reasonable choice for the importance distribution is the transition prior $p(\mathbf{X}_k|\mathbf{X}_{k-1})$ [?].

2.1. PF Algorithm

Let $\mathcal{X}_k = (\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_k)$ and $\mathcal{Y}_k = (\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_k)$ be the stacked vectors of states and observations up to time k and $\{w_k^i, i = 1, 2, \dots, N\}$ represent the weights of the N particles at each time k . The estimation process can be sectioned into three parts: initialization, time update or prediction and the measurement update. Each step is described below for the PF implementation.

1. Initialization at $k = 0$, for $i = 1, \dots, N$

- Sample N particles from the prior $\mathbf{X}_0^i \sim p(\mathbf{X}_0)$
- Calculate the weights of the particles from the initial distribution (assume Gaussian)
 $w_0^i = p(\mathbf{Y}_0|\mathbf{X}_0^i)$
- Normalize the weights $w_0^i = \frac{w_0^i}{w_T}$ using total weight $w_T = \sum_i^N w_0^i$

2. Time Update or Prediction at $k \geq 1$, for $i = 1, \dots, N$

- Predict the particles through the dynamics

$$\mathbf{X}_k^i = f(\mathbf{X}_{k-1}^i) + v_k \quad (5)$$

$$p(\mathbf{X}_k|\mathcal{Y}_{k-1}) = \int p(\mathbf{X}_k|\mathbf{X}_{k-1})p(\mathbf{X}_{k-1}|\mathcal{Y}_{k-1})d\mathbf{X}_{k-1} \quad (6)$$

3. Measurement Update

- Update the weights using the innovation from the measurements assuming that the measurements are Gaussian distributed

$$w_k^i = w_{k-1}^i \frac{p(\mathbf{Y}_k|\mathbf{X}_k^i)p(\mathbf{X}_k^i|\mathbf{X}_{k-1}^i)}{q(\mathbf{X}_k^i|\mathcal{X}_{k-1}^i, \mathcal{Y}_k)} \quad (7)$$

- We assume that the importance distribution $q(\mathbf{X}_k^i|\mathbf{X}_{k-1}^i, \mathcal{Y}_k)$ in this case is equal to the prior density $p(\mathbf{X}_k^i|\mathbf{X}_{k-1}^i)$ [?], so that,

$$w_k^i = w_{k-1}^i \times p(\mathbf{Y}_k|\mathbf{X}_k^i) \quad (8)$$

- Normalize the weights $w_k^i = \frac{w_k^i}{w_T}$ using total weight $w_T = \sum_i^N w_k^i$
- The final posterior density estimate $p(\mathbf{X}_k|\mathbf{Y}_k) = \sum_{i=1}^N w_k^{(i)} \delta(\mathbf{X}_k - \mathbf{X}_k^{(i)})$

Due to the fact that the prior density is usually broader than the likelihood, only a few particles will be assigned higher weight during the measurement update step causing sampling degeneracy. The solution to overcome sampling degeneracy is to resample the particles. Let N_{eff} denote the effective sample size, which can be estimated by [?, ?]: $\hat{N}_{eff} = \frac{1}{\sum_i (\hat{w}_k^i)^2}$. Also let N_{th} denote a lower

threshold for the effective sample size, which is arbitrarily chosen with respect to the accuracy desired; the larger the threshold, the more accurate the PDF results and the more computational cost incurred.

If $N_{eff} > N_{th}$, then the sampling degeneracy is not low enough, the filter continues on to step 2 for the next time update. Otherwise, we replicate the particle with the highest weight to replace the particles falling below the threshold and then renormalize the weights. However, resampling does have its own shortcomings, since the particles that have higher weights are statistically selected each time thus losing the diversity of the samples. This loss of diversity may also cause divergence of the estimates. To avoid the loss of diversity the replicated particles are “jittered/spread-out” by adding process noise to spread the resampled particles [?, ?].

3. Decorrelation, Scaling, and Dimensional Reduction

The PF requires the implementation of a large number particles to achieve a good PDF representation. In addition, the computational and data allocation costs compound with increase in both the number of particles and the dimension of the state. We propose to compress the data with respect to the dimensional size and the number of particles used. However, since our data does not possess a Gaussian distribution, conventional methods that perform orthogonal transformations to obtain linearly uncorrelated variables, such as the principal component analysis, will contribute to the loss of information present in the higher moments. Hence, we propose to apply the method of Independent Component Analysis (ICA). ICA has been developed to be capable of decorrelating non-Gaussian data by finding the local extrema of the kurtosis (fourth order moment) of a linear combination of the states and thus providing the non-Gaussian mutually independent components [?, ?, ?].

Since the states have different scaling in order of magnitudes (i.e. position vs velocity), we normalize the particles for our states based on the canonical units. For distance, we scale the value by the distance unit ($1 DU_{\oplus}$) which is equivalent to the value of the radius of the Earth (6378.145km). For velocity, the scaling metric is given as the distance unit/ time unit ($1 DU_{\oplus}/TU_{\oplus}$) of equivalency 7.90536828 km/s [?]. The implementation of the scaling will help to ensure an equal weighting of information from the velocities that are typically of lower magnitudes.

As an introduction to ICA, we will first review Principal Component Analysis (PCA), commonly used to decorrelate data with a Gaussian distribution.

3.1. Principal Component Analysis

Principal Component Analysis (PCA) is the orthogonal transformation of a set of possibly correlated data into a set of linearly uncorrelated variables known as the principal components [?, ?]. In other words, PCA finds a smaller set of variables with less redundancy [?]. The number of the principal components obtained are usually equal to or less than the dimension of the original data set. The component with the largest variance is deemed as the first component and each succeeding component with a higher variance is computed in that order under the constraint that it be orthogonal to the preceding components [?, ?].

The principal components are obtained by first centering the data set (i.e. state particles \mathbf{X}_k^i) and subtracting its mean to obtain the standardized data set $\tilde{\mathbf{x}}_k^i$

$$\tilde{\mathbf{x}}_k^i = \mathbf{X}_k^i - E[\mathbf{X}_k^i] \quad (9)$$

The eigenvectors \mathbf{V} corresponding to the decreasing eigenvalues \mathbf{D} , represent the direction of the components in order of their decreasing variances.

$$\mathbf{C}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} = E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} \quad (10)$$

$$\mathbf{D} = \mathbf{V}^{-1}\mathbf{C}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{V} \quad (11)$$

In PCA, the components with negligible collective eigenvalues can be considered as redundant or pertaining to negligible information. Thus, the principal components kept are in reference to the directions with the maximum variance and are considered sufficient to represent the original data in a new coordinate frame within a reasonable level of accuracy.

The shortcomings of using the PCA for a non-Gaussian distributed data is that the principal components computed are based on the covariance matrix, which only incorporates second order moments statistics. Hence using the principal components obtained by PCA defeats the purpose of implementing the PF to obtain the non-Gaussian distributed particles of our state estimates.

3.2. Independent Component Analysis

Independent Component Analysis (ICA) is a signal processing technique that is used to represent a set of random variables as linear combinations of statistically independent component variables [?, ?, ?, ?, ?, ?]. The most common demonstration of the application of the ICA is known as the ‘cocktail party problem’. Given a room with three people speaking at the same time to an audience and three microphones recording the mixture of the three speech signals $x_1(t)$, $x_2(t)$, and $x_3(t)$. Each of these recorded signals can be represented as a weighted sum of the speech signals emitted by the three speakers, denoted by $s_1(t)$, $s_2(t)$, and $s_3(t)$. This is expressed as a linear equation:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \quad (12)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \quad (13)$$

$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) \quad (14)$$

where a_{ij} with $i, j = 1, \dots, 3$ are some parameters that depend on the distances of the microphones from the speakers [?, ?]. It would be easy to solve for the speech signals $s_1(t)$, $s_2(t)$, and $s_3(t)$ if we knew a_{ij} , but neither of them are known. However, if we can assume that the signals $s_1(t)$, $s_2(t)$, and $s_3(t)$ are statistically independent at each time instant t , we can use ICA to solve for a_{ij} and in turn solve for $s_1(t)$, $s_2(t)$, and $s_3(t)$.

Beyond the assumption that the signals are statistically independent, the independent components must have non-Gaussian distributions in order for higher order cumulants or moments to be useful in the estimation of the independent components [?, ?]. For Gaussian distributions, the higher order moments are zero. Given that our state estimates using the PF are non-Gaussian, we expect some benefit from implementing the ICA for state decorrelation and dimensional reduction.

Given that at a time k we have an estimated a state vector \mathbf{X}_i of dimension- N , with m particles $\mathbf{X}_i = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)$, we form a linear combination of S -dimensional independent variables $\mathbf{s}_j = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m)$, where $S \leq N$; we then compute a mixing matrix $\mathbf{A}(N \times S)$, such that

$$\mathbf{X} = \mathbf{A}\mathbf{s} \quad (15)$$

where $\forall j, \mathbf{s}_j$ has unit variance. The data \mathbf{X} is pre-whitened to obtain a new set of data $\mathbf{x} = \mathbf{M}\mathbf{X}$ that is uncorrelated and has a unit variance, with \mathbf{M} known as the whitening matrix. We compute the orthonormal matrix \mathbf{B} whose columns w_1, w_2, \dots, w_s are in the direction of the maximum or minimum kurtosis of \mathbf{x} (see Fig. 1).

$$\text{kurt}(w^T \mathbf{x}) = E\{(w^T \mathbf{x})^4\} - 3[E\{(w^T \mathbf{x})^2\}]^2 \quad (16)$$

$$= E\{(w^T \mathbf{x})^4\} - 3\|w\|^4 \quad (17)$$

The columns w_1, w_2, \dots, w_s are solved using the objective function $\mathcal{J}(w)$ under the constraint that $\|w\| = 1$,

$$\mathcal{J}(w) = E\{(w^T \mathbf{x})^4\} - 3\|w\|^4 + F(\|w\|^2) \quad (18)$$

where F is a penalty term due to the constraint. In order to solve for the columns w_1, w_2, \dots, w_s , rapidly, a fixed-point iteration scheme is generated by taking the expectation of Eq. 18, equating it to zero and normalizing the penalty term. We obtain

$$w = \alpha \times (E\{\mathbf{x}(w^T \mathbf{x})^3\} - 3\|w\|^2 w) \quad (19)$$

where α is an arbitrary scalar that diffuses under the normalizing constraint of w .

Hence, the fixed-point iteration scheme is implemented as follows:

1. Take a random initial vector $w(0)$ or norm 1. Let $k = 1$.
2. Let $w(k) = E\{\mathbf{x}(w(k-1)^T \mathbf{x})^3\} - 3w(k-1)$.
3. Divide $w(k)$ by its norm
4. If $|w(k)^T w(k-1)|$ is not close enough to 1, let $k = k + 1$ and go back to step 2, or else output vector w_1 , and continue until w_s is obtained.

The basic idea is that we search for the vector w of norm 1, such that the linear combination of $w^T x$, has maximum or minimum kurtosis. That will give us the independent components \mathbf{s} [?]. Moreover, the number of the independent components desired is declared beforehand, hence the search for the columns of the matrix \mathbf{B} , is maximized based on the stated number of independent components.

A detailed description of the algorithm implementation can be found in references [?, ?]. Hence, our independent components \mathbf{s} are given by

$$\mathbf{x} = \mathbf{M}\mathbf{X} = \mathbf{M}\mathbf{A}\mathbf{s} \quad (20)$$

$$\mathbf{B} = \mathbf{M}\mathbf{A} \quad (21)$$

$$\mathbf{s} = \mathbf{B}^T \mathbf{x} \quad (22)$$

One ambiguity in the ICA method is that the order of the independent components is indeterminate. Because both \mathbf{A} and \mathbf{s} are unknown, the order of the sums in Eq. 12, Eq. 13 and Eq. 14, can be interchanged and hence any independent component can be called the first one [?]. For our application, this is not a hindrance since our goal is to obtain the independent components that are dimensionally reduced and that could be compressed univariately or bivariately.

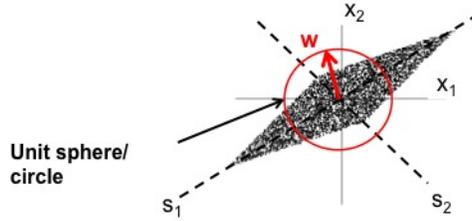


Figure 1. Search for projections that correspond to points on the unit circle, using whitened mixtures of uniformly distributed independent components [?]

4. Problem Description

Given the state estimate particles \mathbf{X}_k^i , we are interested in obtaining the independent components that incorporate higher order moments beyond the second order statistics given by the covariance matrix that is usually implemented using the PCA. We will look at two orbit scenarios and compare the results of reducing the state dimension through using the PCA and the ICA. In both cases, we will reduce the dimensions from a 6-dimensional vector to a 4-dimensional state vector, each using 1000 particles with their respective weights. Retaining the information from the eigenvector matrix and the mean of the data matrix for the PCA, we compute the reconstructed data with respect to the dimensionally reduced principal components. The distributions reconstructed from the ICA and that reconstructed from the PCA will each be compared to the original distribution of the 6-dimensional state vector.

4.1. Scenario 1

For the first scenario, we will be looking at a satellite with a very large eccentricity of 0.8181 and a 24hr orbital period (see Fig. 2). The initial conditions were [7488.36km, 71793.70km, 24219.13km, -0.9275km/s, -0.0257km/s, 0.363km/s], defined as apogee given in the Earth Centered Inertial (ECI) frame in Cartesian coordinates. The standard deviation for the position and the velocity states for the a priori uncertainties were 1m and 1mm/s respectively. The satellite has four opportunities of measurements from 2 Deep Space Network ground stations located at Canberra, Australia and Madrid, Spain for a duration of 75 minutes each, using the Satellite Toolkit (STK) software for a simulated realistic scenario. The last two measurement opportunities are obtained from a space-based observation network called TDRSS (Tracking and Data Relay Satellite System) for a duration of 15 minutes each. NASA Goddard Space Flight Center's Orbit Determination Toolbox (ODTBX) [?] was used in conjunction with the developed PF code and STK. The ODTBX has various capabilities such as implementing ground station and space based measurements as well as sophisticated plotting capabilities.

In Fig. 3, the distribution of the particles for the position states are illustrated in planar X-Y and Y-Z views at apogee after one orbital period. Each particle has its own corresponding weight. The histogram illustrates the scattering of the particles, especially at the tails of the distribution.

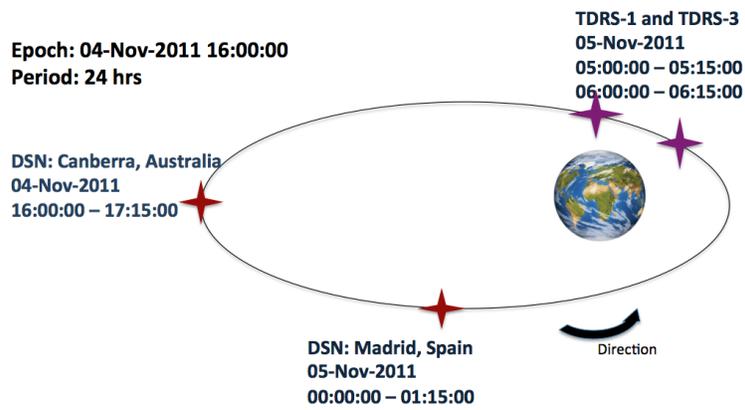


Figure 2. Illustration of the orbit trajectory for scenario 1 during the first orbital period

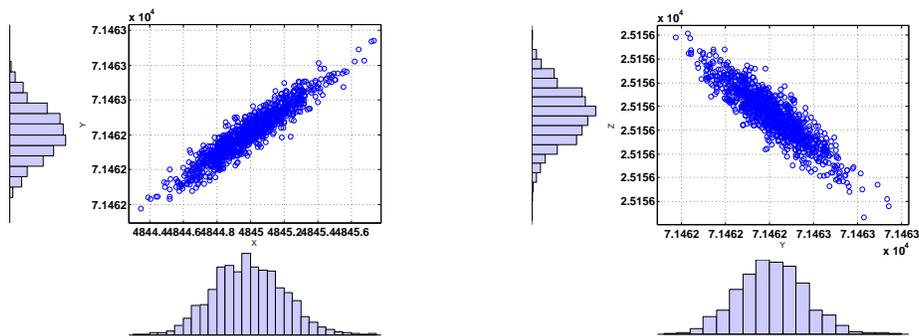


Figure 3. Scenario 1: Scatter and Histogram plot showing the distribution of the particles of the position state vector for Scenario 1 in the X-Y plane(left) and Y-Z plane (right) (All units are in km).

4.2. Scenario 2

For the second scenario, we will look at an orbit of eccentricity 0.2 and an approximate 0.75 day orbital period. The standard deviation for the position and the velocity states for the a prior uncertainties were 1km and 1m/s respectively. The semi-major axis is 35000km, with an initial condition of [28000km, 0km, 0km, 0km/s, 4.133km/s, 0km/s] at perigee. In this scenario, we are interested in observing the evolution of the particles with a larger initial uncertainty that are propagated over the entire orbital period without any measurement update, given a smaller eccentricity. This will illuminate how the region of uncertainty grows over time.

In Fig. 4, the distribution of the particles for the position states are illustrated in planar X-Y and Y-Z views after one orbital period at perigee. Each particle also has its own corresponding weight. The histogram illustrates how heavily skewed the distribution of the particles in the X-Y plane become. This plane is also along the trajectory of the orbit, which illustrates the “banana-shaped” non-Gaussian region of uncertainty. In observing the Y-Z plane, despite that there is no motion in the Z-direction, the uncertainties present are caused by the propagation of the initial particles defined by the region of uncertainty, as well as unmodeled perturbations that are added as process noise.

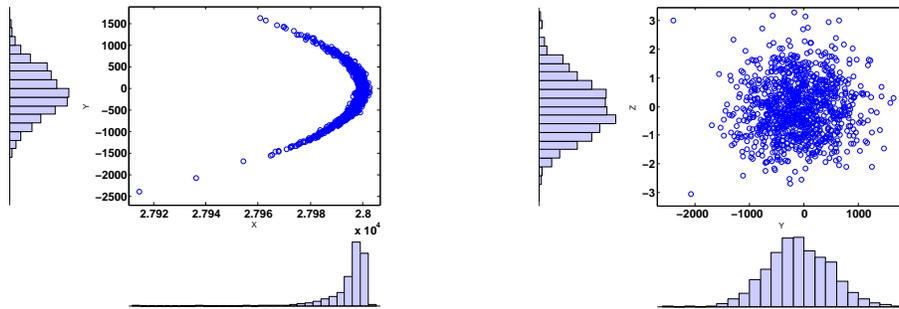


Figure 4. Scenario 2: Scatter and Histogram plot showing the distribution of the particles of the position state vector for Scenario 2 in the X-Y plane (left) and Y-Z plane (right) (All units are in km).

5. Results

In this section, we will look at how accurate the reconstructed states are from the principal components and the independent components. The mean square error is computed between the original

data and the reconstructed data. The mean square error is computed as follows:

$$MSE = \frac{\|\mathbf{X}_{\text{original}}^i - \mathbf{X}_{\text{reconstructed}}^i\|^2}{N} \quad (23)$$

The MSE illuminates the overall accuracy of the errors in the reconstructed data. However, in order to accurately illustrate the goodness-of-fit of the distributions, we will look at the Kolmogorov-Smirnov test statistic.

The Kolmogorov-Smirnov (K-S) test is a nonparametric test that can be used to compare two sample distributions by quantifying the distance between the empirical cumulative distribution functions [?, ?]. The K-S test is useful in comparing two sample distributions as it is sensitive in both the location and shape of the cumulative distribution functions. The K-S test statistic D_{nm} for two samples distributions F_n and G_m is given as:

$$D_{nm} = \sup_{\mathbf{X}} |F_n(\mathbf{X}) - G_m(\mathbf{X})| \quad (24)$$

For our cases, both n and m have 1000 samples, and the distance metric D_{nm} is a scalar value such that the closer it is to zero, the more close the distributions are to each other.

5.1. Scenario 1: PCA vs ICA

In Scenario 1, the reconstructed data for the positional states using the ICA were fully reconstructed with a lesser error compared to the positional reconstruction by PCA (see Fig. 5). However, for the velocity states, we found that the PCA slightly outperformed the ICA based on the values of the mean square errors and the K-S test statistic values in the x and y components, as shown in Tab. 1 and Tab. 2 respectively. Since the ICA computes the independent components based on the maximum or minimum kurtosis of the whitened data, the independent components calculated are based on the maximum kurtosis values of the data. For the PCA, based on the 4 principal components obtained, the kurtosis for these components were 3.13, 2.86, 2.92 and 3.01 while for the ICA the kurtosis for the 4 independent components were 3.27, 3.13, 3.12 and 2.9. It is stated that maximizing non-Gaussianity implies independence [?], hence the higher the values of kurtosis, the more independent the components.

Table 1. Scenario 1: Mean Square Errors for PCA and ICA

MSE	PCA	ICA
$X(DU_{\oplus}^2)$	0.5337×10^{-11}	0.0001×10^{-13}
$Y(DU_{\oplus}^2)$	0.0319×10^{-11}	0.0001×10^{-13}
$Z(DU_{\oplus}^2)$	0.0054×10^{-11}	0.4822×10^{-13}
$V_x(DU_{\oplus}^2/TU_{\oplus}^2)$	0.0001×10^{-14}	0.0463×10^{-13}
$V_y(DU_{\oplus}^2/TU_{\oplus}^2)$	0.0018×10^{-11}	0.0024×10^{-13}
$V_z(DU_{\oplus}^2/TU_{\oplus}^2)$	0.0001×10^{-11}	0.1545×10^{-13}

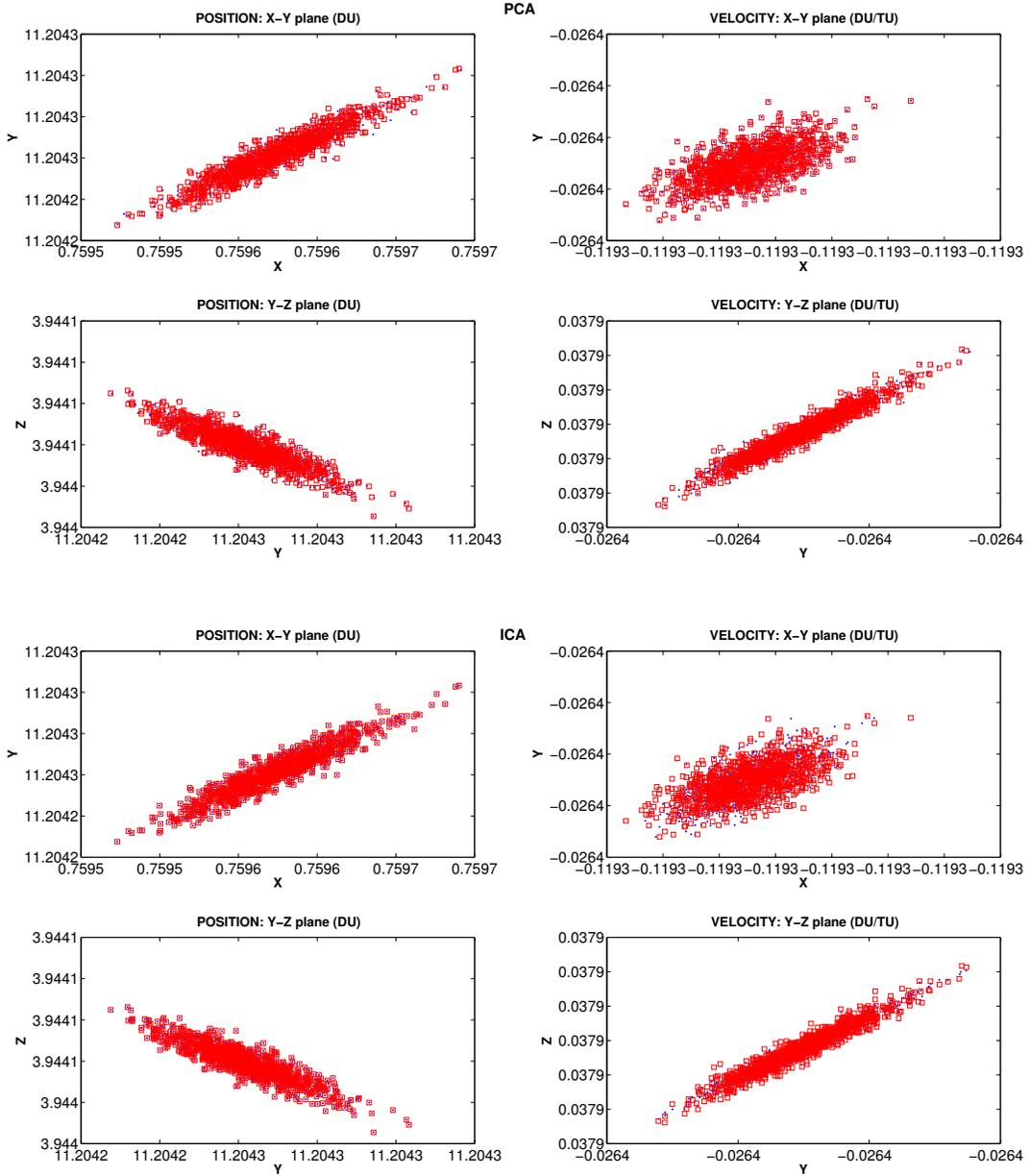


Figure 5. Scenario 1 PCA vs ICA: Original Data (Red Squares) and Reconstructed Data (Blue Dots)

Table 2. Scenario 1: Kolmogorov-Smirnov test statistic for PCA and ICA

K-S statistic	PCA	ICA
X	0.0150	0.0030
Y	0.0110	0.0020
Z	0.0080	0.0020
V_x	0.0060	0.0460
V_y	0.0130	0.0060
V_z	0.0110	0.0240

5.2. Scenario 2: PCA vs ICA

For the second scenario, we observe a more “banana-shaped” region of uncertainty described by the particles. As in scenario 1, the results have the same explanations and conclusions. However, for this case, we see that for the ICA, the reconstructed states in the velocity components are relatively better compared to the PCA reconstructions (see Fig. 6, Tab. 3 and Tab. 4) with the exception of the z component. Moreover, for this non-Gaussian epoch, it is clear that the non-Gaussian information retention given in the values of kurtosis for the components using the ICA (27.39, 5.5, 3.02 and 2.6) confirms that upon further compressing these components univariately, the loss of information will be much less compared to the PCA with kurtosis values of 4.30, 25.66, 3.46 and 2.91.

Table 3. Scenario 2: Mean Square Errors for PCA and ICA

MSE	PCA	ICA
$X(DU_{\oplus}^2)$	1.0722×10^{-8}	0.0001×10^{-8}
$Y(DU_{\oplus}^2)$	0.1365×10^{-4}	0.0001×10^{-8}
$Z(DU_{\oplus}^2)$	0.0051×10^{-12}	0.4822×10^{-8}
$V_x(DU_{\oplus}^2/TU_{\oplus}^2)$	13.941×10^{-8}	0.0463×10^{-8}
$V_y(DU_{\oplus}^2/TU_{\oplus}^2)$	0.0029×10^{-8}	0.0024×10^{-8}
$V_z(DU_{\oplus}^2/TU_{\oplus}^2)$	0.0052×10^{-12}	0.1545×10^{-8}

Table 4. Scenario 2: Kolmogorov-Smirnov test statistic for PCA and ICA

K-S statistic	PCA	ICA
X	0.0740	0.0080
Y	0.0100	0.0020
Z	0.0030	0.0150
V_x	0.0100	0.0060
V_y	0.1110	0.0280
V_z	0.0030	0.0560

6. Conclusion and Future Work

This work has shown that the ICA is capable of fully decorrelating non-Gaussian data and reducing its dimensionality by incorporating higher order moments beyond the second order. The PCA was compared to the ICA in the accuracy of reconstruction of the state vectors. ICA illustrated superiority in the positional states reconstructions to PCA, while PCA outperformed ICA in the velocity states’ reconstructions. Moreover, in considering the values obtained by the K-S test statistic, the ICA illustrated a better reconstruction in the positional states in scenario 1 and in all states but one in scenario 2. Since our goal was to maintain the non-Gaussian information by finding the components that maximized kurtosis, it is clear from the values of kurtosis of the components that the ICA is better equipped in performing this compared to PCA.

In the future, we expect to report on our ongoing work to use decorrelated and possibly dimensionally-reduced particle representations of orbital PDFs as the basis for accurately predicting such PDFs to times of interest, such as conjunctions between space objects. We are studying the potential of methods based on wavelet transforms, characteristic functions, and kernel densities to allow further

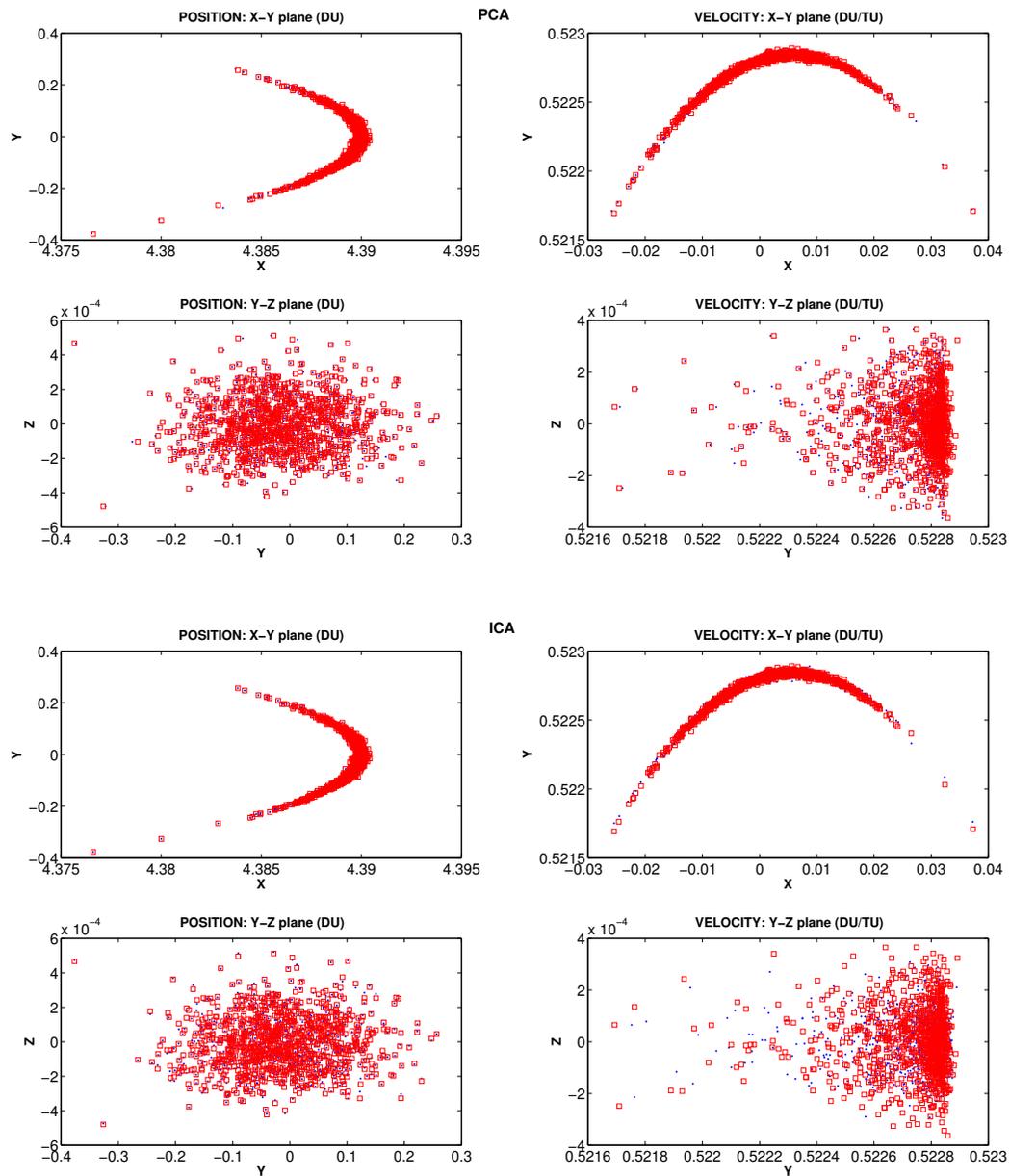


Figure 6. Scenario 2 PCA vs ICA: Original Data (Red Squares) and Reconstructed Data (Blue Dots)

compression, enabling compact PDF representations that could be stored in a catalog of space objects' PDFs. Once the compressed PDF has been transmitted along with other parameters required for reconstruction, the original data can be reconstructed with the goal of maximum information retention.

7. Acknowledgments

This research was funded by NASA's Graduate Student Research Program Fellowship based at Goddard Space Flight Center.