# BAYESIAN ESTIMATION OF A SINGLE MASS CONCENTRATION WITHIN AN ASTEROID

by

Aaron J. Woodard



A Thesis Submitted to the Faculty of the

SYSTEMS AND INDUSTRIAL ENGINEERING DEPARTMENT

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2017

STATEMENT BY AUTHOR

The thesis titled Bayesian Estimation of a Single Mass Concentration Within an Asteroid prepared by Aaron J. Woodard has been submitted in partial fulfillment of requirements for a master's degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Aaron J. Woodard

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

_____ 08/16/2017

*Roberto Furfaro* Date

*Assistant Professor of Systems and Industrial Engineering*

# 1 Contents

## Abstract

Orbit determination has long relied on the use of the Kalman filter, or specifically the extended Kalman filter, as a means of accurately navigating spacecraft. With the advent of cheaper, more powerful computers more accurate techniques such as the particle filter have been utilized. These Bayesian types of filters have in more recent years found their way to other applications. Dr. Furfaro and B. Gaudet have demonstrated the ability of the particle filter to accurately estimate the angular velocity, homogenous density, and rotation angle of a non-uniformly rotating ellipsoid shaped asteroid. This paper extends that work by utilizing a particle filter to accurately estimate the angular velocity and homogenous density of an ellipsoidal asteroid while simultaneously determining the location and mass of a mass concentration modeled as a point mass embedded within the asteroid. This work shows that by taking measurements in several locations around the asteroid, the asteroid's rotation state and mass distribution can be discerned.

# 2 Introduction

Bayesian estimation techniques have frequently been used in the field of astrodynamics for the purposes of navigation. The Kalman filter, an optimal linear Bayesian estimator, has been used in support of the Apollo program, the Agena program and the C-5A Aircraft navigation system[10] .

More recently there has been a growing interest in close proximity operations around asteroids. One such mission is the OSIRIS-REx Asteroid Sample Return Mission, which, as the title suggests, aims to return a sample of the asteroid Bennu to Earth for future study. The endeavor will require the spacecraft to precisely navigate in close proximity of the asteroid Bennu, approach the asteroid, make contact with the surface for a brief moment, collect a sample, and back away from the asteroid.

Hyabusa

Precisely navigating around the asteroid in close proximity and making contact is no simple task. The accuracy of such a maneuver depends heavily on how well the dynamical system is known. A dynamical model that neglects or approximates higher order terms can lead to a poor estimate of the state, which could lead to a critical failure and total loss of the spacecraft.

This thesis presents a method of accurately determining the dynamical system in close proximity of an asteroid by employing a particle filtering technique. The particle filter is given an initial estimate of the state based on a priori knowledge. By measuring the gravitational forces at several points around the asteroid, model parameters can be determined. The model parameters of interest in this paper are the rotational speed of the ellipsoidal asteroid, the density of the homogenous bulk of the ellipsoidal asteroid, and the location and mass of a point mass inhomogeneity embedded within the asteroid.

In Chapter 1, the application of Bayesian estimators in historical navigational systems is introduced. Chapter 1 Section 1 will discuss the motivation for this project and the previous work performed to lay the foundation for this project. The use of particle filtering was then proposed to accurately determine the asteroid's rotational state and mass distribution.

Chapter 2 discusses the deterministic models used for classical orbit determination. Form there, the shortcomings of deterministic models were explored and the need for a more accurate model which could account for random errors was presented. A brief overview of probability theory was presented, which led to the development of the stochastic processes and finally an overview of Bayesian estimation.

In Chapter 3, the Kalman filter is introduced as an optimal linear Bayesian estimator. A full derivation of the Kalman filter is presented, followed by an overview of the algorithm. Then a more convenient form of the Kalman filter is derived that makes numerical integration of the system easier. Finally the extended Kalman filter is derived, which is a sub-optimal filter that allows the Kalman filter to be applied to systems that are non-linear.

Chapter 4 will present the background for the development of the Sequential Monte Carlo filter, also known as the particle filter. The particle filter is then completely derived. An overview of the particle filter algorithm is then presented.

The methodology is then presented in Chapter 5. An overview of the goal of the project was presented, followed by a description of the tools utilized. The mathematical model of the system dynamics was then presented. The measurement approaches used as well as an overview of each particle filter were then outlined.

Chapter 6 presents an overview of the results from the experiments.

Finally Chapter 7 discusses the results from the experiments and offers ideas to improve the performance of the particle filter. Additionally, future work is discussed to demonstrate how the results from the particle filter could be utilized to accurately map the gravitational field in close proximity to an asteroid.

## 2.1 Motivation

As discussed in the section above, close proximity asteroid missions are becoming evermore popular. These missions are much riskier and present a greater danger to the spacecraft. The possibility of mission failure, such as failing to collect a sample, or complete loss of the spacecraft is much higher than that of a typical orbiter or lander mission due to the uncertainty of the asteroid's gravitational field. To reduce the possibility of critical spacecraft failure it is necessary to understand the mass distribution within the asteroid.

In the paper *Estimation of Asteroid Model Parameters*, Dr. Furfaro and B. Gaudet laid the foundation by showing it is possible to utilize particle filtering to determine the rotational state and homogenous density of an ellipsoidal asteroid.

Dr. Furfaro and B. Gaudet considered a non-uniformly rotating ellipsoidal asteroid that was given an angular velocity oriented at an angle $\theta = 75°$ from the Z-axis with magnitude of $\omega_0 = 4.06x10^{-4}$ rad/s. The ellipsoidal asteroid had semi-major axes of 250m, 287m and 350m and homogenous density of 1400kg/m$^3$.

Twelve measurements were taken over four nutation periods around the asteroid at two locations. Measurements were corrupted with bias and noise error. The initial conditions were given an amount of bias akin to approximating the state with some prior knowledge.

Over several iterations the particle filter was able to determine the parameters of the system to within 1-3% depending on the amount of bias given to the measurement errors.

This thesis will expand upon the previous work by demonstrating the use of particle filtering to estimate the location and mass of a point mass concentration. Additionally, the rotational state and density of homogenous bulk of the asteroid will simultaneously be determined using particle filtering. Once the mass concentration is estimated using a point mass model, it will be possible to swap out the simplistic gravitational model with a more accurate model such as the polyhedron gravity model developed by Dan Scheeres[9] . Then instead of estimating a point mass, the average density and components of the density gradient can be estimated. This will provide a very accurate mapping of the gravitational field.

# 3 Literature Review

The following sections provide a brief history on deterministic models and present the basics of probability theory, stochastic processes, and Bayesian estimation.

## 3.1 History

Classical mechanics typically has employed a deterministic approach for estimating the state of a system. The idea of using deterministic models to explain physical phenomenon is simple: by measuring a sufficient number of parameters about the system, the system dynamics model can be used to completely determine the state of the system. The deterministic approach has been utilized by many scientists, one such notable example Johannes Kepler. By utilizing the observations made by Tycho Brahe, Kepler was able to formulate his three laws of planetary motion, which were later built upon by Newton. Orbit determination to this day relies heavily on the principles and foundations laid by Kepler.

That being said, as technology has progressed and as humans have become capable of exploring the solar system, there is an ever growing need for more accurate models of physical phenomenon that classical deterministic mechanics cannot provide. A major flaw of deterministic mechanics is that it is not able to model errors within the system. These errors can come from many sources, as Maybeck describes:

- Approximations in the system dynamics model
- Disturbances in the system dynamics model that cannot be modeled deterministically
- Inaccurate measurements of the system

To illustrate the first point, consider the two-body problem. The two-body problem is a system dynamics model derived from Newton's law of universal gravitation and describes the motion of two point mass under the influence of each other's gravitational forces. This model has been used

the world over to describe the orbits of many celestial bodies and is still commonly used for preliminary orbit design. However, the two-body problem is only a simplification of the more general n-body problem. The two-body problem cannot model the effects of the Sun on the Moon's orbit around the Earth, nor can it model the gravitational effects of the Earth's non-spherical shape on Earth satellites. Over time, the approximations of the system will cause error to accumulate, leading to larger and larger differences between the predicted state and the true state of the system.

Furthermore, the state of the system is not only determined by the system dynamics and the control inputs but also from the disturbances within the system that cannot be controlled or modeled deterministically[12] . For example, consider the control surface of an aircraft. The orientation of the aircraft depends on the elevation of the control surface. However, the elevation of the control surface cannot be precisely controlled. There will always be some amount of uncertainty in the elevation of the control surface. So if it is desired to elevate the control surface by 10°, the actual elevation may be $10 \pm 1°$ or worse if the control surface or actuator mechanism were poorly designed. Thus, the uncertainty in the control surface will lead to a disturbances in the system dynamics and result in a difference between the expected value of the aircraft's orientation versus its actual value, none of which can be deterministically modeled.

Finally, no sensor can ever provide a perfect measurement of a system[12] . Sensors are susceptible to disturbances from other undesired signals and introduce inaccuracies from their own system dynamics into the observation. Thus, the state can never be precisely measured.

As the need for more accurate models grows, a new approach is needed to accurately model the errors that arise from the probabilistic disturbances that occur in physical phenomenon. This has lead to the development of stochastic system models, thus the motivation for the Bayesian estimator.

## 3.2 Probability Background

It is assumed that the reader has some understanding of a probability density function and random variables. This section will briefly cover probability topics and define important equations that will be useful for the remainder of this paper.

Consider the random variable $X_i$ for i = 1,2,...,n. The expectation or mean of the random variable $X_i$ is defined by equation Eq. 1. $f(x)$ is the probability density function of the random variable $X_i$. The expectation can be described as the value obtained by averaging all the outcomes from an experiment[12] .

$$E[X_i] = \int_{-\infty}^{\infty} x f(x) dx \qquad \textbf{Eq. 1}$$

Thus, the expectation or mean vector of a random vector $\mathbf{x} = [X_1, X_2, \ldots, X_n]^T$ is given by Eq. 2.

$$E[\mathbf{x}] = \left[E[X_1], E[X_2], \ldots E[X_n]\right] \qquad \textbf{Eq. 2}$$

The covariance matrix can then be defined by Eq. 3. An element whose position is given by i and j in the covariance matrix describes the covariance of the ith and jth element of the random vector **x**. The covariance is a measure of how much a change in one element of **x** is associated with a change in a different element of **x**. Note the diagonal elements of the covariance matrix, where i = j, are simply variance of the ith parameter of the random vector **x**, which describes the density of the values about the mean.

$$\mathbf{P_{xx}} = E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^T] \qquad \textbf{Eq. 3}$$

If **x** is time dependent, that is **x** = **x**(t), the covariance matrix becomes:

$$\mathbf{P_{xx}}(t) = E[(\mathbf{x}(t) - E[\mathbf{x}(t)])(\mathbf{x}(t) - E[\mathbf{x}(t)])^T] \qquad \textbf{Eq. 4}$$

Then the covariance of **x**(t) can be calculated for **x**($t_1$) and **x**($t_2$) to measure the variability of **x**(t) with previous values of **x**(t).

$$\mathbf{P_{xx}}(t_1, t_2) = \mathrm{E}[(\mathbf{x}(t_1) - \mathrm{E}[\mathbf{x}(t_1)])(\mathbf{x}(\mathbf{t_2}) - \mathrm{E}[\mathbf{x}(\mathbf{t_2})])^{\mathrm{T}}] \quad \textbf{Eq. 5}$$

This leads to the concept of correlation. Correlation is defined as[12] :

$$\Psi_{xx}(t_1, t_2) = \mathrm{E}[\mathbf{x}(t_1)x(t_2)^{\mathrm{T}}] \quad \textbf{Eq. 6}$$

$$= \mathbf{P_{xx}}(t_1, t_2) + \mathrm{E}[\mathbf{x}(t_1)]\mathrm{E}[\mathbf{x}(t_2)]^{\mathrm{T}}$$

Correlation is the measure of the dependence of the current value of **x**(t) with previous values of **x**(t). If **x**(t) is uncorrelated in time then the correlation matrix is defined by the following equation. Term "white" is used to specify a process that is uncorrelated in time for all $t_1 \neq t_2$.

$$\Psi_{xx}(t_1, t_2) = \mathrm{E}[\mathbf{x}(t_1)]\mathrm{E}[\mathbf{x}(t_2)]^{\mathrm{T}}$$

$$\mathbf{P_{xx}}(t_1, t_2) = \mathbf{0}$$

Finally, there is a need to describe the idea of a Brownian motion process vector. A Brownian motion process vector $\boldsymbol{\beta}(\cdot,\cdot)$ is a vector that meets the following criteria[12] :

1. A Brownian motion process is a process with independent increments:
    a. $\boldsymbol{\beta}(t_1,\cdot) - \boldsymbol{\beta}(t_0,\cdot)$, $\boldsymbol{\beta}(t_2,\cdot) - \boldsymbol{\beta}(t_1,\cdot)$, … $\boldsymbol{\beta}(t_n,\cdot) - \boldsymbol{\beta}(t_{n-1},\cdot)$, are independent for

[1] $t_0 < t_1 < t_2 < \cdots < t_n$

2. Increments of a Brownian motion process are Gaussian random variables such that:
    a. $E\left[\left(\boldsymbol{\beta}(t_2,\cdot) - \boldsymbol{\beta}(t_1,\cdot)\right)\right] = 0$
    b. $E\left[\left(\boldsymbol{\beta}(t_2,\cdot) - \boldsymbol{\beta}(t_1,\cdot)\right)\left(\boldsymbol{\beta}(t_2,\cdot) - \boldsymbol{\beta}(t_1,\cdot)\right)^T\right] = \int_{t_1}^{t_2} \mathbf{Q}(t)dt$
3. $\boldsymbol{\beta}(t_0, \omega_t) = 0$ for all $\omega_i \epsilon \Omega$

Brownian motion is central to the derivation of stochastic integrals as derived by Maybeck. The concept of a stochastic integral will be necessary for the solution of a linear stochastic differential equation $\dot{\mathbf{x}}(t)$, as will be seen in the following sections.

Finally the following equation is Bayes' Theorem (or Bayes' Rule):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Here P(A) and P(B) are the probabilities of events A and B occurring without regard to each other. P(A|B) is the probability of event A occurring given that event B has occurred. This is a conditional probability. Likewise, P(B|A) is the probability of event B occurring given that event A has occurred. Bayes' Theorem will be a crucial tool in the following sections as the Kalman and particle filters are derived.

## 3.3 Stochastic Processes

Stochastic processes are families of random variables $\mathbf{x}(t_1)$, $\mathbf{x}(t_2)$, … $\mathbf{x}(t_N)$ defined on a sample space $\Omega$. These processes can be discrete or continuous in time. The set of random variables $\mathbf{x}(t)$ can then be characterized by their joint probability distribution.

Ultimately, the goal in using a stochastic model is to provide an accurate representation for the state of some system of interest, taking into account and minimizing the errors in the system model and observations. The developments made in the deterministic model can be extended to the stochastic system model by rewriting the deterministic model in terms of random variables and appending a random disturbance term to the model.

Consider the linear Gaussian stochastic system model[12] :

$$\dot{\mathbf{x}}(\mathrm{t}) = \mathbf{F}(\mathrm{t})\mathbf{x}(\mathrm{t}) + \mathbf{G}(\mathrm{t})\mathbf{w}(\mathrm{t})$$

Here, $\dot{\mathbf{x}}(\mathrm{t})$ is the stochastic representation of the system dynamics model and. $\mathbf{w}(\mathrm{t})$ is a Gaussian random vector of measurement noises processes[12] . In this form, the equations of motion are too general to be of any practical use. However, the stochastic system can still be evaluated by considering the joint probability distribution[12] .

$$F_{\mathbf{x}(\mathrm{t_i})|\mathrm{x}(\mathrm{t_{i-1}}),\ldots,\mathrm{x}(\mathrm{t_0})}(\xi_\mathrm{i}|\mathbf{x}_{\mathrm{i}-1}, \ldots, \mathbf{x}_0)$$

Here $\boldsymbol{\xi}_\mathbf{i}$ is a realization of $\mathbf{x}(\mathrm{t_i})$. Typically this distribution as it is written is too cumbersome to work with. Therefore the processes considered here will be strictly Markov processes. A Markov process is a process where the previous value of the state encompasses all information needed to estimate the current state[12] . Thus

$$F_{\mathbf{x}(\mathrm{t_i})|\mathrm{x}(\mathrm{t_{i-1}}),\ldots,\mathrm{x}(\mathrm{t_0})}(\xi_\mathrm{i}|\mathbf{x}_{\mathrm{i}-1}, \ldots, \mathbf{x}_0) = F_{\mathbf{x}(\mathrm{t_i})|\mathrm{x}(\mathrm{t_{i-1}})}(\xi_\mathrm{i}|\mathbf{x}_{\mathrm{i}-1})$$

Given the linear definition of the system equations and assumptions above, the distribution $F_{\mathbf{x}(\mathrm{t_i})|\mathrm{x}(\mathrm{t_{i-1}})}(\boldsymbol{\xi}_\mathbf{i}|\mathbf{x}_{\mathrm{i}-1})$ is therefore a Gaussian-Markov distribution. As such the distribution can be completely defined by its mean and covariance. Thus, an estimate can be generated from the mean of the distribution.

With a complete description of the assumed Gaussian distribution $F_{\mathbf{x}(t_i)|x(t_{i-1})}(\xi_i|\mathbf{x}_{i-1})$, it is now possible to create an estimate of the state $\mathbf{x}(t_i)$ at a given time $t_i$ by determining the mean of the distribution. However, suppose more information about the system, such as measurements, were available. How could this be used to better improve the estimate of the state $\mathbf{x}(t_i)$? Consider the linear observation process

$$\mathbf{z}(t_i) = \mathbf{H}(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i)$$

Where measurements $\mathbf{z}(t_i)$ are available at discrete times $t_i, t_{i-1}, \dots, t_0$ and $\mathbf{v}(t_i)$ is a Gaussian random noise process. Given these measurements, the distribution $F_{\mathbf{x}(t_i)|x(t_{i-1})}(\xi_i|\mathbf{x}_{i-1})$ could be improved by conditioning it on all previous measurements $\mathbf{Z}(t_i) = [\mathbf{z}(t_i), \mathbf{z}(t_{i-1}), \dots \mathbf{z}(t_0)]$. Thus the distribution can be defined given all of the information possible as

$$F_{\mathbf{x}(t_i)|x(t_{i-1}),\mathbf{z}(t_i)}(\xi_i|\mathbf{x}_{i-1}, \mathbf{Z_i})$$

From this conditional distribution, which is still assumed Gaussian, an improved estimate can now be calculated from mean of the distribution.

Note that an analytical definition of the joint distribution $F_{\mathbf{x}(t_i)|x(t_{i-1}),\mathbf{z}(t_i)}(\xi_i|\mathbf{x}_{i-1})$ is only possible if $\dot{\mathbf{x}}(t)$ and $\mathbf{z}(t_i)$ are linear and Gaussian, as will be seen in the derivation of the Kalman filter. For nonlinear, non-Gaussian processes, other tools must be developed, as will be seen in the derivation of the Sequential Monte Carlo filter.

## 3.4 Bayesian Statistics

From the previous section, it is obvious that the objective here is to generate an estimate for the current state from the distribution $F_{\mathbf{x}(t_i)|x(t_{i-1}),\mathbf{z}(t_i)}(\xi_i|\mathbf{x}_{i-1})$**,** known as the posterior distribution. Very rarely will one have access to this distribution directly, and so this distribution must be computed from a known prior distribution $F_{x(t_{i-1})|\mathbf{z}(t_{i-1})}(\boldsymbol{\xi}_{\mathbf{i-1}}|\mathbf{Z}_{i-1})$. Thus, the goal here is to

propagate the information in the prior distribution and produce a posterior distribution from which an estimate can be generated. This leads into the topic of Bayesian statistics.

Bayesian statistics is a branch of statistics that is interested in taking a prior distribution, propagating it forward using all relevant data about the system, and creating a posterior distribution[6] . Thus the goal in Bayesian estimation is to develop a definition of the posterior distribution from the prior distribution and, using all relevant information about that system, create an estimate from the posterior distribution. Two such Bayesian techniques described here are the Kalman filter and the Sequential Monte Carlo filter.

An estimator is a function that creates an estimate of a random variable given the information about it. The estimate is then a logical choice of the mean, median or mode of the distribution[12] . In the case of a linear Gaussian system, the mean, median, and mode are all equivalent.

# 4 The Kalman Filter

This section will provide a brief overview and history of the Kalman filter. A detailed derivation of the Kalman filter is then presented, followed by a short description of the extended Kalman filter.

## 4.1 Overview

The Kalman filter was developed by Rudolf Emil Kálmán in 1960. The Kalman filter is an optimal recursive estimation algorithm that produces highly accurate estimates of an unknown state. The The Kalman filter is an optimal linear estimator. Given that the system dynamics are linear and all errors are Gaussian, the Kalman filter provides an exact analytical solution to the estimation problem.

The Kalman filter works by using a prediction-correction scheme that relies on both the mathematical model and several measurements of the system. Once the Kalman filter is initialized with an initial condition, the state is then propagated to a future time using the information provided by the mathematical model. This predicted estimate has accumulated some amount of error due to the fact that the mathematical model of the system is not complete.

Once the predicted estimate of the state is calculated, a measurement is taken at a discrete time $t_i$. This measurement will be used to help correct for errors that accumulated due to generating the state estimate from the mathematical model. However, it is important to note that this measurement itself is not perfect due to the reasons explained above.

After the state has been propagated using the mathematical model and a measurement has been taken at time $t_i$, the Kalman filter uses the uncertainty from both the measurement and the propagated state to generate a correction term that is then added to the propagated state. When uncertainty in the measurement is less than the uncertainty in the predicted value, the value from

the measurement has more influence on the corrected estimated, and likewise, if the uncertainty in the predicted value is less that of the measurement, the predicted value has more influence on the corrected estimate.

The Kalman filter, or rather the extended Kalman filter as will be explained later, has had numerous applications in the field of orbit determination and spacecraft navigation. Since its creation the Kalman filter has become an integral part in almost every guidance and navigation system (Leondes). One example of the use of the Kalman filter is the Apollo spacecraft (Leondes).

The Kalman filter is explicitly derived in the following sections.

## 4.2 Derivation of the Kalman Filter

Consider a state process vector **x**(t) that satisfies Eq. 7, where **x**(t) is an n-by-1 state process vector. For this derivation, an unforced linear system is used.

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{w}(t) \qquad \textbf{Eq. 7}$$

**F**(t) is an n-by-n dynamics matrix. **G**(t) is an n-by-s noise input matrix. **w**(t) is an n-by-1 Gaussian white noise vector. **w**(t) represents the time derivative of the Brownian motion process vector $\frac{d\boldsymbol{\beta}}{dt}$ (although a Brownian motion process vector is not differentiable). **w**(t) has an expectation and variance of

$$E[\mathbf{w}(t)] = 0$$

$$E[\mathbf{w}(t)\mathbf{w}(t')^T] = \mathbf{Q}(t)\delta(t - t')$$

$\mathbf{Q}(t)$ is an s-by-s matrix of piecewise continuous functions such that $\mathbf{Q}(t)$ is symmetric and positive semidefinite. $\delta(t - t')$ is the Dirac delta function, where $\delta(0) = 1$ and $\delta(y) = 0$ for all $y \neq 0$.

Measurements of the system can be described by using a measurement process $\mathbf{z}(t_i)$, where $\mathbf{z}(t_i)$ provides measurements of the state process $\mathbf{x}(t)$ at discrete times $t_i$. The measurement process $\mathbf{z}(t_i)$ can then be related to $\mathbf{x}(t_i)$ by the equation:

$$\mathbf{z}(t_i) = \mathbf{H}(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i) \qquad \textbf{Eq. 8}$$

$\mathbf{H}(t_i)$ is an m-by-n measurement that maps the state vector process to the measurement process. $\mathbf{v}(t_i)$ is an m-by-1 measurement noise vector. $\mathbf{v}(t_i)$ is modeled as zero-mean, white Gaussian noise vector. The expectation and covariance of $\mathbf{v}(t_i)$ are given as (respectively):

$$E[\mathbf{v}(\mathbf{t})] = 0$$

$$E\left[\mathbf{v}(t_i)\mathbf{v}(t_j)^T\right] = \begin{cases} \mathbf{R}(t_i) & t_i = t_j \\ 0 & t_i \neq t_j \end{cases}$$

Where $\mathbf{R}(t_i)$ is a positive definite, symmetric matrix.

For convenience, a process vector that represents the entire measurement history will be denoted:

$$\mathbf{Z}(t_i) = \begin{bmatrix} \mathbf{z}(t_1) \\ \mathbf{z}(t_2) \\ \vdots \\ \mathbf{z}(t_i) \end{bmatrix}$$

$\mathbf{Z_i}$ will be used to denote a realization of the process vector $\mathbf{Z}(\mathbf{t_i})$**,** and $\boldsymbol{\xi}$ will be used to denote a realization of $\mathbf{x}(t_i)$

The Kalman filter creates optimal estimates of the state process $\mathbf{x}(t)$ by first propagating the state from a time denoted $t_{i-1}$ to a future time $t_i$ (prediction), using the information of measurements $\mathbf{z}(t_1), \mathbf{z}(t_2), \ldots, \mathbf{z}(t_{i-1})$ and then taking a measurement $\mathbf{z}(t_i)$ at the time $t_i$ and using that measurement to update the prediction and provide a better estimate of the state. Times before

correction will be denoted with a "-" superscript, $t^-$, while times after the estimate has been corrected will be denoted with a "+" superscript, $t^+$.

For the derivation of the Kalman filter, it will be assumed that the conditional probability density function of $\mathbf{x}(t_{i-1})$, denoted $f_{\mathbf{x}(t_{i-1})|\mathbf{Z}(\mathbf{t_{i-1}})=\mathbf{Z_{i-1}}}(\boldsymbol{\xi}|\mathbf{Z}_{i-1} = \mathbf{Z_{i-1}})$, is defined by a Gaussian density. This is required in order to show that the probability density function of $\mathbf{x}(t_i)$ is also Gaussian, which makes defining the probability density of $\mathbf{x}(t_i)$ possible.

To begin, it is assumed that the conditional mean $\hat{\mathbf{x}}(t_{\mathbf{i-1}}^+)$ and its conditional covariance matrix $\mathbf{P}(t_{\mathbf{i-1}}^+)$ are known. Here $\hat{\mathbf{x}}(t_{\mathbf{i-1}}^+)$ and $\mathbf{P}(t_{\mathbf{i-1}}^+)$ have been corrected with the measurements $\mathbf{Z}(t_{i-1})$. The estimated state and covariance are defined by:

$$\hat{\mathbf{x}}(t_{\mathbf{i-1}}^+) = \mathrm{E}[\mathbf{x}(t_{i-1})|\, \mathbf{Z}(t_{i-1})] \qquad \textbf{Eq. 9}$$

$$\mathbf{P}(t_{\mathbf{i-1}}^+) = \mathrm{E}\left[\left(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\right)\left(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\right)^{\mathrm{T}} |\, \mathbf{Z}(t_{i-1})\right] \qquad \textbf{Eq. 10}$$

The general solution to Eq. 7 is

$$\mathbf{x}(t_i) = \boldsymbol{\Phi}(t_i, t_{i-1})\mathbf{x}(t_{i-1}) + \mathbf{w}_{\mathrm{d}}(t_{i-1}) \qquad \textbf{Eq. 11}$$

Here $\boldsymbol{\Phi}(t_i, t_{i-1})$ is the state transition matrix. The term $\mathbf{w_d}(t_{i-1})$ is a stochastic process defined by the stochastic integral:

$$\mathbf{w_d}(t_{i-1}) = \int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i, \tau)\mathbf{G}(\tau)\mathbf{d}\boldsymbol{\beta}(\tau) \qquad \textbf{Eq. 12}$$

First it must be shown that the probability density function $\mathbf{x}(t_i)$ conditioned on $\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}$, denoted $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}}(\boldsymbol{\xi}|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1})$, is Gaussian. Since $\mathbf{x}(t_i)$ is a linear combination of $\mathbf{w_d}(t_{i-1})$ and $\mathbf{x}(t_{i-1})$, the conditional probability density of $\mathbf{x}(t_i)$ can be shown to be Gaussian if

it can be demonstrated that the joint conditional probability density $f_{\mathbf{x}(t_i),\mathbf{w_d}(t_i)|\mathbf{Z}(\mathbf{t_{i-1}})=\mathbf{Z}_{i-1}}(\boldsymbol{\xi},\boldsymbol{\rho}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1})$ is Gaussian. This can be done by applying Bayes' rule to the joint conditional probability density of $\mathbf{w_d}(t_{i-1})$ and $\mathbf{x}(t_{i-1})$.

$$f_{\mathbf{x}(t_i),\mathbf{w_d}(t_i)|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}}(\boldsymbol{\xi},\boldsymbol{\rho}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1})=\frac{f_{\mathbf{x}(t_i),\mathbf{w_d}(t_i),\mathbf{Z}(\mathbf{t_{i-1}})}(\boldsymbol{\xi},\boldsymbol{\rho},\mathbf{Z}_{i-}}{f_{\mathbf{Z}(\mathbf{t_{i-1}})}(\mathbf{Z}_{i-1})} \quad \textbf{Eq. 13}$$

Since $\mathbf{w_d}(t_{i-1})$ is independent of $\mathbf{x}(t_{i-1})$ and $\mathbf{Z}(t_{i-1})$, their joint density can be separated like so:

$$f_{\mathbf{x}(t_i),\mathbf{w}(t_i)|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}}(\boldsymbol{\xi},\boldsymbol{\rho}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1})=\frac{f_{\mathbf{x}(t_i),\mathbf{Z}(\mathbf{t_{i-1}})}(\boldsymbol{\xi},\mathbf{Z}_{i-1})f_{\mathbf{w_d}(t_i)}(}{f_{\mathbf{Z}(\mathbf{t_{i-1}})}(\mathbf{Z}_{i-1})} \quad \textbf{Eq. 14}$$

Reorganizing the terms on the right-hand side

$$f_{\mathbf{x}(t_i),\mathbf{w}(t_i)|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}}(\boldsymbol{\xi},\boldsymbol{\rho}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}) \quad \textbf{Eq. 15}$$
$$=f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}}(\boldsymbol{\xi}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1})f_{\mathbf{w_d}(t_i)}(\boldsymbol{\rho})$$

Because $f_{\mathbf{x}(t_i)|\mathbf{Z}(\mathbf{t_{i-1}})=\mathbf{Z}_{i-1}}(\boldsymbol{\xi}|\mathbf{Z}(\mathbf{t_{i-1}})=\mathbf{Z}_{i-1})$ was assumed Gaussian and $f_{\mathbf{w_d}(t_i)}(\boldsymbol{\rho})$ is Gaussian due to the definition of a Brownian motion process, their product must also be Gaussian. Thus $f_{\mathbf{x}(t_i),\mathbf{w}(t_i)|\mathbf{Z}(\mathbf{t_{i-1}})=\mathbf{Z}_{i-1}}(\boldsymbol{\xi},\boldsymbol{\rho}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1})$ is Gaussian as well.

Given that $f_{\mathbf{x}(t_i)|\mathbf{Z}(\mathbf{t_{i-1}})}\left(\boldsymbol{\xi}|\mathbf{Z}(t_{i-1})\right)$ is Gaussian, the density can be completely described by mean and covariance. The expectation of $\mathbf{x}(t_i)$ conditioned on $\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}$ is given as:

$$\mathrm{E}[\mathbf{x}(t_i)|\mathbf{Z_{i-1}}]=\mathrm{E}[\Phi(t_i,t_{i-1})\mathbf{x}(t_{i-1})+\ \mathbf{w_d}(t_{i-1})|\ \mathbf{Z}(t_{i-1})]$$

This can be split apart like so:

$$\mathrm{E}[\mathbf{x}(t_i)|\mathbf{Z_{i-1}}]=\ \mathrm{E}[\Phi(t_i,t_{i-1})\mathbf{x}(t_{i-1})|\ \mathbf{Z}(t_{i-1})]+\mathrm{E}[\mathbf{w_d}(t_{i-1})|\ \mathbf{Z}(t_{i-1})]$$

$\mathbf{w_d}(t_{i-1})$ is zero-mean white Gaussian noise vector, as shown in the derivation of stochastic integrals[12] . This reduces the equation to:

$$E[\mathbf{x}(t_i)|\mathbf{Z_{i-1}}] = E[\Phi(t_i, t_{i-1})\mathbf{x}(t_{i-1})|\, \mathbf{Z}(t_{i-1})]$$

Since $\Phi(t_i, t_{i-1})$ is not a random process, it can be taken out of the expectation operator, yielding:

$$E(\mathbf{x}(t_i)|\mathbf{Z_{i-1}}) = \Phi(t_i, t_{i-1})E\big(\mathbf{x}(t_{i-1})|\, \mathbf{Z}(t_{i-1})\big) \quad \textbf{Eq. 16}$$

The conditional expectation of $\mathbf{x}(t_{i-1})$ is then defined as:

$$E\big(\mathbf{x}(t_{i-1})|\, \mathbf{Z}(t_{i-1})\big) = \hat{\mathbf{x}}(t_{i-1}^{+}) \quad \textbf{Eq. 17}$$

Then plugging Eq. 17 into Eq. 16 gives:

$$\hat{\mathbf{x}}(t_i^{-}) = \Phi(t_i, t_{i-1})\hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+}) \quad \textbf{Eq. 18}$$

The covariance can be derived in a similar way starting with:

$$P(t_i^{-}) = E\Big[\big(\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i)\big)\big(\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i)\big)^{\mathbf{T}}|\, \mathbf{Z}(t_{i-1})\Big] \quad \textbf{Eq. 19}$$

By inserting Eq. 11 and Eq. 18 into Eq. 19 yields:

$$\begin{aligned} P(t_i^{-}) = E\Big[&\big(\mathbf{\Phi}(t_i, t_{i-1})\mathbf{x}(t_{i-1}) + \mathbf{w_d}(t_{i-1}) \\ &- \Phi(t_i, t_{i-1})\hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+})\big)\big(\mathbf{\Phi}(t_i, t_{i-1})\mathbf{x}(t_{i-1}) + \mathbf{w_d}(t_{i-1}) \\ &- \Phi(t_i, t_{i-1})\hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+})\big)^{\mathbf{T}}|\, \mathbf{Z}(t_{i-1})\Big] \end{aligned}$$

Rearranging terms leads to:

$$\begin{aligned} P(t_i^{-}) = E\Big[&\Big(\mathbf{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+})\big) \\ &+ \mathbf{w_d}(t_{i-1})\Big)\Big(\mathbf{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+})\big) + \mathbf{w_d}(t_{i-1})\Big)^{\mathbf{T}}|\, \mathbf{Z}(t_{i-1})\Big] \end{aligned}$$

Now multiplying through gives:

$$P(t_i^-) = E\Big[\Big(\boldsymbol{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)^T \boldsymbol{\Phi}(t_i, t_{i-1})^T + \boldsymbol{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)\mathbf{w_d}(t_{i-1})^T + \mathbf{w_d}(t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)^T \boldsymbol{\Phi}(t_i, t_{i-1})^T \Big) + \mathbf{w_d}(t_{i-1})\mathbf{w_d}(t_{i-1})^T \,|\, \mathbf{Z}(t_{i-1})\Big]$$

The expectation operator can then be split up like so:

$$P(t_i^-) = E\Big[\boldsymbol{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)^T \boldsymbol{\Phi}(t_i, t_{i-1})^T \,|\, \mathbf{Z}(t_{i-1})\Big] \qquad \textbf{Eq. 20}$$

$$+ E\big[\boldsymbol{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)\mathbf{w_d}(t_{i-1})^T \,|\, \mathbf{Z}(t_{i-1})\big]$$

$$+ E\Big[\mathbf{w_d}(t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)^T \boldsymbol{\Phi}(t_i, t_{i-1})^T \,|\, \mathbf{Z}(t_{i-1})\Big]$$

$$+ E[\mathbf{w_d}(t_{i-1})\mathbf{w_d}(t_{i-1})^T \,|\, \mathbf{Z}(t_{i-1})]$$

Eq. 20 will now be evaluated term by term. The first term can reorganized like so, noting that $\boldsymbol{\Phi}(t_i, t_{i-1})$ is not a random process.

$$E\Big[\boldsymbol{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)^T \boldsymbol{\Phi}(t_i, t_{i-1})^T \,|\, \mathbf{Z}(t_{i-1})\Big]$$

$$= \boldsymbol{\Phi}(t_i, t_{i-1})\, E\Big[\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)^T \,|\, \mathbf{Z}(t_{i-1})\Big] \boldsymbol{\Phi}(t_i, t_{i-1})^T$$

By substituting Eq. 10 the first term of the $P(t_i^-)$ becomes:

$$\boldsymbol{\Phi}(t_i, t_{i-1})\, E\Big[\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)^T \,|\, \mathbf{Z}(t_{i-1})\Big] \boldsymbol{\Phi}(t_i, t_{i-1})^T$$

$$= \boldsymbol{\Phi}(t_i, t_{i-1})\, \mathbf{P}(t_{i-1}^+) \boldsymbol{\Phi}(t_i, t_{i-1})^T$$

The second and third terms can easily be evaluated by noticing that the state estimator, $\boldsymbol{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)\big)$, and process noise, $\mathbf{w_d}(t_{i-1})$, are uncorrelated with each other.

$$\mathrm{E}\Big[\,\boldsymbol{\Phi}(t_i, t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^{+})\big)\mathbf{w_d}(t_{i-1})^{\mathrm{T}} |\, \mathbf{Z}(t_{i-1})\Big]$$

$$= \mathrm{E}\Big[\mathbf{w_d}(t_{i-1})\big(\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^{+})\big)^{\mathrm{T}}\boldsymbol{\Phi}(t_i, t_{i-1})^{\mathrm{T}} |\, \mathbf{Z}(t_{i-1})\Big] = 0$$

Lastly the fourth term can be evaluated using the properties of the stochastic integral described by Maybeck:

$$\mathrm{E}[\mathbf{w_d}(t_{i-1})\mathbf{w_d}(t_{i-1})^{\mathrm{T}} |\, \mathbf{Z}(t_{i-1})] = \int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i, \tau)\mathbf{G}(\tau)\mathbf{Q}(t)\mathbf{G}(\tau)^{\mathrm{T}}\boldsymbol{\Phi}(t_i, \tau)^{\mathrm{T}}\mathrm{d}\tau$$

$\mathbf{Q}(t)$ is a positive semi-definite diffusion matrix that comes from the definition of a Brownian motion process.

Finally the covariance matrix of $\mathbf{x}(t_i)$ becomes:

$$\mathbf{P}(t_i^{-}) = \boldsymbol{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{i-1}^{+})\boldsymbol{\Phi}(t_i, t_{i-1})^{\mathrm{T}} + \int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i, \tau)\mathbf{G}(\tau)\mathbf{Q}(t)\mathbf{G}(\tau)^{\mathrm{T}}\boldsymbol{\Phi}(t_i, \tau)^{\mathrm{T}}\mathrm{d}\tau \qquad \textbf{Eq. 21}$$

The probability density of $\mathbf{x}(t_i)$ is completely defined.

Now that the prediction step of the Kalman filter has been derived, the correction step can be derived. To correct the predicted value $\mathbf{x}(t_i)$ the conditional probability density $f_{\mathbf{x}(t_i)|\mathrm{Z}(t_i)}$ must shown to be Gaussian and fully defined. This is accomplished using Bayes' rule multiple times.

$$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)} = \frac{f_{\mathbf{x}(t_i),\mathbf{Z}(t_i)}}{f_{\mathbf{Z}(t_i)}} \qquad \textbf{Eq. 22}$$

By recalling the definition of $\mathbf{Z}(t_i)$, Eq. 22 can be expanded to give:

$$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)} = \frac{f_{\mathbf{x}(t_i),\mathbf{z}(t_i),\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i)\mathbf{Z}(t_{i-1})}}$$

Then the numerator can be expanded twice using Bayes' rule to give

$$f_{\mathbf{x}(\mathbf{t_i}),\mathbf{z}(\mathrm{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})} = f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})} f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})} f_{\mathbf{Z}(\mathrm{t_{i-1}})}$$

Likewise, the denominator can also be expanded using Bayes' rule

$$f_{\mathbf{z}(\mathrm{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})} = f_{\mathbf{z}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})} f_{\mathbf{Z}(\mathrm{t_{i-1}})}$$

Putting this all together results in:

$$f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_i})} = \frac{f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})} f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}}{f_{\mathbf{z}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}} \qquad \textbf{Eq. 23}$$

Now the probability density $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_i})}$ can be analyzed by analyzing the three probability density functions $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$, $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})}$ and $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$.

The density $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ has already been shown to be Gaussian and has been fully defined in deriving the prediction step. Explicitly, $\mathrm{f}_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ can be written as:

$$f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}(\boldsymbol{\xi}, \mathbf{Z}_{\mathrm{i-1}}) \qquad \textbf{Eq. 24}$$

$$= \frac{\exp\left(-\frac{1}{2}[\boldsymbol{\xi} - \hat{\mathbf{x}}(\mathrm{t_i^-})]^{\mathrm{T}} \mathbf{P}(\mathrm{t_i^-})^{-1} [\boldsymbol{\xi} - \hat{\mathbf{x}}(\mathrm{t_i^-})]\right)}{\sqrt{(2\pi)^{\mathrm{n}} \left|\mathbf{P}(\mathrm{t_i^-})\right|}}$$

Moving on to the density $\mathrm{f}_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})}$, recall the measurement model is described by:

$$\mathbf{z}(\mathrm{t_i}) = \mathbf{H}(\mathrm{t_i})\mathbf{x}(\mathrm{t_i}) + \mathbf{v}(\mathrm{t_i}) \qquad \textbf{Eq. 8}$$

$\mathbf{z}(\mathrm{t_i})$ is a linear combination of $\mathbf{x}(\mathrm{t_i})$ and $\mathbf{v}(\mathrm{t_i})$. Therefore the to show the conditional density $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})}$ is Gaussian, it must be shown that the joint density $f_{\mathbf{v}(\mathrm{t_i}),\mathbf{x}(\mathbf{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ is Gaussian.

Using Bayes rule:

$$f_{\mathbf{v}(t_i),\mathbf{x}(\mathbf{t_i})|\mathbf{Z}(t_{i-1})} = \frac{f_{\mathbf{v}(t_i),\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})}}{f_{\mathbf{Z}(t_{i-1})}}$$

$\mathbf{v}(t_i)$ is independent of $\mathbf{x}(t_i)$ and $\mathbf{Z}(t_{i-1})$, therefore the joint probability can be represented as:

$$f_{\mathbf{v}(t_i),\mathbf{x}(\mathbf{t_i})|\mathbf{Z}(t_{i-1})} = \frac{f_{\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})} f_{\mathbf{v}(\mathbf{t_i})}}{f_{\mathbf{Z}(t_{i-1})}}$$

Using Bayes' rule again to form:

$$f_{\mathbf{v}(t_i),\mathbf{x}(\mathbf{t_i})|\mathbf{Z}(t_{i-1})} = f_{\mathbf{x}(\mathbf{t_i})|\mathbf{Z}(t_{i-1})} f_{\mathbf{v}(\mathbf{t_i})}$$

This shows that $f_{\mathbf{v}(t_i),\mathbf{x}(\mathbf{t_i})|\mathbf{Z}(t_{i-1})}$ is a product of two Gaussian density. Therefore it is Gaussian as well, which means $f_{\mathbf{z}(t_i)|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})}$ is also Gaussian.

To define the density $f_{\mathbf{z}(t_i)|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})}$ relies on knowing that $\mathbf{x}(t_i)$ must have assumed a realization, $\mathbf{x}(t_i) = \boldsymbol{\xi}$. This means that $\mathbf{x}(t_i)$ is no longer a random process here and is now a fixed value $\mathbf{x}(t_i) = \boldsymbol{\xi}$.

The mean is given as:

$$\mathrm{E}[\mathbf{z}(t_i)|\mathbf{x}(t_i) = \boldsymbol{\xi}, \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] = \mathrm{E}[\mathbf{H}\mathbf{x}(t_i) + \mathbf{v}(t_i)|\, \mathbf{x}(t_i) = \boldsymbol{\xi}, \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}]$$

Given $\mathbf{H}(t_i)$ is not a random, $\mathbf{x}(t_i)$ is fixed, and $\mathbf{v}(t_i)$ is a zero-mean white Gaussian noise, the mean reduces to simply:

$$\mathrm{E}[\mathbf{z}(t_i)|\mathbf{x}(t_i) = \boldsymbol{\xi}, \mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] = \mathbf{H}(t_i)\boldsymbol{\xi} \qquad \textbf{Eq. 25}$$

The covariance is:

$$\mathrm{E}[(\mathbf{z}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi})(\mathbf{z}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi})^{\mathrm{T}} | \, \mathbf{x}(\mathrm{t_i}) = \boldsymbol{\xi}, \mathbf{Z}(\mathrm{t_{i-1}}) = \mathbf{Z_{i-1}}]$$

$$= \mathrm{E}[(\mathbf{H}(\mathrm{t_i})\mathbf{x}(\mathrm{t_i}) + \mathbf{v}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi})(\mathbf{H}(\mathrm{t_i})\mathbf{x}(\mathrm{t_i}) + \mathbf{v}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi})^{\mathrm{T}} | \, \mathbf{x}(\mathrm{t_i})$$

$$= \boldsymbol{\xi}, \mathbf{Z}(\mathrm{t_{i-1}}) = \mathbf{Z_{i-1}}]$$

But $\mathbf{x}(\mathrm{t_i}) = \boldsymbol{\xi}$, and the measurement noise is uncorrelated with $\mathbf{x}(\mathrm{t_i})$, thus the covariance simplifies to:

$$\mathrm{E}[(\mathbf{z}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi})(\mathbf{z}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi})^{\mathrm{T}} | \, \mathbf{x}(\mathrm{t_i}) = \boldsymbol{\xi}, \mathbf{Z}(\mathrm{t_{i-1}}) = \mathbf{Z_{i-1}}] = \mathrm{E}[\mathbf{v}(\mathrm{t_i})\mathbf{v}(\mathrm{t_i})^{\mathrm{T}}]$$

Recall that the covariance of $\mathbf{v}(\mathrm{t_i})$:

$$\mathrm{E}\left[\mathbf{v}(\mathrm{t_i})\mathbf{v}(\mathrm{t_j})^{\mathrm{T}}\right] = \begin{cases} \mathrm{R}(\mathrm{t_i}) & \mathrm{t_i} = \mathrm{t_j} \\ 0 & \mathrm{t_i} \neq \mathrm{t_j} \end{cases}$$

This makes the conditional covariance of the measurement $\mathbf{z}(\mathrm{t_i})$:

$$\mathrm{E}[(\mathbf{z}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi})(\mathbf{z}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi})^{\mathrm{T}} | \, \mathbf{x}(\mathrm{t_i}) = \boldsymbol{\xi}, \mathbf{Z}(\mathrm{t_{i-1}}) = \mathbf{Z_{i-1}}] = \mathbf{R}(\mathrm{t_i})$$

Therefore, the conditional density $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})}$ is fully defined and is written explicitly as:

$$f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),=\boldsymbol{\xi},\mathbf{Z}(\mathrm{t_{i-1}})=\mathbf{Z_{i-1}}} \qquad \textbf{Eq. 26}$$

$$= \frac{\exp\left(-\frac{1}{2}[\boldsymbol{\zeta} - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi}]^{\mathrm{T}}\mathbf{R}(\mathrm{t_i})^{-1}[\boldsymbol{\zeta} - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi}]\right)}{\sqrt{(2\pi)^{\mathrm{m}}|\mathbf{R}(\mathrm{t_i})|}}$$

Lastly the density $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ must be shown to be Gaussian. Since $\mathbf{z}(\mathrm{t_i})$ is a linear combination of $\mathbf{x}(\mathrm{t_i})$ and $\mathbf{v}(\mathrm{t_i})$, $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ can be shown to be Gaussian if $f_{\mathbf{x}(\mathrm{t_i}),\mathbf{v}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ are shown to be Gaussian.

Applying Bayes' rule to $f_{\mathbf{x}(\mathrm{t_i}),\mathbf{v}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ gives:

$$f_{\mathbf{x}(\mathrm{t_i}),\mathbf{v}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})} = f_{\mathbf{v}(\mathrm{t_i})|\mathbf{x}(\mathrm{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})} f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$$

Since $\mathbf{v}(t_i)$ is Gaussian and independent of $\mathbf{x}(t_i)$ and $\mathbf{Z}(t_{i-1})$, their joint densitys can be separated.

$$f_{\mathbf{x}(t_i),\mathbf{v}(t_i)|\mathbf{Z}(t_{i-1})} = f_{\mathbf{v}(t_i)} f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}$$

$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}$ has been shown to be Gaussian and $f_{\mathbf{v}(t_i)}$ is assumed to be Gaussian. Thus, $f_{\mathbf{x}(t_i),\mathbf{v}(t_i)|\mathbf{Z}(t_{i-1})}$ is a product of two Gaussian densities, and therefore, $f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}$ is Gaussian.

The conditional expectation $\mathrm{E}[\mathbf{z}(t_i)|\mathbf{Z}_{i-1}]$ can be calculated as:

$$\mathrm{E}[\mathbf{z}(t_i)|\mathbf{Z}_{i-1}] = \mathrm{E}\left[\left(\mathbf{H}(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i)\right)|\mathbf{Z}_{i-1}\right]$$

Which is can be evaluated by expanding the expectation operator and noting that $\mathbf{H}(t_i)$ is not random, and $\mathbf{v}(t_i)$ is a zero-mean white Gaussian measurement noise vector.

$$\mathrm{E}[\mathbf{z}(t_i)|\mathbf{Z}_{i-1}] = \mathbf{H}(t_i)\mathrm{E}[\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}]$$

Which can be rewritten as:

$$\mathrm{E}[\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}] = \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)$$

The conditional covariance is then:

$$\begin{aligned}
&\mathrm{E}\left[\left(\mathbf{z}(t_i) - \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{z}(t_i) - \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\right] \\
&\quad = \mathrm{E}\Big[\left(\mathbf{H}(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i)\right. \\
&\quad \left. - \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{H}(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i) - \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}|\mathbf{Z}(t_{i-1}) = \mathbf{Z}_{i-1}\Big]
\end{aligned}$$

Multiplying through and expanding out the terms, noting that $\mathbf{v}(t_i)$ is uncorrelated with $\mathbf{x}(t_i)$, gives:

$$\mathrm{E}\left[\left(\mathbf{z}(t_i)-\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{z}(t_i)-\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}|\mathbf{Z}_{i-1}\right]$$

$$=\ \mathrm{E}\left[\mathbf{H}(t_i)\left(\mathbf{x}(t_i)-\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{x}(t_i)-\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}\mathbf{H}(t_i)^{\mathrm{T}}\right.$$

$$\left.+\,\mathbf{v}(t_i)\mathbf{v}(t_i)^{\mathrm{T}}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}\right]$$

The expectation operator can then be split like so:

$$\mathrm{E}\left[\left(\mathbf{z}(t_i)-\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{z}(t_i)-\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}|\mathbf{Z}_{i-1}\right]$$

$$=\ \mathrm{E}\left[\mathbf{H}(t_i)\left(\mathbf{x}(t_i)-\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{x}(t_i)-\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}\mathbf{H}(t_i)^{\mathrm{T}}\right]$$

$$+\,\mathrm{E}[\mathbf{v}(t_i)\mathbf{v}(t_i)^{\mathrm{T}}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}]$$

Since $\mathbf{H}(t_i)$ is not random it can be taken out of the expectation operator.

$$\mathrm{E}\left[\left(\mathbf{z}(t_i)-\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{z}(t_i)-\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}\right]$$

$$=\ \mathbf{H}(t_i)\mathrm{E}\left[\left(\mathbf{x}(t_i)-\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{x}(t_i)-\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}\right]\mathbf{H}(t_i)^{\mathrm{T}}$$

$$+\,\mathrm{E}[\mathbf{v}(t_i)\mathbf{v}(t_i)^{\mathrm{T}}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}]$$

Then $\mathbf{P}(t_i^-)$ can be substituted for $\mathrm{E}\left[\left(\mathbf{x}(t_i)-\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{x}(t_i)-\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}\right]$

$$\mathrm{E}\left[\left(\mathbf{z}(t_i)-\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)\left(\mathbf{z}(t_i)-\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-)\right)^{\mathrm{T}}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}\right]$$

$$=\ \mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}}+\mathrm{E}[\mathbf{v}(t_i)\mathbf{v}(t_i)^{\mathrm{T}}|\mathbf{Z}(t_{i-1})=\mathbf{Z}_{i-1}]$$

Finally recall that the covariance of $\mathbf{v}(t_i)$ was defined as:

$$\mathrm{E}\left[\mathbf{v}(t_i)\mathbf{v}(t_j)^{\mathrm{T}}\right]=\begin{cases}\mathrm{R}(t_i) & t_i=t_j\\ 0 & t_i\neq t_j\end{cases}$$

Which can be inserted into the equation above to yield:

$$\mathrm{E}\left[\left(\mathbf{z}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\hat{\mathbf{x}}(\mathrm{t_i^-})\right)\left(\mathbf{z}(\mathrm{t_i}) - \mathbf{H}(\mathrm{t_i})\hat{\mathbf{x}}(\mathrm{t_i^-})\right)^{\mathrm{T}} | \mathbf{Z}(\mathrm{t_{i-1}}) = \mathbf{Z_{i-1}}\right] = \mathbf{H}(\mathrm{t_i})\mathbf{P}(\mathrm{t_i^-})\mathbf{H}(\mathrm{t_i})^{\mathrm{T}} + \mathbf{R}(\mathrm{t_i})$$

The density $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ has now been fully defined and is explicitly written as:

$$f_{\mathbf{z}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})=\mathbf{Z_i}} \qquad \textbf{Eq. 27}$$

$$= \frac{\exp\left(-\frac{1}{2}[\boldsymbol{\zeta} - \mathbf{H}(\mathrm{t_i})\hat{\mathbf{x}}(\mathrm{t_i^-})]^{\mathrm{T}}[\mathbf{H}(\mathrm{t_i})\mathbf{P}(\mathrm{t_i^-})\mathbf{H}(\mathrm{t_i})^{\mathrm{T}} + \mathbf{R}(\mathrm{t_i})]^{-1}[\boldsymbol{\zeta} - \mathbf{H}(\mathrm{t_i})\hat{\mathbf{x}}(\mathrm{t_i^-}\right.}{\sqrt{(2\pi)^{\mathrm{m}}\left|\mathbf{H}(\mathrm{t_i})\mathbf{P}(\mathrm{t_i^-})\mathbf{H}(\mathrm{t_i})^{\mathrm{T}} + \mathbf{R}(\mathrm{t_i})\right|}}$$

By showing that $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})=\mathbf{Z_{i-1}}}$, $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})=\mathbf{Z_{i-1}}}$ and $f$, it has been shown that $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_i})=\mathbf{Z_i}}$ is also Gaussian. Now all that is left is to explicitly define $\mathrm{f}_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_i})=\mathbf{Z_i}}$ and extract the mean and covariance matrix.

Recall:

$$f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_i})=\mathbf{Z_i}} = \frac{f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(\mathrm{t_{i-1}})} f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}}{f_{\mathbf{z}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}} \qquad \textbf{Eq. 28}$$

Plugging Eq. 24, Eq. 26, and Eq. 27 for the densities on the right-hand side of Eq. 28 results in:

$$f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_i})=\mathbf{Z_i}} = \frac{\sqrt{\left|\mathbf{H}(\mathrm{t_i})\mathbf{P}(\mathrm{t_i^-})\mathbf{H}(\mathrm{t_i})^{\mathrm{T}} + \mathbf{R}(\mathrm{t_i})\right|}}{\sqrt{(2\pi)^{\mathrm{n}}\left|\mathbf{P}(\mathrm{t_i^-})\right||\mathbf{R}(\mathrm{t_i})|}} \exp\left(-\frac{1}{2}\{[\boldsymbol{\zeta}\right. \qquad \textbf{Eq. 29}$$

$$- \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi}]^{\mathrm{T}}\mathbf{R}(\mathrm{t_i})^{-1}[\boldsymbol{\zeta} - \mathbf{H}(\mathrm{t_i})\boldsymbol{\xi}]$$

$$+ [\boldsymbol{\xi} - \hat{\mathbf{x}}(\mathrm{t_i^-})]^{\mathrm{T}}\mathbf{P}(\mathrm{t_i^-})^{-1}[\boldsymbol{\xi} - \hat{\mathbf{x}}(\mathrm{t_i^-})]$$

$$- [\boldsymbol{\zeta} - \mathbf{H}(\mathrm{t_i})\hat{\mathbf{x}}(\mathrm{t_i^-})]^{\mathrm{T}}[\mathbf{H}(\mathrm{t_i})\mathbf{P}(\mathrm{t_i^-})\mathbf{H}(\mathrm{t_i})^{\mathrm{T}} + \mathbf{R}(\mathrm{t_i})]^{-1}[\boldsymbol{\zeta}$$

$$\left. - \mathbf{H}(\mathrm{t_i})\hat{\mathbf{x}}(\mathrm{t_i^-})]\}\right)$$

Now begins the rigorous algebra to transform Eq. 29 into something that looks more obviously Gaussian.

First expand the terms in the exponential in Eq. 29. For readability, the time notation ($t_i$) has been removed from the following equations. $\hat{\mathbf{x}}^-$ and $\mathbf{P}^-$ are understood to mean $\hat{\mathbf{x}}(t_i^-)$ and $\mathbf{P}(t_i^-)$.

$$\begin{aligned}\{[\boldsymbol{\zeta}-\mathbf{H}\boldsymbol{\xi}]^{\mathrm{T}}\mathbf{R}^{-1}[\boldsymbol{\zeta}-\mathbf{H}\boldsymbol{\xi}]+[\boldsymbol{\xi}-\hat{\mathbf{x}}^-]^{\mathrm{T}}\mathbf{P}^{--1}[\boldsymbol{\xi}-\hat{\mathbf{x}}^-]-[\boldsymbol{\zeta}-\mathbf{H}\hat{\mathbf{x}}^-]^{\mathrm{T}}[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}[\boldsymbol{\zeta}-\mathbf{H}\hat{\mathbf{x}}^-]\}\\ =\{\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}-\boldsymbol{\xi}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}-\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\boldsymbol{\xi}+\boldsymbol{\xi}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\boldsymbol{\xi}+\boldsymbol{\xi}^{\mathrm{T}}\mathbf{P}^{--1}\boldsymbol{\xi}-\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{P}^{--1}\boldsymbol{\xi}\\ -\boldsymbol{\xi}^{\mathrm{T}}\mathbf{P}^{--1}\hat{\mathbf{x}}^-+\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{P}^{--1}\hat{\mathbf{x}}^--\boldsymbol{\zeta}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}\boldsymbol{\zeta}+\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}\boldsymbol{\zeta}\\ +\boldsymbol{\zeta}[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}\mathbf{H}\hat{\mathbf{x}}^--\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}\,\mathbf{H}\hat{\mathbf{x}}^-\}\end{aligned}$$

Then group like terms. Note that each term is a scalar, and that the transpose of a scalar is just the scalar itself. This equation then becomes

$$\begin{aligned}&[\boldsymbol{\zeta}-\mathbf{H}\boldsymbol{\xi}]^{\mathrm{T}}\mathbf{R}^{-1}[\boldsymbol{\zeta}-\mathbf{H}\boldsymbol{\xi}]+[\boldsymbol{\xi}-\hat{\mathbf{x}}^-]^{\mathrm{T}}\mathbf{P}^{--1}[\boldsymbol{\xi}-\hat{\mathbf{x}}^-] \qquad \textbf{Eq. 30}\\ &\quad-[\boldsymbol{\zeta}-\mathbf{H}\hat{\mathbf{x}}^-]^{\mathrm{T}}[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}[\boldsymbol{\zeta}-\mathbf{H}\hat{\mathbf{x}}^-]\\ &\quad=\boldsymbol{\zeta}^{\mathrm{T}}[\mathbf{R}^{-1}-[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}]\boldsymbol{\zeta}\\ &\quad+2\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}\boldsymbol{\zeta}\\ &\quad+\hat{\mathbf{x}}^{-\mathrm{T}}[\mathbf{P}^{--1}-\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^-\mathbf{H}+\mathbf{R}]^{-1}\mathbf{H}]\hat{\mathbf{x}}^-\\ &\quad+\boldsymbol{\xi}^{\mathrm{T}}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}+\mathbf{P}^{--1}]\boldsymbol{\xi}-2\boldsymbol{\xi}^{\mathrm{T}}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}+\mathbf{P}^{--1}\hat{\mathbf{x}}^-]\end{aligned}$$

Now consider the first term on the right hand side of Eq. 30:

$$\boldsymbol{\zeta}^{\mathrm{T}}[\mathbf{R}^{-1}-[\mathbf{H}\mathbf{P}^-\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}]\boldsymbol{\zeta}$$

The matrix inversion lemma, described by Maybeck, can be used to show the following matrix identities:

$$[\mathbf{P}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1} = \mathbf{P} - \mathbf{P}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\mathbf{H}\mathbf{P} \quad \textbf{Eq. 31}$$

$$\mathbf{H}[\mathbf{P}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}} = \mathbf{R} - \mathbf{R}[\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\mathbf{R} \quad \textbf{Eq. 32}$$

$$[\mathbf{P}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{P}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1} \quad \textbf{Eq. 33}$$

By pre- and post-multiplying Eq. 32 by $\mathbf{R}^{-1}$ the equation becomes:

$$[\mathbf{R}^{-1} - [\mathbf{H}\mathbf{P}^{-}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}] = \mathbf{R}^{-1}\mathbf{H}[\mathbf{P}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$$

Which can be plugged into the first term on the right-hand side of Eq. 30:

$$\boldsymbol{\zeta}^{\mathrm{T}}[\mathbf{R}^{-1} - [\mathbf{H}\mathbf{P}^{-}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}]\boldsymbol{\zeta} = \boldsymbol{\zeta}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}[\mathbf{P}^{--1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}$$

Now consider the second term in equation Eq. 30:

$$2\hat{\mathbf{x}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{-}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\boldsymbol{\zeta}$$

This term can be attained by pre-multiplying Eq. 33 by $\mathbf{P}^{-1}$:

$$2\hat{\mathbf{x}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{-}\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}\boldsymbol{\zeta} = 2\hat{\mathbf{x}}^{\mathrm{T}}\mathbf{P}^{--1}[\mathbf{P}^{--1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}$$

Lastly, consider the third term in Eq. 30:

$$\hat{\mathbf{x}}^{-\mathrm{T}}[\mathbf{P}^{--1} - \mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{-}\mathbf{H} + \mathbf{R}]^{-1}\mathbf{H}]\hat{\mathbf{x}}^{-1}$$

By pre- and post-multiplying Eq. 31 by $\mathbf{P}^{-1}$, the third term in Eq. 30 can be written as:

$$\hat{\mathbf{x}}^{-\mathrm{T}}[\mathbf{P}^{--1} - \mathbf{H}^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{-}\mathbf{H} + \mathbf{R}]^{-1}\mathbf{H}]\hat{\mathbf{x}} = \hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{P}^{--1}[\mathbf{P}^{--1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{P}^{--1}\hat{\mathbf{x}}^{-}.$$

The three terms that have been rearranged using the three matrix identities can be combined like so, remembering again that these terms are all scalars:

$$\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}[\mathbf{P}^{--1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}+\ 2\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{P}^{--1}[\mathbf{P}^{--1}+$$

$$\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}+\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{P}^{--1}[\mathbf{P}^{--1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{P}^{--1}\hat{\mathbf{x}}^{-}=\ [\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}+\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{P}^{--1}][\mathbf{P}^{--1}+$$

$$\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}+\mathbf{P}^{--1}\hat{\mathbf{x}}^{-1}]$$

Now these terms can be plugged back into the original expansion:

$$\begin{aligned}\{\,[\boldsymbol{\zeta}-\mathbf{H}\boldsymbol{\xi}]^{\mathrm{T}}\mathbf{R}^{-1}[\boldsymbol{\zeta}-\mathbf{H}\boldsymbol{\xi}]+[\boldsymbol{\xi}-\hat{\mathbf{x}}^{-}]^{\mathrm{T}}\mathbf{P}^{--1}[\boldsymbol{\xi}-\hat{\mathbf{x}}^{-}] \\ -[\boldsymbol{\zeta}-\mathbf{H}\hat{\mathbf{x}}^{-}]^{\mathrm{T}}[\mathbf{H}\mathbf{P}^{-}\mathbf{H}^{\mathrm{T}}+\mathbf{R}]^{-1}[\boldsymbol{\zeta}-\mathbf{H}\hat{\mathbf{x}}^{-}]\} \\ =\{\boldsymbol{\xi}^{\mathrm{T}}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}+\mathbf{P}^{--1}]\boldsymbol{\xi}-2\boldsymbol{\xi}^{\mathrm{T}}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}+\mathbf{P}^{--1}\hat{\mathbf{x}}^{-}] \\ +[\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}+\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{P}^{--1}][\mathbf{P}^{--1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta} \\ +\mathbf{P}^{--1}\hat{\mathbf{x}}^{-}]\}\end{aligned} \qquad \textbf{Eq. 34}$$

The n-vector **a** and the n-by-n matrix **A** will be defined like so to make the algebra easier to follow.

$$\mathbf{a}=\ [\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}+\mathbf{P}^{--1}\hat{\mathbf{x}}^{-}]$$

$$\mathbf{A}=\ [\mathbf{P}^{--1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{--1}\mathbf{H}]$$

Eq. 34 can then be written simply as:

$$\boldsymbol{\xi}^{\mathrm{T}}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}+\mathbf{P}^{--1}]\boldsymbol{\xi}-2\boldsymbol{\xi}^{\mathrm{T}}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}+\mathbf{P}^{--1}\hat{\mathbf{x}}^{-}]$$

$$+[\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}+\hat{\mathbf{x}}^{-\mathrm{T}}\mathbf{P}^{--1}][\mathbf{P}^{--1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1}[\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\boldsymbol{\zeta}+\mathbf{P}^{--1}\hat{\mathbf{x}}^{-}]$$

$$=\boldsymbol{\xi}^{\mathbf{T}}\mathbf{A}\boldsymbol{\xi}-2\boldsymbol{\xi}^{\mathbf{T}}\mathbf{a}+\mathbf{a}^{\mathbf{T}}\mathbf{A}^{-1}\mathbf{a}$$

Note that this is a quadratic equation in $\boldsymbol{\xi}$ and by multiplying the second and third terms by $\mathbf{A}\mathbf{A}^{-1}$, the equation above can be equivalently written as:

$$\boldsymbol{\xi}^{\mathbf{T}}\mathbf{A}\boldsymbol{\xi}-2\boldsymbol{\xi}^{\mathbf{T}}\mathbf{a}+\mathbf{a}^{\mathbf{T}}\mathbf{A}^{-1}\mathbf{a}=\ \boldsymbol{\xi}^{\mathbf{T}}\mathbf{A}\boldsymbol{\xi}-2\boldsymbol{\xi}^{\mathbf{T}}\mathbf{A}\mathbf{A}^{-1}\mathbf{a}+\mathbf{a}^{\mathbf{T}}\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{a}$$

Which can be factored to yield:

$$\boldsymbol{\xi}^{\mathbf{T}}\mathbf{A}\boldsymbol{\xi} - 2\boldsymbol{\xi}^{\mathbf{T}}\mathbf{A}\mathbf{A}^{-1}\mathbf{a} + \mathbf{a}^{\mathbf{T}}\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{a} = (\boldsymbol{\xi} - \mathbf{A}^{-1}\mathbf{a})^{\mathrm{T}}\mathbf{A}(\boldsymbol{\xi} - \mathbf{A}^{-1}\mathbf{a})$$

Thus $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}$ can be written as:

$$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)=\mathbf{z}_i} = \frac{\sqrt{\left|\mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}} + \mathbf{R}(t_i)\right|}}{\sqrt{(2\pi)^n\left|\mathbf{P}(t_i^-)\right||\mathbf{R}(t_i)|}} \exp(\boldsymbol{\xi} - \mathbf{A}^{-1}\mathbf{a})^{\mathrm{T}}\mathbf{A}(\boldsymbol{\xi} - \mathbf{A}^{-1}\mathbf{a}) \qquad \textbf{Eq. 35}$$

Time notation will now be resumed

It can be seen from Eq. 35 that mean and covariance of $\mathbf{x}(t_i)$ conditioned on $\mathbf{Z}(t_i) = \mathbf{Z}_i$**:**

$$\begin{aligned}\hat{\mathbf{x}}(t_i^+) = \ & \mathrm{E}[\mathbf{x}(t_i)|\mathbf{z}(t_i) = \mathbf{z}_i] = \mathbf{A}^{-1}\mathbf{a} \\ = \ & [\mathbf{P}(t_i^-)^{-1} + \mathbf{H}(t_i)^{\mathrm{T}}\mathbf{R}(t_i)^{-1}\mathbf{H}(t_i)]^{-1}[\mathbf{H}(t_i)^{\mathrm{T}}\mathbf{R}(t_i)^{-1}\mathbf{z}_i \\ & + \mathbf{P}(t_i^-)^{-1}\hat{\mathbf{x}}(t_i^-)\ ]\end{aligned} \qquad \textbf{Eq. 36}$$

And:

$$\begin{aligned}\mathbf{P}(t_i^+) = \ & \mathrm{E}\left[\left(\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)\right)\left(\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)\right)^{\mathrm{T}}|\mathbf{z}(t_i) = \mathbf{z}_i\right] = \mathbf{A}^{-1} \\ & = [\mathbf{P}(t_i^-)^{-1} + \mathbf{H}(t_i)^{\mathrm{T}}\mathbf{R}(t_i)^{-1}\mathbf{H}(t_i)]^{-1}\end{aligned} \qquad \textbf{Eq. 37}$$

Expand Eq. 36:

$$\begin{aligned}\hat{\mathbf{x}}(t_i^+) = \ & [\mathbf{P}(t_i^-)^{-1} + \mathbf{H}(t_i)^{\mathrm{T}}\mathbf{R}(t_i)^{-1}\mathbf{H}(t_i)]^{-1}\mathbf{P}(t_i^-)^{-1}\hat{\mathbf{x}}(t_i^-) \\ & + [\mathbf{P}(t_i^-)^{-1} + \mathbf{H}(t_i)^{\mathrm{T}}\mathbf{R}(t_i)^{-1}\mathbf{H}(t_i)]^{-1}\mathbf{H}(t_i)^{\mathrm{T}}\mathbf{R}(t_i)^{-1}\mathbf{z}_i\end{aligned}$$

Then substituting Eq. 31 and Eq. 33 gives:

$$\begin{aligned}\hat{\mathbf{x}}(t_i^+) = & [\mathbf{P}(t_i^-) - \mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}}[\mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}} + \mathbf{R}(t_i)]^{-1}\mathbf{H}(t_i)\mathbf{P}(t_i^-)]\mathbf{P}(t_i^-)^{-1}\hat{\mathbf{x}}(t_i^-) \\ & + [\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}}[\mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}} + \mathbf{R}(t_i)]^{-1}]\ \mathbf{z}_i\end{aligned}$$

Then post-multiplying the $\mathbf{P}(t_i^-)^{-1}\hat{\mathbf{x}}(t_i^-)$ through and regrouping terms gives:

$$\hat{\mathbf{x}}(t_i^+) = \hat{\mathbf{x}}(t_i^-) + [\mathbf{P}(t_i^-)\mathbf{H}(t_i)^T[\mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^T + \mathbf{R}(t_i)]^{-1}][\mathbf{z_i} - \mathbf{H}(t_i)\mathbf{x}(t_i^-)]$$

The same can be done for $\mathbf{P}(t_i^+)$ by substituting Eq. 31.

$$\mathbf{P}(t_i^+) = \mathbf{P}(t_i^-) - \mathbf{P}(t_i^-)\mathbf{H}(t_i)^T[\mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^T + \mathbf{R}(t_i)]^{-1}\mathbf{H}(t_i)\mathbf{P}(t_i^-)$$

Note that $\mathbf{P}(t_i^-)$ was assumed to be positive definite in this derivation. If $\mathbf{P}(t_{i-1}^-)$ is assumed to be positive definite, then by analyzing:

$$P(t_i^-) = \mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^+)\mathbf{\Phi}(t_i, t_{i-1})^T + \int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i, \tau)\mathbf{G}(\tau)\mathbf{Q}(t)\mathbf{G}(\tau)^T\mathbf{\Phi}(t_i, \tau)^T d\tau$$

It can be seen that $\mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^+)\mathbf{\Phi}(t_i, t_{i-1})^T$ is positive definite since the state transition matrix is always positive definite, and $\int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i, \tau)\mathbf{G}(\tau)\mathbf{Q}(t)\mathbf{G}(\tau)^T\mathbf{\Phi}(t_i, \tau)^T d\tau$ is at worst positive semi-definite since $\mathbf{Q}(t)$ is at worst positive semidefinite[12] . This makes $\mathbf{P}(t_i^-)$, and its inverse $\mathbf{P}(t_i^-)^{-1}$, positive definite if $\mathbf{P}(t_{i-1}^-)$ is positive definite.

Then by analyzing Eq. 37 it can be easily seen that if $\mathbf{P}(t_i^{-1})$ is positive definite then $\mathbf{P}(t_i^+)$ must be positive definite since $\mathbf{H}(t_i)^T\mathbf{R}(t_i)^{-1}\mathbf{H}(t_i)$ must be positive definite since $\mathbf{R}(t_i)$ was assumed to be positive definite:

$$\mathbf{P}(t_i^+) = [\mathbf{P}(t_i^-)^{-1} + \mathbf{H}(t_i)^T\mathbf{R}(t_i)^{-1}\mathbf{H}(t_i)]^{-1}$$

Therefore if $\mathbf{P}(t_0)$ is positive definite, the derivation holds.

## 4.3 Kalman filter algorithm

The Kalman filter algorithm works by using a prediction-correction scheme. First, the current state estimate and covariance matrix are predicted by propagating the previous state estimate and covariance matrix using Eq. 38 and Eq. 39:

$$\hat{\mathbf{x}}(t_i^-) = \boldsymbol{\Phi}(t_i, t_{i-1})\hat{\mathbf{x}}(t_{\mathbf{i-1}}^+) \quad \text{Eq. 38}$$

$$\mathbf{P}(t_i^-) = \boldsymbol{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^+)\boldsymbol{\Phi}(t_i, t_{i-1})^T + \int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i, \tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}(\tau)^T\boldsymbol{\Phi}(t_i, \tau)^T d\tau \quad \textbf{Eq. 39}$$

Then after taking a measurement, the current state estimate and covariance are corrected by applying Eq. 40 and Eq. 41:

$$\hat{\mathbf{x}}(t_i^+) = \hat{\mathbf{x}}(t_i^-) + \mathbf{K}(t_i)[\mathbf{z_i} - \mathbf{H}(t_i)\mathbf{x}(t_i^-)] \quad \textbf{Eq. 40}$$

$$\mathbf{P}(t_i^+) = \mathbf{P}(t_i^-) - \mathbf{K}(t_i)\mathbf{H}(t_i)\mathbf{P}(t_i^-) \quad \textbf{Eq. 41}$$

Where $\mathbf{K}(t_i)$ is the Kalman filter gain defined as:

$$\mathbf{K}(t_i) = \mathbf{P}(t_i^-)\mathbf{H}(t_i)^T[\mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^T + \mathbf{R}(t_i)]^{-1}$$

The Kalman filter gain is a weighting term that generates a correction term to update $\hat{\mathbf{x}}(t_i^-)$.

Note that as $\mathbf{P}(t_i^-) \to 0$ (i.e. the estimate $\hat{\mathbf{x}}(t_i^-)$ is more certain), $\mathbf{K}(t_i) \to 0$ and $\hat{\mathbf{x}}(t_i^+) = \hat{\mathbf{x}}(t_i^-)$. However as $\mathbf{R}(t_i) \to 0$ (i.e. the measurement is more certain), $\mathbf{K}(t_i) \to \mathbf{H}(t_i)^-$ and thus $\hat{\mathbf{x}}(t_i^+) = \mathbf{z_i}$.

Eq. 38 and Eq. 39 make up the prediction step of the Kalman filter. However, it is worth noting that it would be more convenient to represent Eq. 38 and Eq. 39 in the form of a differential equation. This would make it easier to propagate the state estimate and the covariance matrix through a numerical integrator such as MATLAB's ODE45 function. To get Eq. 38 and Eq. 39 in the form of differential equations, it is necessary to define terms $\hat{\mathbf{x}}(t/t_{i-1})$ and $\mathbf{P}(t/t_{i-1})$ that are valid for any $t \in [t_{i-1}, t_i)$. That way the differential equation can be integrated from time $t_{i-1}$ to times between $t_{i-1}$ and $t_i$.

$$\hat{\mathbf{x}}(t/t_{i-1}) = \boldsymbol{\Phi}(t, t_{i-1})\hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+})$$

$$\mathbf{P}(t/t_{i-1}) = \boldsymbol{\Phi}(t, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^{+})\boldsymbol{\Phi}(t, t_{i-1})^{T} + \int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t,\tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}(\tau)^{T}\boldsymbol{\Phi}(t,\tau)^{T}d\tau$$

The derivatives can be taken like so:

$$\frac{d}{dt}\left(\hat{\mathbf{x}}(t/t_{i-1})\right) = \frac{d}{dt}\left(\boldsymbol{\Phi}(t, t_{i-1})\hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+})\right)$$

$$\frac{d}{dt}\left(\hat{\mathbf{x}}(t/t_{i-1})\right) = \mathbf{F}(t)\boldsymbol{\Phi}(t, t_{i-1})\hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+})$$

Note that $\boldsymbol{\Phi}(t, t_{i-1})\hat{\mathbf{x}}(t_{\mathbf{i-1}}^{+}) = \hat{\mathbf{x}}(t/t_{i-1})$ thus:

$$\frac{d}{dt}\left(\hat{\mathbf{x}}(t/t_{i-1})\right) = \mathbf{F}(t)\hat{\mathbf{x}}(t \qquad \textbf{Eq. 42}$$

$$/t_{i-1})$$

For the covariance matrix:

$$\frac{d}{dt}\left(\mathbf{P}(t/t_{i-1})\right) = \boldsymbol{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^{+})\boldsymbol{\Phi}(t_i, t_{i-1})^{T} + \int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i,\tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}(\tau)^{T}\boldsymbol{\Phi}(t_i,\tau)^{T}d\tau$$

$$\frac{d}{dt}\left(\mathbf{P}(t/t_{i-1})\right) = \mathbf{F}(t)\boldsymbol{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^{+})\boldsymbol{\Phi}(t_i, t_{i-1})^{T} + \boldsymbol{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^{+})\boldsymbol{\Phi}(t_i, t_{i-1})^{T}\mathbf{F}(t_i)^{T}$$

$$+ \int_{t_{i-1}}^{t_i} \mathbf{F}(t)\boldsymbol{\Phi}(t_i,\tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}(\tau)^{T}\boldsymbol{\Phi}(t_i,\tau)^{T}d\tau$$

$$+ \int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i,\tau)\mathbf{G}(\tau)\mathbf{Q}(\tau)\mathbf{G}(\tau)^{T}\boldsymbol{\Phi}(t_i,\tau)^{T}\mathbf{F}(t)d\tau + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}(t)$$

Note that Leibniz's rule was used for the integral term.

Grouping terms yields:

$$\frac{d}{dt}\left(\mathbf{P}(t/t_{i-1})\right) = \mathbf{F}(t)\left[\mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^{+})\mathbf{\Phi}(t_i, t_{i-1})^{T}\right.$$

$$\left. + \int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i, \tau)\mathbf{G}(\tau)\mathbf{Q}(t)\mathbf{G}(\tau)^{T}\mathbf{\Phi}(t_i, \tau)^{T}d\tau\right]$$

$$+ \left[\mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{\mathbf{i-1}}^{+})\mathbf{\Phi}(t_i, t_{i-1})^{T}\right.$$

$$\left. + \int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i, \tau)\mathbf{G}(\tau)\mathbf{Q}(t)\mathbf{G}(\tau)^{T}\mathbf{\Phi}(t_i, \tau)^{T}d\tau\right]\mathbf{F}(t)^{T} + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}(t)$$

Substituting $\mathbf{P}(t/t_{i-1})$ for the terms in brackets results in:

$$\frac{d}{dt}\left(\mathbf{P}(t/t_{i-1})\right) = \mathbf{F}(t)\,\mathbf{P}(t/t_{i-1}) + \mathbf{P}(t/t_{i-1})\mathbf{F}(t)^{T} + \mathbf{G}(t)\mathbf{Q}(t)\mathbf{G}(t) \qquad \textbf{Eq. 43}$$

Eq. 42 and Eq. 43 can readily be integrated with a numerical integrator starting from the initial conditions:

$$\hat{\mathbf{x}}(t_{i-1}/t_{i-1}) = \hat{\mathbf{x}}(t_{i-1})$$

$$\mathbf{P}(t_{i-1}/t_{i-1}) = \mathbf{P}(t_{i-1})$$

## 4.4 Extensions of the Kalman Filter

The Kalman filter is an optimal estimation algorithm for linear systems. For linear systems with Gaussian noise, the Kalman filter is able to minimize the mean square error of the estimate, which provides an optimal estimate of the system. However, in many applications the system dynamics and observation equations are nonlinear. The Kalman filter as derived is not suitable for nonlinear estimation. This problem has lead to the development of the extended Kalman filter (EKF).

The EKF helps resolves the nonlinear estimation issue by linearizing the system dynamics and the observation equations around the current predicted value, although this approximation results in a non-optimal estimate.

The system dynamics are given by the nonlinear differential equation:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{w}(t) \quad \textbf{Eq. 44}$$

The observation equation is given by:

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), t) + \mathbf{v}(t) \quad \textbf{Eq. 45}$$

As before, it is desired to propagate both the state estimate and the covariance matrix to the next time instance so the Kalman gain $\mathbf{K}(t_i)$ can be computed. However, the system is nonlinear, which means the conditional probability density, $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}$, can no longer be assumed Gaussian. This creates a problem, as it may require a boundless number of moments to be calculated to define the distribution (Julier), which means an optimal estimate cannot be feasibly computed.

To work around this, it is first assumed that the initial state process vector $\mathbf{x}(t_0)$ is Gaussian with mean $\hat{\mathbf{x}}(t_0)$ and covariance $\mathbf{P}(t_0)$. Then the initial state can be propagated forward integrating the system dynamics.

$$\hat{\mathbf{x}}(t_i^-) = \mathbf{f}(\hat{\mathbf{x}}(t_{i-1}^+), t) \quad \textbf{Eq. 46}$$

For the initial iteration, $\hat{\mathbf{x}}(t_{i-1}^+) = \mathbf{x}(t_0)$.

Propagating the covariance matrix this time around is slightly more difficult, as the function $\mathbf{f}(\hat{\mathbf{x}}(t_{i-1}^+), t)$ cannot be directly applied. To generate the predicted covariance matrix, the EKF linearizes the system dynamics and the observation equations around the current estimate.

$$\mathbf{F}\left(\hat{\mathbf{x}}(t_{i-1}^{+})\right) = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}(t_{i-1}^{+})}$$

The covariance matrix is then propagated by:

$$\mathbf{P}(t_i^{-}) = \mathbf{F}\left(\hat{\mathbf{x}}(t_{i-1}^{+})\right)\mathbf{P}(t_{i-1})\mathbf{F}\left(\hat{\mathbf{x}}(t_{i-1}^{+})\right)^{T} + \mathbf{Q}(t_{i-1}) \qquad \textbf{Eq. 47}$$

Then the state estimate and the covariance matrix are updated with the Kalman gain. However, the observation equation cannot be directly applied here either. The observation equation is linearized around $\hat{\mathbf{x}}(t_i^{-})$. Which gives the matrix:

$$\mathbf{H}\left(\hat{\mathbf{x}}(t_i^{-})\right) = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\hat{\mathbf{x}}(t_i^{-})}$$

This is then used to compute the Kalman gain.

$$\mathbf{K}(t_i) = \mathbf{P}(t_i^{-})\mathbf{H}\left(\hat{\mathbf{x}}(t_i^{-})\right)^{T}\left[\mathbf{H}\left(\hat{\mathbf{x}}(t_i^{-})\right)\mathbf{P}(t_i^{-})\mathbf{H}\left(\hat{\mathbf{x}}(t_i^{-})\right)^{T} + \mathbf{R}(t_i)\right]^{-1}$$

Finally the Kalman Gain is used to update the estimate and the covariance matrix:

$$\hat{\mathbf{x}}(t_i^{+}) = \hat{\mathbf{x}}(t_i^{-}) + \mathbf{K}(t_i)\left[\mathbf{z_i} - \mathbf{H}\left(\hat{\mathbf{x}}(t_i^{-})\right)\mathbf{x}(t_i^{-})\right] \qquad \textbf{Eq. 48}$$

$$\mathbf{P}(t_i^{+}) = \mathbf{P}(t_i^{-}) - \mathbf{K}(t_i)\mathbf{H}\left(\hat{\mathbf{x}}(t_i^{-})\right)\mathbf{P}(t_i^{-}) \qquad \textbf{Eq. 49}$$

# 5 The Sequential Monte Carlo Filter (Particle filter)

This section will provide a brief overview of the Sequential Monte Carlo filter followed by a detailed derivation.

## 5.1 Overview

The Monte Carlo method is a statistical sampling technique that has been utilized for decades to successfully solve scientific problems that cannot be solved analytically[5] . The Monte Carlo method was pioneered by Stanislaw Ulam in 1946 during his time at the Los Alamos Laboratory[5] . Ulam pondered the use of statistical sampling to determine the odds of winning a game of solitaire. He then considered a more general application of statistical sampling to other physical phenomenon, namely neutron diffusion[5] . With the advent of more powerful computers, the concept was implemented by John von Neumann to predict the explosive behavior of fission weapons[5] .

In the Sequential Monte Carlo method, statistical samples are drawn from a distribution just like the Monte Carlo method. This time, however, a genetic weighting technique assigns a weight to each sample. The weight represents the probability of that particle being selected during the update step. Samples that are taken from this set of samples represent samples from a posterior distribution.

In 1993, N.J. Gordon et al developed a Bayesian variant of the Sequential Monte Carlo method called the bootstrap filter that would not be restricted to the linear assumptions of the Kalman filter[8] .

An explicit derivation of the particle filter is given in the following section.

## 5.2 Derivation

Note the derivation here was adapted from several sources to be consistent with the notation in Maybeck. Samples are denoted with the superscript "k", whereas the subscript "i" denotes discrete times when measurements are available.

Consider a general dynamical system and observation equation

$$\mathbf{x}(\mathrm{t_i}) = \mathbf{f}(\mathbf{x}(\mathrm{t_{i-1}}), \mathbf{w}(\mathrm{t_{i-1}}))$$

$$\mathbf{z}(\mathrm{t_i}) = \mathbf{h}(\mathbf{x}(\mathrm{t_{i-1}}), \mathbf{v}(\mathrm{t_{i-1}}))$$

Where $\mathbf{w}(\mathrm{t})$ and $\mathbf{v}(\mathrm{t})$ are two zero-mean white Gaussian process and measurement noise vectors, respectively, independent of both $\mathbf{x}(\mathrm{t_i})$ and $\mathbf{z}(\mathrm{t_i})$, with covariance matrices $\mathbf{Q}(\mathrm{t})$ and $\mathbf{R}(\mathrm{t})$, respectively.

Ultimately the goal of recursive Bayesian estimation is to estimate to some certainty the state of the system, $\mathbf{x}(\mathrm{t_i})$, given the measurements $\mathbf{Z}(\mathrm{t_i})$. Thus we are interested in determining the probability density function $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_i})}$. The probability function $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_i})}$ can be obtained using a prediction-correction scheme. $\mathbf{Z_i}$ will again be used to denote a realization of the process vector $\mathbf{Z}(\mathbf{t_i})$.

First consider the conditional probability density function $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-i}})}$. This conditional probability density function can be obtained by marginalizing the following joint conditional density function:

$$f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-i}})} = \int_{\infty}^{-\infty} f_{\mathbf{x}(\mathrm{t_i}),\mathbf{x}(\mathrm{t_{i-1}})|\mathbf{Z}(\mathrm{t_{i-1}})} d\mathbf{x}(\mathrm{t_{i-1}})$$

The joint conditional density function $f_{\mathbf{x}(\mathrm{t_i}),\mathbf{x}(\mathrm{t_{i-1}})|\mathbf{z}(\mathrm{t_{i-1}})}$ can then be split up using Bayes' rule like so:

$$f_{\mathbf{x}(t_i),\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})} = \frac{f_{\mathbf{x}(t_i),\mathbf{x}(t_{i-1}),\mathbf{Z}(t_{i-1})}}{f_{\mathbf{Z}(t_{i-1})}} = \frac{f_{\mathbf{x}(t_i)|\mathbf{x}(t_{i-1}),\mathbf{Z}(t_{i-1})} f_{\mathbf{x}(t_{i-1}),\mathbf{Z}(t_{i-1})}}{f_{\mathbf{Z}(t_{i-1})}}$$

$$f_{\mathbf{x}(t_i),\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})} = \frac{f_{\mathbf{x}(t_i)|\mathbf{x}(t_{i-1}),\mathbf{Z}(t_{i-1})} f_{\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})} f_{\mathbf{Z}(t_{i-1})}}{f_{\mathbf{Z}(t_{i-1})}}$$

And finally:

$$f_{\mathbf{x}(t_i),\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})} = f_{\mathbf{x}(t_i)|\mathbf{x}(t_{i-1})} f_{\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})}$$

Note that here that the substitution $f_{\mathbf{x}(t_i)|\mathbf{x}(t_{i-1}),\mathbf{Z}(t_{i-1})} = f_{\mathbf{x}(t_i)|\mathbf{x}(t_{i-1})}$ was made because the system is assumed to be a Markov process. The previous state encapsulates all of the required information needed in computing the next state. Thus $f_{\mathbf{x}(t_i)\mathrm{Z}(t_{i-i})}$ becomes:

$$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-i})} = \int_{\infty}^{-\infty} f_{\mathbf{x}(t_i)|\mathbf{x}(t_{i-1})} f_{\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})} d\mathbf{x}(t_{i-1})$$

Then the density updated $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}$ can be obtained using Bayes' rule as before:

$$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)} = \frac{f_{\mathbf{z}(t_i)|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})} f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}}$$

If the system and observation equations are linear and Gaussian, the distributions in the integrand above are completely defined by their respective means and covariances. As shown in the Kalman filter derivation:

$$f_{\mathbf{x}(t_i)|\mathbf{x}(t_{i-1})} \sim N\left(\hat{\mathbf{x}}(t_i^-), \mathbf{P}(t_i^-)\right)$$

$$f_{\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})} \sim N(\hat{\mathbf{x}}(t_{i-1}^+), \mathbf{P}(t_{i-1}^+))$$

$$f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})} \sim N(\mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-), \mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}} + \mathbf{R}(t_i))$$

Thus the prediction and correction equations can be solved for analytically, leading to the Kalman filter.

Unfortunately, the system dynamics and observation equations of interest are rarely ever linear and Gaussian. Thus an analytical solution is not possible. Many sub-optimal approaches have been developed, such as the EKF, to combat this problem, relying on linear approximations of the model to generate an estimate.

The approach described here is a numerical approach known as the Sequential Monte Carlo filter or the particle filter.

Suppose the independent and identically distributed samples $\mathbf{x}(\mathrm{t_{i-1}})^{(k)}$, for k = 1, 2, 3…*N*, are drawn from the distribution $f_{\mathbf{x}(\mathrm{t_{i-1}})|\mathbf{Z}(\mathrm{t_{i-1}})}$.

If $\mathbf{x}(\mathrm{t_{i-1}})^{(k)}$ is a sample distributed as $f_{\mathbf{x}(\mathrm{t_{i-1}})|\mathbf{Z}(\mathrm{t_{i-1}})}$ and $\mathbf{w}(\mathrm{t_{i-1}})^{(k)}$ is a sample distributed as $f_{\mathbf{w}(\mathrm{t_{i-1}})}$, then the sample $\mathbf{x}(\mathrm{t_i})^{(k)} = \mathbf{f}(\mathbf{x}(\mathrm{t_{i-1}})^{(k)}, \mathbf{w}(\mathrm{t_{i-1}})^{(k)})$, must be distributed as $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$[8] . Essentially, samples from $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ can be obtained by sampling from $f_{\mathbf{x}(\mathrm{t_{i-1}})|\mathbf{Z}(\mathrm{t_{i-1}})}$ and propagating each sample through the system dynamics equation.

For a very large *N*, the distribution $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$ can be approximated informally as a sum of Dirac functions (Godsill).

$$f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})} \approx \frac{1}{N}\sum_{k=1}^{N} \delta_{\mathbf{x}(\mathrm{t})^{(k)}}(\mathbf{x})$$

Dirac functions have the following properties

$$\delta(\mathrm{t}) = \begin{cases} 0 & \mathrm{t} \neq 0 \\ \infty & \mathrm{t} = 0 \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(\mathrm{t}) d\mathrm{t} = 1$$

It is understood that Dirac functions are not real functions, as no function can simultaneously satisfy both of the conditions above[4] .

Thus the expectation of $\mathbf{x}(\mathrm{t_i})$ conditioned on $\mathbf{Z}(\mathrm{t_{i-1}})$ can be computed:

$$\mathrm{E}\big(\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})\big) = \int_{-\infty}^{\infty} \mathbf{x}\, f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})} d\mathbf{x} \approx \int_{-\infty}^{\infty} \mathbf{x} \frac{1}{N} \sum_{k=1}^{N} \delta_{\mathbf{x}(\mathrm{t_i})^{(k)}}(\mathbf{x})\, d\mathbf{x}$$

Due to the properties of the Dirac delta function this becomes a familiar representation of a sample mean.

$$\mathrm{E}\big(\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})\big) \approx \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}(\mathrm{t_i})^{(k)}$$

Thus the estimate from the prediction step is simply the sample mean.

$$\hat{\mathbf{x}}(\mathrm{t_i^-}) = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}(\mathrm{t_i})^{(k)}$$

With a plethora of samples from $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{Z}(\mathrm{t_{i-1}})}$, an estimate $\hat{\mathbf{x}}(\mathrm{t_i^-})$**,** and a recently acquired measurement $\mathbf{z}(\mathrm{t_i})$, it is now desired to take these samples and update them into samples from $f_{\mathbf{x}(\mathrm{t_i})|\mathbf{z}(\mathrm{t_i})}$. How can this be done? Smith and Gordon demonstrate how this can be achieved[1] .

Consider a scenario where there a set of samples $\{x^{(i)}: i = 1, \dots, N\}$ are drawn from a distribution $g(x)$. It is desired to take these samples from $g(x)$ and form samples from a distribution $h(x)$. Given the samples $\{x^{(i)}: i = 1, \dots, N\}$, the distribution $g(x)$ and a positive function $f(x)$ that can

be normalized to the distribution $h(x) = \frac{f(x)}{\int_{-\infty}^{\infty} f(x)dx}$, how can one create samples of $h(x)$, or at least approximate samples of $h(x)$, from samples of $g(x)$? The answer is yes.

By applying a weight $\omega^{(i)} = f\left(x^{(i)}\right)/g\left(x^{(i)}\right)$ to each of the samples $\{x^{(i)}: i = 1, \dots, N\}$ and then drawing samples from $\{x^{(i)}: i = 1, \dots, N\}$ with probability $\omega^{(i)} = f\left(x^{(i)}\right)/g\left(x^{(i)}\right)$, the resampled particles can will approximately from the distribution $h(x)$.

To understand how this works, first consider the cumulative distribution function of a given sample $x^*$ taken from $g(x)$ by using an unweighted sampling approximation.

$$\Pr(x^* \leq a) = \frac{1}{N}\sum_{i=1}^{N} 1_{(-\infty,a]}\left(x^{(i)}\right)$$

The term $1_{(-\infty,a]}\left(x^{(i)}\right)$ is known as the indicator function. Essentially $1_{(-\infty,a]}\left(x^{(i)}\right)$ = 1 where ever $x^{(i)}$ is in the set $(-\infty, a]$ and $1_{(-\infty,a]}\left(x^{(i)}\right) = 0$ everywhere else. Intuitively, this just means that the probability of a specific event occurring is equal to the number of times it occurred in a sample divided by the total number of samples.

If $N \to \infty$ this equation then becomes an expectation:

$$\frac{1}{N}\sum_{i=1}^{N} 1_{(-\infty,a]}\left(x^{(i)}\right) \quad \underset{N \to \infty}{\to} \quad \mathrm{E}\left(1_{(-\infty,a]}(x)\right)$$

Using the definition of the expectation operator:

$$\mathrm{E}\left(1_{(-\infty,a]}(x)\right) = \int_{-\infty}^{\infty} 1_{(-\infty,a]}(x)g(x)dx$$

Now this can be simplified using the properties of the indicator function.

$$\mathrm{E}\left(1_{(-\infty,a]}(x)\right) = \int_{-\infty}^{a} g(x)dx$$

Thus as $N \to \infty$:

$$\Pr(x \leq a) = \int_{-\infty}^{a} g(x)dx$$

This intuitively makes sense. It is simply the relationship between the cumulative density function and its probability density function.

Now consider the cumulative distribution function of $x^*$ from $g(x)$ given a weighted sampling approximation:

$$\Pr(x^* \leq a) = \sum_{i=1}^{N} q^{(i)} 1_{(-\infty,a]}\left(x^{(i)}\right)$$

Here the parameter $q^{(i)}$ is defined as:

$$q^{(i)} = \frac{\omega^{(i)}}{\sum_{j=1}^{N} \omega^{(j)}}$$

And $\omega^{(i)}$ as:

$$\omega^{(i)} = \frac{f\left(x^{(i)}\right)}{g(x^{(i)})}$$

Substituting the definition of $q^{(i)}$ into the cumulative distribution function gives:

$$\Pr(x^* \leq a) = \sum_{i=1}^{N} \frac{\omega^{(i)}}{\sum_{j=1}^{N} \omega^{(j)}} 1_{(-\infty,a]}\left(x^{(i)}\right)$$

The term $1_{(-\infty,a]}\left(x^{(i)}\right)$ is known as the indicator function. Essentially, $1_{(-\infty,a]}\left(x^{(i)}\right) = 1$ everywhere $x^{(i)}$ is in the set $(-\infty, a]$ and $1_{(-\infty,a]}\left(x^{(i)}\right) = 0$ everywhere else.

Note that $\omega^{(j)}$ does not depend on the parameter $i$ and can be removed from the summation like so:

$$\Pr(x^* \leq a) = \frac{\sum_{i=1}^{N} \omega^{(i)} 1_{(-\infty,a]}\left(x^{(i)}\right)}{\sum_{j=1}^{N} \omega^{(j)}}$$

Multiplying the numerator and denominator by $1/N$ results in:

$$\Pr(x^* \leq a) = \frac{\frac{1}{N}\sum_{i=1}^{N} \omega^{(i)} 1_{(-\infty,a]}\left(x^{(i)}\right)}{\frac{1}{N}\sum_{j=1}^{N} \omega^{(j)}}$$

Note that summation terms resemble the sample mean. Then as $N \to \infty$, the equations becomes:

$$\omega^{(i)} = \frac{f\left(x^{(i)}\right)}{g(x^{(i)})} \quad \underset{N \to \infty}{\to} \quad \frac{f(x)}{g(x)}$$

$$\Pr(x^* \leq a) = \frac{\frac{1}{N}\sum_{i=1}^{N} \omega^{(i)} 1_{(-\infty,a]}\left(x^{(i)}\right)}{\frac{1}{N}\sum_{j=1}^{N} \omega^{(j)}} \quad \underset{N \to \infty}{\to} \quad \frac{\mathrm{E}\left(\frac{f(x)}{g(x)} 1_{(-\infty,a]}(x)\right)}{\mathrm{E}\left(\frac{f(x)}{g(x)}\right)}$$

Then the expression for the expectation operator can be substituted to yield:

$$\Pr(x^* \leq a) = \frac{\int_{-\infty}^{\infty} \frac{f(x)}{g(x)} 1_{(-\infty,a]}(x) g(x) dx}{\int_{-\infty}^{\infty} \frac{f(x)}{g(x)} g(x) dx}$$

The $g(x)$ cancel in the integrands:

$$\Pr(x^* \leq a) = \frac{\int_{-\infty}^{\infty} f(x) 1_{(-\infty,a]}(x) dx}{\int_{-\infty}^{\infty} f(x) dx}$$

Then the integrand in the numerator can be simplified using the properties of the indicator function:

$$\Pr(x^* \leq a) = \frac{\int_{-\infty}^{a} f(x) dx}{\int_{-\infty}^{\infty} f(x) dx}$$

And the grand finale:

$$\Pr(x^* \leq a) = \frac{\int_{-\infty}^{a} f(x) dx}{\int_{-\infty}^{\infty} f(x) dx} = \int_{-\infty}^{a} h(x) dx$$

Note that $\int_{-\infty}^{\infty} f(x) dx$ is a constant and can be pulled out of the integrand.

What this means is that as $N \to \infty$, samples drawn from the set $\{x^{(i)} : i = 1, \ldots, N\}$ with probability masses $q^{(i)}$ tend to the distribution $h(x)$. Thus:

$$q^{(i)} g(x) \propto h(x)$$

Or:

$$q^{(i)} \propto \frac{h(x)}{g(x)}$$

This result is invaluable as it is desired to take samples from a distribution $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}$ and obtain samples from $f_{\mathbf{x}(t_i)|\mathbf{z}(t_i)}$. Now all that remains is applying this result to the estimation problem.

If $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}$ is taken as $h(x)$ and $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}$ as $g(x)$ then:

$$q^{(k)} \propto \frac{f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}}{f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}}$$

Recall the expression for $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}$ using Bayes' rule.

$$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)} = \frac{f_{\mathbf{z}(t_i)|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})} f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}}$$

Substituting this in results in:

$$q^{(i)} \propto \frac{\frac{f_{\mathbf{z}(t_i)|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})} f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}}}{f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}}$$

Which can be simplified to:

$$q^{(i)} \propto \frac{f_{\mathbf{z}(t_i)|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}}$$

Then note that $f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}$ is a constant because it will never change as the probability of the measurements $\mathbf{z}(t_i)$ conditioned on a sample $\mathbf{x}(t_i)^{(k)}$ are evaluated. Thus the weighted term is just the likelihood of the measurement.

$$q^{(i)} \propto f_{\mathbf{z}(t_i)|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})}$$

Thus sampling from the set of points $\mathbf{x}(t_i)^{(i)}$ from the distribution $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}$ with probabilities equal to the likelihood $f_{\mathbf{z}(t_i)|\mathbf{x}(\mathbf{t_i}),\mathbf{Z}(t_{i-1})}$ will result in samples approximately from $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}$, with the caveat that $N$ must be very large for the approximation to be accurate, especially if the distribution $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}$ does not closely resemble $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}$ (Smith).

## 5.3 Particle filter algorithm

To summarize, the particle filter algorithm works like so:

1. Given $N$ samples $\mathbf{x}(\mathrm{t_{i-1}})^{(k)}$ from $f_{\mathbf{x}(\mathrm{t_{i-1}})|\mathbf{Z}(\mathrm{t_{i-1}})}$, where $\mathbf{Z}(\mathrm{t_{i-1}})$ represents a vector of all the previous measurements up to the current time step, propagate each sample $\mathbf{x}(\mathrm{t_{i-1}})^{(k)}$ through the system dynamics model $\mathbf{x}(\mathrm{t_i})^{(k)} = \mathbf{f}(\mathbf{x}(\mathrm{t_{i-1}})^{(i)}, \mathbf{w}(\mathrm{t_{i-1}}))$. The predicted estimate will be the sample mean of this new set of samples $\mathbf{x}(\mathrm{t_i})^{(i)}$

$$\hat{\mathbf{x}}(\mathrm{t_i^-}) = \frac{1}{N}\sum_{k=1}^{N}\mathbf{x}(\mathrm{t_i})^{(k)}$$

2. Now given the newly acquired measurement $\mathbf{z}(\mathrm{t_i})$, apply a weight $q^{(k)}$ to each sample $\mathbf{x}(\mathrm{t_i})^{(k)}$. The weight $q^{(k)}$ will be equal to the likelihood $f_{\mathbf{z}(\mathrm{t_i})|\mathbf{x}(\mathrm{t_i})^{(k)},\mathbf{Z}(\mathrm{t_{i-1}})}$.
3. Take $N$ samples from the set of $\mathbf{x}(\mathrm{t_i})^{(k)}$ with probability $q^{(k)}$, call these new samples $\mathbf{x}(\mathrm{t_i})^{(k)}$. This is the updated set of samples. The updated estimate will be the sample mean of this new set $\mathbf{x}(\mathrm{t_i})^{(j)}$.

$$\hat{\mathbf{x}}(\mathrm{t_i^-}) = \frac{1}{N}\sum_{j=1}^{N}\mathbf{x}(\mathrm{t_i})^{(j)}$$

## 5.4 Advantages and Disadvantages

The primary advantage of the particle filter is that it is not constrained to linear dynamical models. Given a dynamics model, a measurement model, measurements, and a suitable number of particles, the particle filter can estimate any number of parameters in a dynamical system. The Kalman filter cannot be used for non-linear, non-Gaussian systems, and the extended Kalman filter[8] may not produce accurate results for non-linear, non-Gaussian systems[9] .

One disadvantage is that the number of particles required to estimate the state grows exponentially as the dimensions of the state grows[9] . Consequently, a larger state space may take exponentially more time to estimate.

# 6 Methodology

The goal of this project was to expand upon the work of Dr. Furfaro and B. Gaudet by using particle filtering to detect the location and mass of a mass concentration embedded within the asteroid modeled as a point mass. Additionally, the ellipsoid's angular velocity and homogenous density were simultaneously estimated using the particle filter.

The asteroid was modeled as a rotating ellipsoid with a mass concentration embedded within its homogenous bulk. The angular velocity initially aligned with the Z-axis. The inhomogeneity was modeled as a point mass. The point mass was given a mass of 1/1000 the total mass of the homogenous bulk.

The spacecraft is set up to hover at various points above the asteroids surface at various times. By measuring the commanded thrust required to cancel out the gravitational forces acting on the spacecraft, the particle filter could be used to determine the location and mass of the point mass inhomogeneity, in addition to the density of the ellipsoidal mass and angular velocity.

The method of generating measurements is similar to the technique described by Dr. Furfaro and Gaudet. In order for the spacecraft to hover above a point in the asteroid body frame, the commanded thrust must on average cancel out with the gravitational forces. In this simulation it was assumed that the gravitational forces and the commanded thrust cancel each other out, and by measuring the amount of commanded thrust, the gravitational force on the spacecraft could be measured.

Table 1 gives an overview of the dimensions of the ellipsoidal asteroid. Figure 1 illustrates the ellipsoidal asteroid (shown in blue) and the point mass concentration (red asterisk).

| Parameter | Value | Description |
|---|---|---|
| **a** | 348 | Length of first ellipsoid semi axis meters (X-axis). |
| **b** | 287 | Length of second ellipsoid semi axis meters (Y-axis). |
| **c** | 250 | Length of third ellipsoid semi axis meters. (Z-axis). |

**Table 1: Asteroid Dimensions**

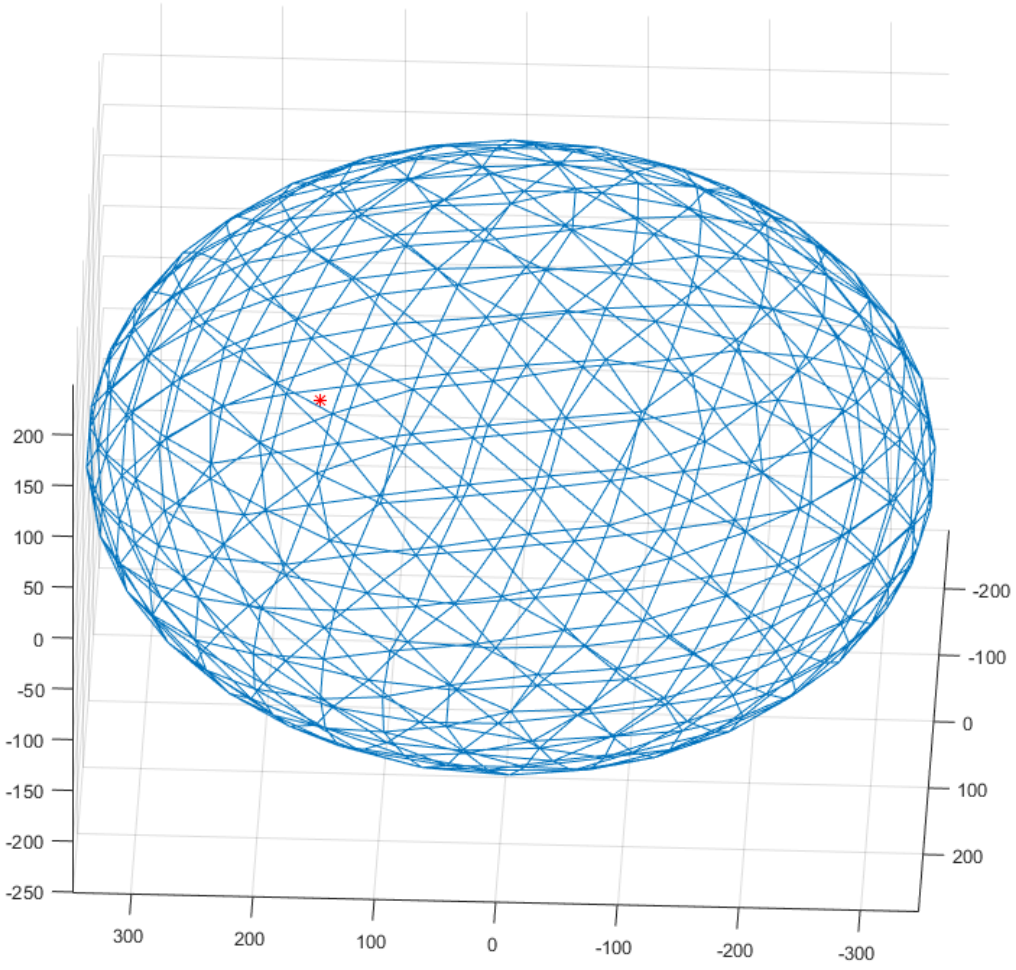


**Figure 1: Plot of the ellipsoidal asteroid with point mass inhomogeneity (red *)**

## 6.1 Tools

### 6.1.1 MathWorks MATLAB

MathWorks MATLAB is a software language and integrated development editor that provides a plethora of tools for engineers and scientists. The primary tools used in this paper are the numerical integrator function ODE45 and the plot function. The particle filters created in this paper were written in the MATLAB language.

## 6.2 Mathematical Model

Gravitational forces acting on the spacecraft were from two sources, the rotating homogenous bulk of the asteroid and the point mass inhomogeneity.

The equations of motion of the spacecraft hovering around the asteroid are computed using the equations of motion derived by Sawai et al[3] , modified to include the point mass inhomogeneity.

$$\ddot{\mathbf{r}} = \mathbf{a}_g + \mathbf{a}_{pm} - \mathbf{2}\begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}\dot{\mathbf{r}} - \begin{bmatrix} -\omega_y^2 - \omega_z^2 & \omega_x\omega_y & \omega_x\omega_z \\ \omega_x\omega_y & -\omega_x^2 - \omega_z^2 & \omega_y\omega_z \\ \omega_x\omega_z & \omega_y\omega_z & -\omega_x^2 - \omega_y^2 \end{bmatrix}\mathbf{r}$$

$$\omega_x = \omega_0 \sin(\theta)\cos(\omega_n t)$$

$$\omega_y = \omega_0 \sin(\theta)\sin(\omega_n t)$$

$$\omega_z = \omega_0 \cos(\theta)$$

$$\omega_n = \frac{I_{33} - I_{11}}{I_{11}}\omega_0 \cos(\theta)$$

$\mathbf{r}$ and $\dot{\mathbf{r}}$ are the spacecraft position and velocity in the rotating body frame respectively. $\boldsymbol{\omega_x}$, $\boldsymbol{\omega_y}$, and $\boldsymbol{\omega_z}$ represent the x, y, and z components of the ellipsoid's angular velocity. Note that in these simulations, $\dot{\mathbf{r}} = \mathbf{0}$ because the spacecraft is hovering (not moving) in the body frame. θ is the angle between the angular velocity vector and the Z-axis. For these simulations θ is initially set to

zero. When θ is set to zero, it can be seen that the X and Y components of the angular velocity are zero as well. Thus $\omega_z = \omega_o$ when $\theta = 0$.

$I_{nn}$ are the diagonal elements of the ellipsoid. The ellipsoid was assumed axially symmetric for simplicity. Note that the equations for the components of angular velocity assume that the ellipsoid has a homogenous density and a diagonal inertia matrix. With the addition of the point mass, the assumption of the diagonal inertia matrix no longer holds. It will be seen in the Section 6.3.1 that a non diagonal inertia matrix will not matter for the particle filter described in Section 6.3.1 because the particle filter is only looking at the system at one instant in time. For the particle filter derived in Section 6.3.2, however, this will matter as numerical integration of the asteroids rotational equations of motion will be necessary.

$\mathbf{a}_g$ represents the gravitational force due to the homogenous bulk of the asteroid. The gravitational force created by the homogenous bulk mass was computed using the MATLAB function *Ellipsoid_Gravity_field.m*, written by Dario Cersosimo. The function uses Ivory's method to compute the gravitational forces of a homogenous ellipsoid based on the semi-principle axes and the density of the ellipsoid. The components of the gravitational forces are in the rotating body frame.

The gravitational forces produced by the inhomogeneity are modeled as a point mass, which is described by the equation:

$$\mathbf{a}_{\mathbf{p}m} = \frac{-mG\left(\mathbf{r} - \mathbf{r}_{\mathbf{pm}}\right)}{\left\|\mathbf{r} - \mathbf{r}_{\mathbf{pm}}\right\|^{\mathbf{3}}}$$

Here, $\mathbf{a}_{pm}$ is acceleration due to the point mass, m is the mass of the point mass inhomogeneity, $\mathbf{r}$ is the position of the spacecraft in the rotating body frame, $\mathbf{r}_{\mathbf{pm}}$ is the location of the point mass

inhomogeneity in the rotating body frame, and G is the gravitational constant 6.67 x $10^{-11}$ $m^3/(kgs^2)$.

## 6.3 Measurement Techniques

As stated previously, the method used to measure the gravitational forces acting on the spacecraft is similar to that described in Gaudet/Furfaro. The spacecraft was set to hover above a point on the asteroid and the commanded thrust, which is assumed to be equal in magnitude but opposite direction to the gravitational force, was measured to provide the particle filter with measurements with which it can estimate the asteroid parameters and the position and mass of the point mass in homogeneity. This method was implemented two different ways, which are described in the sections below.

### 6.3.1 Several Measurements Simultaneously

Initially a very simple method was used to generate simulated measurements. In this method all measurements were taken at the same instant in time from different locations. This model would be equivalent to having several spacecraft hovering around the asteroid, each taking a single measurement at once. Essentially, the particle filter will be reconstructing the asteroid's dynamical system by using several snapshots taken from different locations at a single time.

The particle filter is first initialized by defining the state that will be considered truth. The initial guess is created by adding some bias to the true state. The maximum amount of bias for each parameter in the initial guess was set at 10% and 30% in this experiment.

The values to be estimated in the simulation are the X, Y and Z coordinates of the mass concentration in the body frame, the mass of the mass concentration, the density of the homogenous bulk of the ellipsoidal asteroid, and the initial magnitude of the angular velocity.

The true state is then inserted into our mathematical model of the gravitational field and the commanded thrust required to cancel out the gravitational forces is calculated. These simulated measurements are then corrupted by adding a small amount of bias and random noise to them. The maximum amount of noise and bias added to the measurements can be adjusted. In the experiment the bias was set at 0%, 1% and 3%. The random noise was limited to $1/100^{th}$ the magnitude of the measurement.

The particle filter is then initialized by scattering particles around the initial guess. Each particle represents a possible state for the system. Each particle is then inserted into the system model, where the commanded thrust required to cancel out the gravitational forces from that particle is computed. Once computed, the commanded thrust is compared to the corrupted measurements generated from the true state.

Based on the difference between the commanded thrust computed from each of the particles' states and the commanded thrust computed from the true state, each particle is assigned a weight. The particles are then resampled by sampling from the set of particles with probability equal to their assigned weight. Particles with higher weights were selected more frequently in the new set. Each particle was then perturbed by a small amount to ensure no two particles were exactly the same.

The process was repeated until the standard deviation dropped below the prescribed amount (10% of the initial value).

Note that because the particle filter described above is only estimating the system from a single instant in time, all the parameters are constant from the particle filter's point of view. Therefore, it is not necessary to consider the effects of the point mass concentration on the angular velocity vectors in this simulation. This simulation was run at time $t = 0$ with $\theta$ initially set to zero. The

particle filter then estimated the angular velocity $\omega_0 = \omega_z$. The X and Y components of angular velocity were known to be zero and were not part of the estimation process.

Figure 2 illustrates a functional diagram of the filter used for this section.

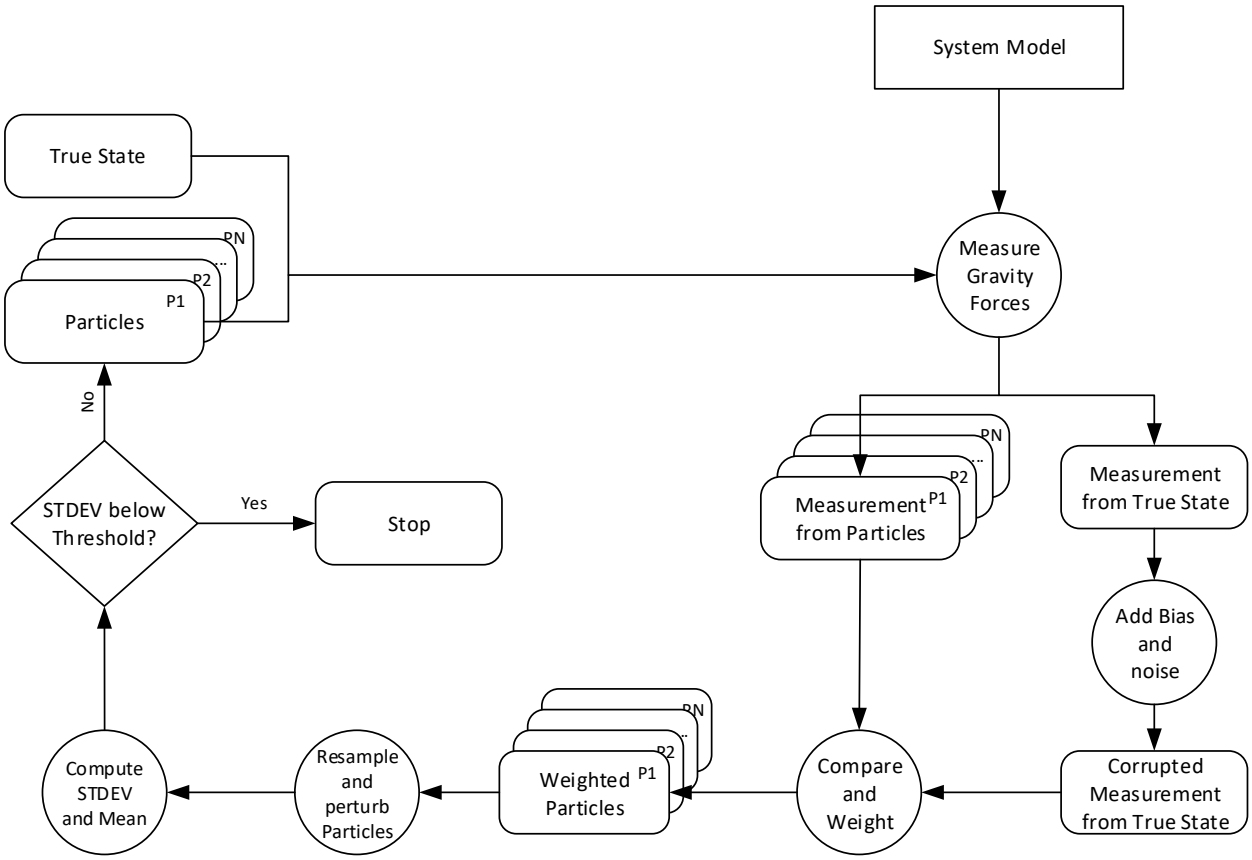


**Figure 2: Functional diagram of the particle filter**

The parameters to be estimated in this simulation are the homogenous density of the asteroid, the magnitude of angular velocity, the mass of the mass concentration, and the X, Y, and Z coordinates of the mass concentration. Table 2 describes values of true state to be estimated in this simulation.

| Parameter | Value | Description |
|---|---|---|
| **density** | 7870 | Density of the homogenous bulk of ellipsoid in $kg/m^{3.}$ |
| **$\omega_0$** | $4.06 \times 10^{-5}$ | Initial magnitude of ellipsoid angular momentum relative to inertial frame in rad/s (aligned with the Z-axis) |
| **m** | $8.15 \times 10^{8}$ | Mass of point mass inhomogeneity in kg. |
| **r** | [150, 150, 150] | Position of point mass inhomogeneity in the body fixed frame in meters. |

**Table 2: Description of true state.**

#### 6.3.1.1 Positioning of Measurements

The locations where the simulated measurements were taken proved to be crucial to locating the inhomogeneity. The simulated measurements had to be unique enough such that the particle filter could not confuse the true position and mass of the inhomogeneity for a point mass that was of similar size and location. If the simulated measurements were not unique enough such that the particle filter could not rule out the force generated by one point mass over another, the particle filter would likely diverge or, in some cases, not be able to converge to a high degree of certainty.

The problem is intuitive to understand. Consider a point mass that is 50 kg and a distance of 10 meters away from an observer. Then consider a mass that is 40 kg but is closer to the observer such that the gravitational force exerted by the 40kg mass on the observer is the same as the 50 kg mass. Such a scenario would make it impossible to determine which point mass could be the true state. To improve the particle filter's convergence, multiple measurements need to be taken at various locations such that it is possible to hone in on the location and mass of the inhomogeneity.

An initial design of where to take measurements was created by drawing shapes in the 3D modeling program Blender and using the vertices from the shape as the locations where the measurements would be taken. This initial design was far from optimal and had 174 measurement locations. This drastically slowed down the time to run the particle filter. However, it did help generate a good initial guess as to where the measurements should be taken.

Once the initial placement was chosen, the number of simulated measurements was reduced to 25 by removing vertices from the 3D shape, greatly improving speed.

### 6.3.2 Several Measurements Created over Time

This method is more similar to a realistic space mission around an asteroid. In this method each measurement is taken from one location at one time. The spacecraft then moves to a different location at a different time to take another measurement. This process is repeated as many times as is necessary to reconstruct the dynamical system.

The particle filter starts by initializing the particles around the initial guess. The initial guess is created by adding some amount of bias to the true state. The maximum amount of bias for each parameter in the initial guess was set at 10% and 30% in this experiment.

The values to be estimated in the simulation are the X, Y and Z coordinates of the mass concentration in the body frame, the mass of the mass concentration, the density of the homogenous bulk of the ellipsoidal asteroid, and the X, Y and Z components of the asteroid's angular velocity relative to the inertial frame in body frame coordinates.

Once initialized, the true state and the particles are propagated through the numerical integrator *ode45.m* using the asteroids rotational equations of motion. Note that the only parameters that change with time are the components of the asteroid's angular velocity. The density does not change with time. The mass of the mass concentration does not change with time, and the location of the mass concentration is fixed with the rotating body. Thus when gravity force measurements are taken in the body-fixed frame, the coordinates of the mass concentration will remain unchanged.

Once the particles and the true state are propagated, the commanded thrust required to cancel the gravitational forces for each particle is measured using the system dynamics model. In this approach, a single measurement was created for each particle from a single location. The

measurement from the true state are corrupted with random noise and bias. The maximum amount of noise and bias added to the measurements can be adjusted. In the experiment the bias was set at 0%, 1% and 3%. The random noise was limited to $1/100^{th}$ the magnitude of the measurement. The measurements created from the particles is then compared against corrupted measurements from the true state. Based on the difference between the measurements created from the particles and the corrupted measurements from the true state, a weight was assigned to each particle.

Once each particle was assigned a weight, the particles were resampled by sampling from the set of particles with probability equal to their weight. Particles with higher weights had a better chance of beings elected.

The estimate for the true state was computed by taking the mean of the particles. The standard deviation was then computed and used as a metric to determine if the particle filter had converged. If the standard deviation had dropped below the prescribed threshold (10% of its initial value) the particle filter was deemed to have converged. If the standard deviation had not fallen below that threshold, the resample particles were then run through the system again until the standard deviation had dropped below the threshold. Each time the particles are resampled, they are perturbed by a small amount to ensure that there are no particles that are exactly the same. Also, each measurement is taken from one of 25 selected locations at a given time. These locations are the same as the locations used in the previous method. These locations were repeatedly used until a solution could be found or the particle filter was confirmed to have diverged.

This method introduced another level of complexity that did not have to be considered in the previous technique nor in the particle filter created by Dr. Furfaro and B. Gaudet. The addition of the point mass causes the changes to the moment of inertia of the asteroid in such a way that it causes the asteroid to wobble. This meant all three components of the asteroid's angular velocity

changed over time. Unfortunately, the change in angular velocity could not be parameterized in a single parameter in the state vector as it was in the particle filter created by Dr. Furfaro and B. Gaudet. To parameterize the angular velocity in the state vector would require knowledge of the asteroid's moment of inertia, which would require knowledge of the location and mass of the mass concentration, which would defeat the purpose of this project. Thus, it is necessary to add the three components of the angular velocity to the state vector and estimate them in time.

Figure 3 functional diagram of the particle filter used in this section. Note the addition of the *ode45.m* block which was not necessary in the previous method because the previous method only considered the dynamical system at one instant in time.

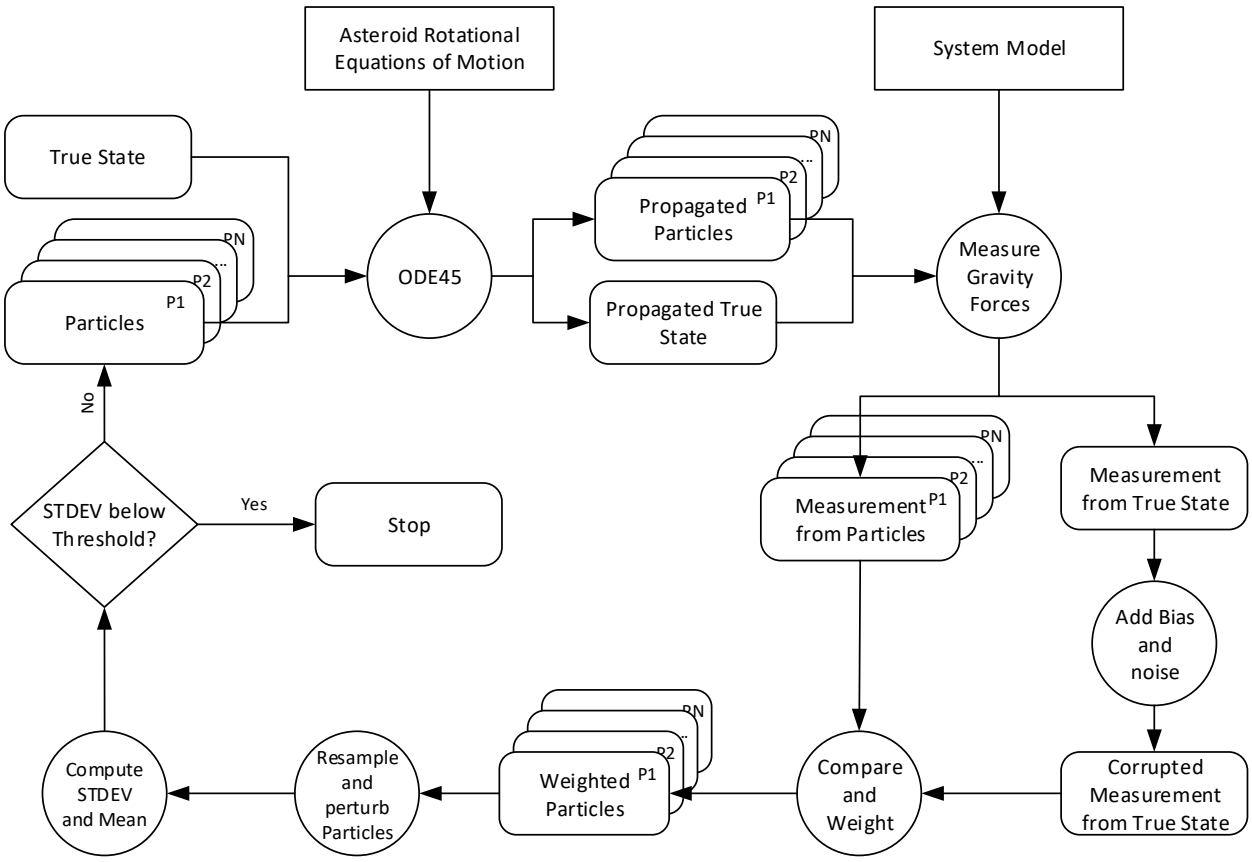


**Figure 3: Functional diagram of the particle filter.**

The parameters to be estimated in this simulation are the homogenous density of the asteroid, the X, Y, and Z components of angular velocity, the mass of the mass concentration, and the X, Y, and Z coordinates of the mass concentration. The table below describes values of true state to be estimated. Note the angular velocity has the same magnitude but is now aligned in the direction [1, 1, 1]. This is to alleviate problems with dividing by zero when the total error is calculated.

| Parameter | Value | Description |
|---|---|---|
| **density** | 7870 | density of the homogenous bulk of ellipsoid in $kg/m^{3.}$ |
| $\omega_0$ | $[2.34, 2.34, 2.34] \times 10^{-5}$ | Initial magnitude of ellipsoid angular momentum relative to inertial frame in rad/s. |
| **m** | $8.15 \times 10^{8}$ | Mass of point mass inhomogeneity in kg. |
| **r** | [150, 150, 150] | Position of point mass inhomogeneity in the body fixed frame in meters. |

**Table 3: Description of True State**

#### 6.3.2.1 Integrating the Asteroids Rotational Equations of Motion

Integration of the asteroid's rotational equations of motion was necessary to propagate the components of angular velocity. This was done using MATLAB's ODE45 integrator.

The asteroid was set up to rotate about the center of mass of the homogenous bulk. Without the inclusion of the inhomogeneity, this center of mass was taken to be at the body frame origin. By including the mass of the inhomogeneity, the center of mass moved slightly but only by 26 cm, which was considered to be negligible given the size of the asteroid and the distance of the spacecraft.

The moment of inertia of the bulk homogenous mass was computed about the center of mass by using the following equations for a the moment of inertia for an ellipsoidal object:

The moment of inertia of the point mass inhomogeneity was calculated about body-frame origin using Equation 4.14 of Shaub/Junkins[2] .

$$\mathbf{I} = m_{pm}\begin{bmatrix} r_y^2 + r_z^2 & -r_x r_y & -r_x r_z \\ -r_x r_y & r_x^2 + r_z^2 & -r_y r_z \\ -r_x r_z & -r_y r_z & r_x^2 + r_y^2 \end{bmatrix}$$

These two were summed together to give the total moment of inertia for the asteroid with a mass concentration.

The rotation of the asteroid was simulated in ODE45 by integrating the asteroid's rotational equations of motion and integrating the kinematic differential equation for quaternions. The asteroid's rotational equations of motion are derived in Shuab/Junkins[2] Section 4.2.3. The final equation is given by Equation 4.32 of Shaub/Junkins[2]:

$$\mathbf{I}\dot{\boldsymbol{\omega}} = -\tilde{\boldsymbol{\omega}}\mathbf{I}\boldsymbol{\omega} + \mathbf{L}_c$$

It was assumed here that there are no external torques acting on the asteroid and that solar radiation pressure is negligible, $\mathbf{L}_c = \mathbf{0}$. $\tilde{\boldsymbol{\omega}}$ is defined as:

$$\tilde{\boldsymbol{\omega}} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}$$

Figure 4 shows a plot of the simulated asteroid's angular velocity and energy over 25,000 seconds. Initially the angular velocity is aligned with the Z-axis, as can be seen from the graph. The components of angular velocity, $\omega_1$, $\omega_2$, and $\omega_3$, change in time demonstrating that the asteroid wobbles due to contribution of the mass concentration. The energy, however, remains constant. This is to be expected since there are no external torques acting on the asteroid.

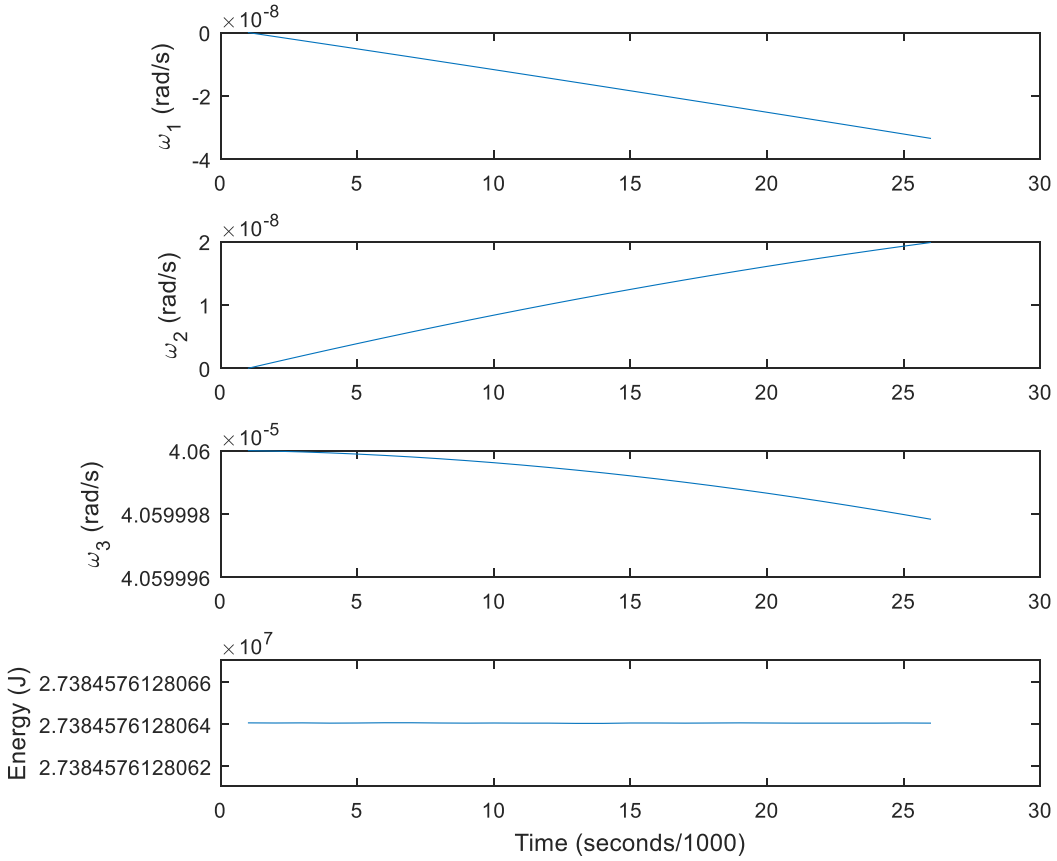


**Figure 4: Plot of components of angular velocity and energy over time**

Figure 5 shows what the simulation would be like if the point mass inhomogeneity is neglected. This is to ensure the simulation is working properly. Note the energy is lower here because there is less mass in this simulation rotating around the axis.

The angular velocity remains constant as the asteroid is rotating around the Z-axis. The angular velocity remains constant because the object is rotating about one of its principle axes, the moment of inertia matrix is diagonal in the body frame, and there is no energy transfer.

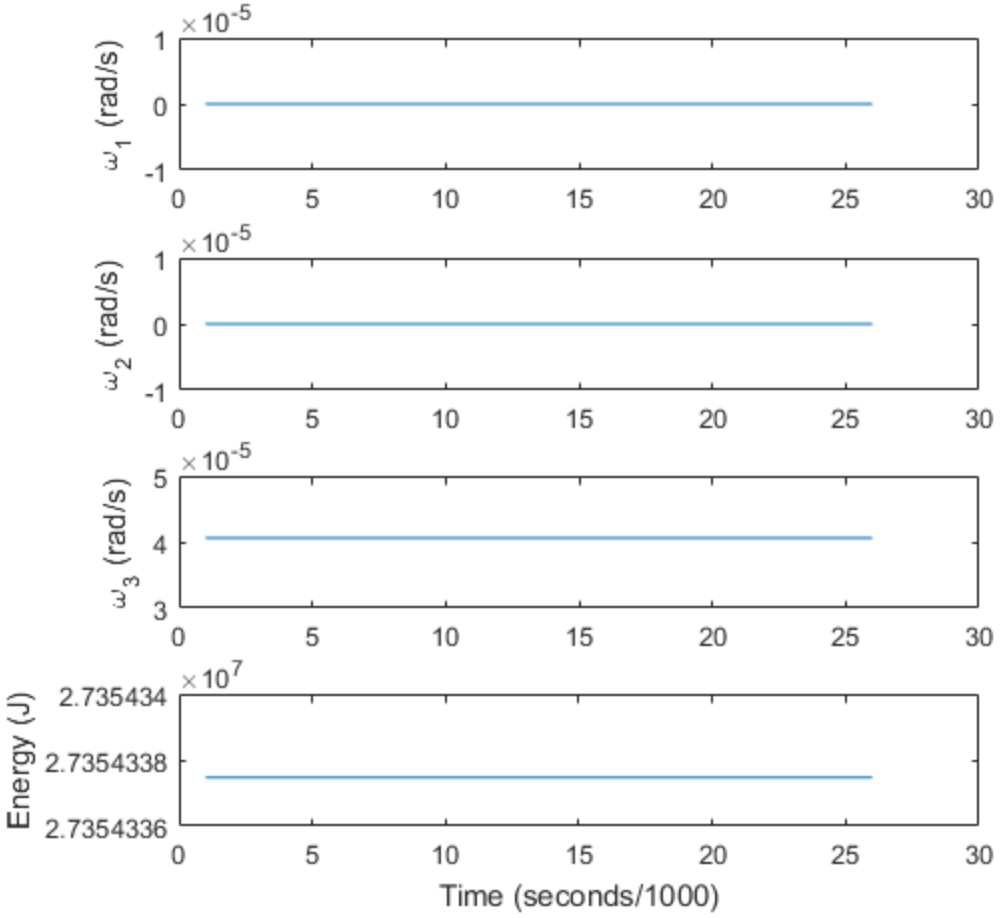


**Figure 5: Comparison of the rotating asteroid simulation without the point mass concentration**

From the comparison of each simulation, it appears the simulation is working properly. The results are as expected.

# 7 Results

## 7.1 Several Measurements Simultaneously

The initial run of the particle filter used the first measurement technique, where multiple measurements were taken at a single time. This approach can be thought of as taking a snapshot of the asteroid from multiple positions all at once and rebuilding the dynamical system based on these snapshots. 10,000 particles were used and 25 measurements were taken simultaneously around the asteroid.

The values of the true state are as follows:

| Parameter | Value | Description |
|---|---|---|
| **density** | 7870 | density of the homogenous bulk of ellipsoid in $kg/m^{3.}$ |
| **$\omega_0$** | $4.06 \times 10^{-5}$ | Initial magnitude of ellipsoid angular momentum relative to inertial frame in rad/s (aligned with the Z-axis) |
| **m** | $8.15 \times 10^{8}$ | Mass of point mass inhomogeneity in kg. |
| **r** | [150, 150, 150] | Position of point mass inhomogeneity in the body fixed frame in meters. |

The bias of the initial guess and the measurement bias were varied for each run. Table 4 provides a summary of the results.

| Run | Initial bias (max) | Measurement Bias | Initial Error | Final Error | Initial STDEV | Final STDEV | No. Iterations | Converged? |
|---|---|---|---|---|---|---|---|---|
| **1** | 30% | 3% | 26.48% | 6.68% | 86.9e6 | 8.73e6 | 42 | Yes |
| **2** | 10% | 3% | 11.56% | 4.05% | 85.1e6 | 8.80e6 | 63 | Yes |
| **3** | 30% | 1% | 32.78% | 6.74% | 82.1e6 | 8.34e6 | 85 | Yes |
| **4** | 10% | 1% | 14.57% | 4.26% | 79.5e6 | 8.69e6 | 280 | Yes |
| **5** | 30% | 0% | 24.64% | 6.28% | 72.5e6 | 7.11e6 | 44 | Yes |
| **6** | 10% | 0% | 10.96% | 5.79% | 82.6.8e6 | 8.54e6 | 61 | Yes |

**Table 4: Summary of Results**

As a side note, it can be seen that the final standard deviation is not exactly 10% of the initial standard deviation. This is a programming error that had no significant impact on the simulation. The initial standard deviation in the table is actually the standard deviation of the particles after they have been filtered through the first iteration. The actual initial standard deviation, which would be the standard deviation from the zeroth iteration, was stored in a variable, but was not

recorded. Thus, the final standard deviation is slightly larger than 10% of the initial standard deviation presented in the table.

Figure 6, Figure 7, Figure 8, and Figure 10 show each of the estimated parameters plotted along an axis for an initial bias of 30% and a measurement bias of 3%. The red diamond in the center of the plots is the true value for that state. The blue asterisk * is the estimated state, which is calculated as the mean of all of the values for that parameter. The nebulous cloud of points around the mean state is a plot of every individual particle. These points are color coded corresponding to their weight to give an idea of how the particle filter is weighting points. The colors are as follows:

| Color | Green | Blue | Magenta | Yellow | Red | Black |
|---|---|---|---|---|---|---|
| **% of Max Weight** | 90 | 80 | 70 | 60 | 50 | < 50 |

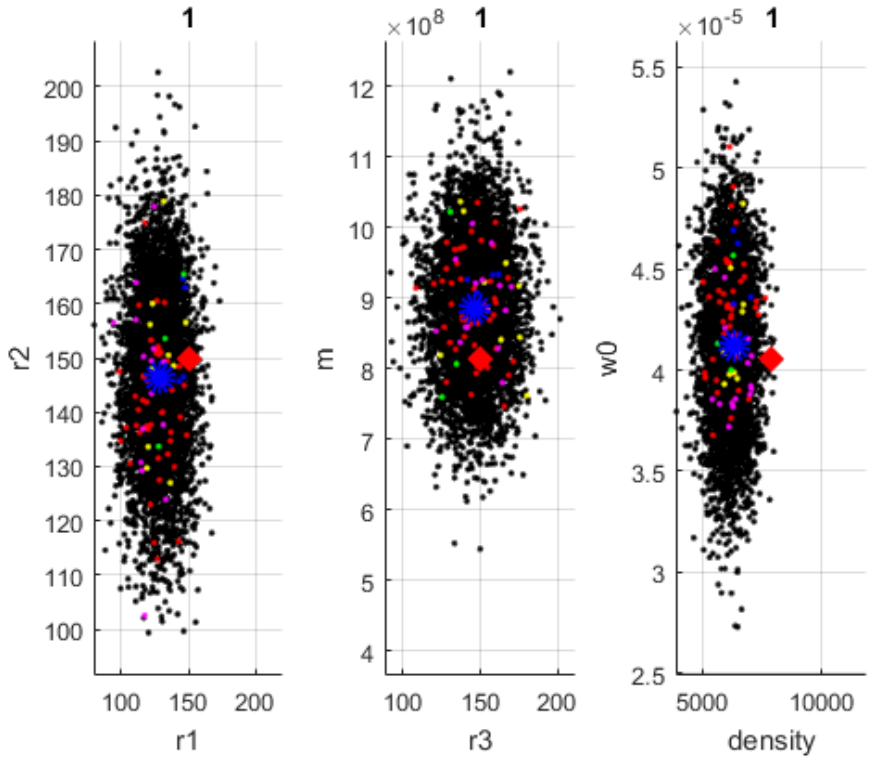


**Figure 6: Plot of all particles, estimated state, and true value. Iteration 1.**

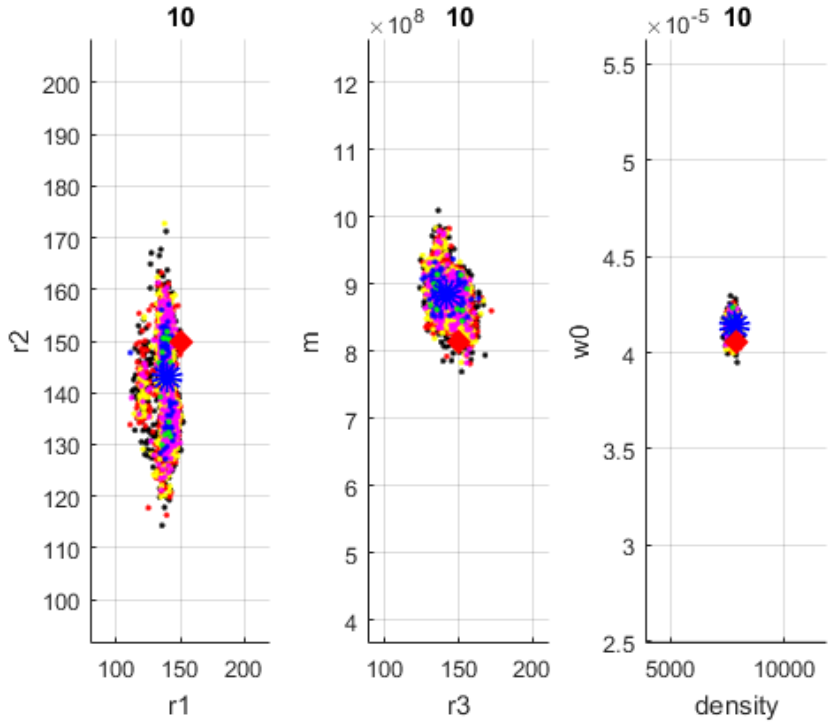


**Figure 7: Plot of all particles, the estimated state, and true values. Iteration 10.**

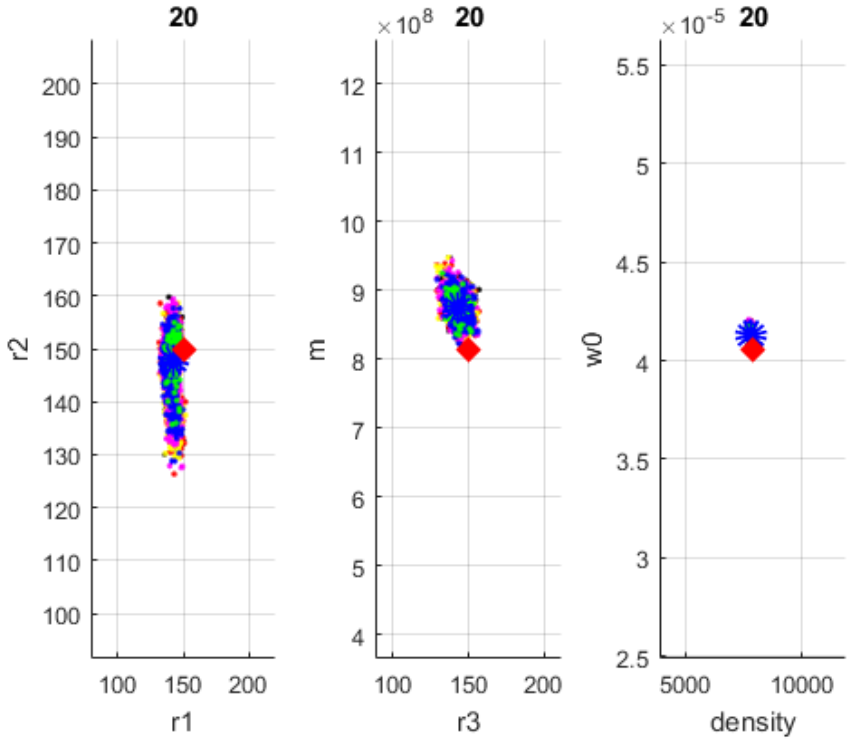


**Figure 8: Plot of all particles, the estimated state, and true values. Iteration 20**

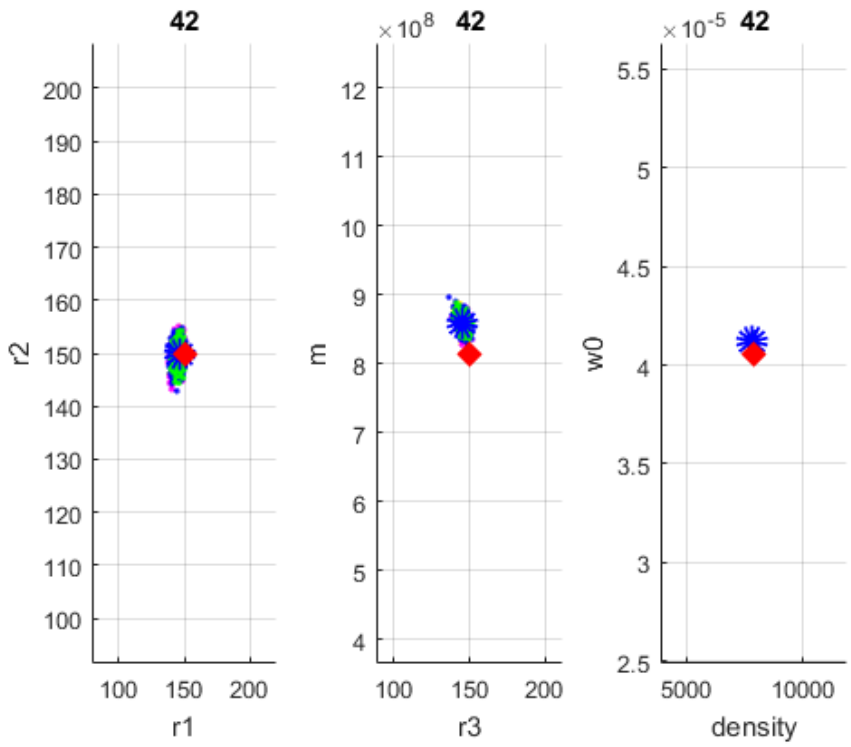


**Figure 9: Plot of all particles, the estimated state, and true values. Iteration 60.**

In iteration 1 (Figure 6), the particle filter has been initialized to a highly incorrect state. The nebulous cloud of particles surrounding the estimated state was quite large, indicating the estimate is not very certain. Many of the particles in the first iteration are black, indicating that many do no have very much weight

In iteration 10 (Figure 7), the particle filter began to hone in on the true values. Clusters of highly weighted (green and blue) can be seen forming near true state.

In iteration 20 (Figure 8), the particle filter rapidly converged around the correct solution for the angular velocity and the density. Few black particles can be seen, and a cluster of green and blue particles formed around the true state.

In iteration 42 (Figure 9), the particle filter has completely converged. The standard deviation is below 10% its original value. The cloud of particles has become much smaller as the particle filter

closed in on the true value. Note, however, that the particle filter was not able to discern the inhomogeneity parameters with as high a degree of certainty as it has with the bulk density and angular velocity terms. This is explained in the discussion section.

Figure 10 shows the percent error for each of the parameters and the total error of the entire state. Initially the error is 26.48%, but it is reduced to 6.68% in the final iteration. The red line on top represents the absolute value of all errors summed together. Note that individual parameter errors can be negative as the absolute value of the error was not taken for individual parameters. A negative value represents an undershoot of the estimated parameter.

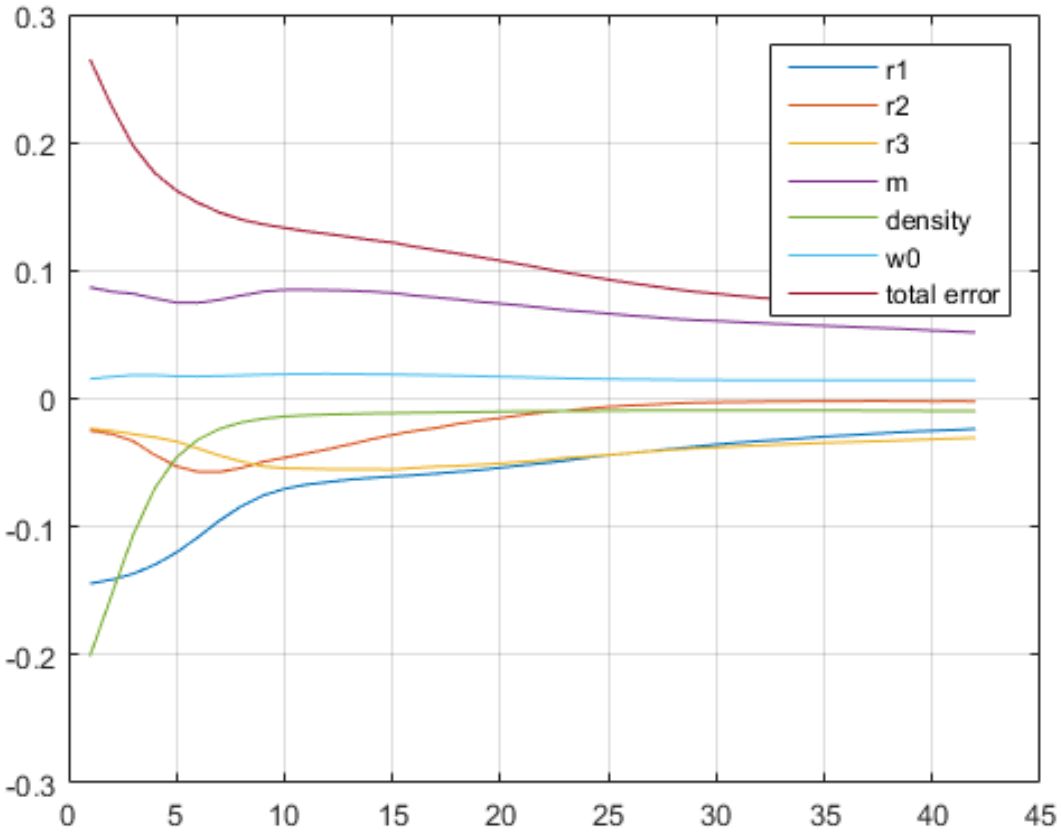


**Figure 10: Plot of the % state error over time**

## 7.2 Several Measurements Taken Over Time

In this run, the particle filter took measurements from different locations and different times. This approach can be thought of as a single spacecraft hovering around the asteroid at given positions and times taking measurements. 10,000 particles were used.

The values of the true state is as follows:

| Parameter | Value | Description |
|---|---|---|
| density | 7870 | density of the homogenous bulk of ellipsoid in $kg/m^{3.}$ |
| $\omega_0$ | $[2.34, 2.34, 2.34] \times 10^{-5}$ | Initial angular velocity of ellipsoid relative to inertial frame in rad/s. |
| m | $8.15 \times 10^{8}$ | Mass of point mass inhomogeneity in kg. |
| r | [150, 150, 150] | Position of point mass inhomogeneity in the body fixed frame in meters. |

Table 5 summarizes the results for each value of initial and measurement bias.

| Run | Initial bias (max) | Measurement Bias | Initial Error | Final error | Initial STDEV | Final STDEV | No. Iterations | Converged? |
|---|---|---|---|---|---|---|---|---|
| **1** | 30% | 3% | 49.0% | 154% | 94.3e6 | 454e6 | 300 | No |
| **2** | 10% | 3% | 20.0 | 126% | 75.4e6 | 284e6 | 300 | No |
| **3** | 30% | 1% | 31.5% | 141% | 89.3e6 | 411e6 | 300 | No |
| **4** | 10% | 1% | 16.8% | 127% | 87.5e6 | 321e6 | 300 | No |
| **5** | 30% | 0% | 48.0% | 138% | 98.4e6 | 393e6 | 300 | No |
| **6** | 10% | 0% | 19.7% | 170% | 88.0e6 | 353e6 | 300 | No |

**Table 5: Summary of results**

Figure 11, Figure 12, and Figure 13 show a plot of the estimated state, the true state and each of the particles for run 2, with an initial bias of 10% and a measurement bias of 3%. The colors and symbols are the same as in the previous simulation.

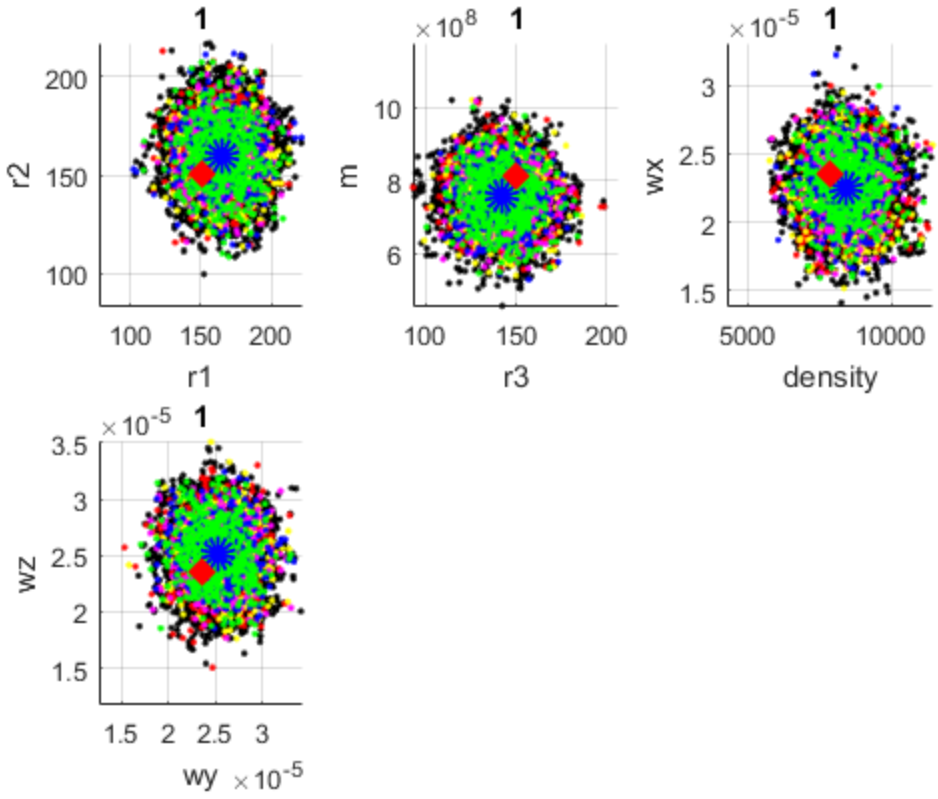


**Figure 11: Plot of all particles, the estimated state, and true values. Iteration 1.**

From Figure 11, it can be seen that there is a large green cluster around the estimated state in the first iteration. This is typically undesirable as it indicates that the particle filter is not able to hone in on the correct solution.

In Figure 12, the cluster of particles has not begun converging around the correct state. Many of the particles around the estimated state have a low probability, and there are no clear clusters of high probability areas.

Finally, in Figure 13, it is clear that the particle filter is not able to find the actual state. The plot of the X, Y, and Z components of angular velocity show the particle has clearly diverged.

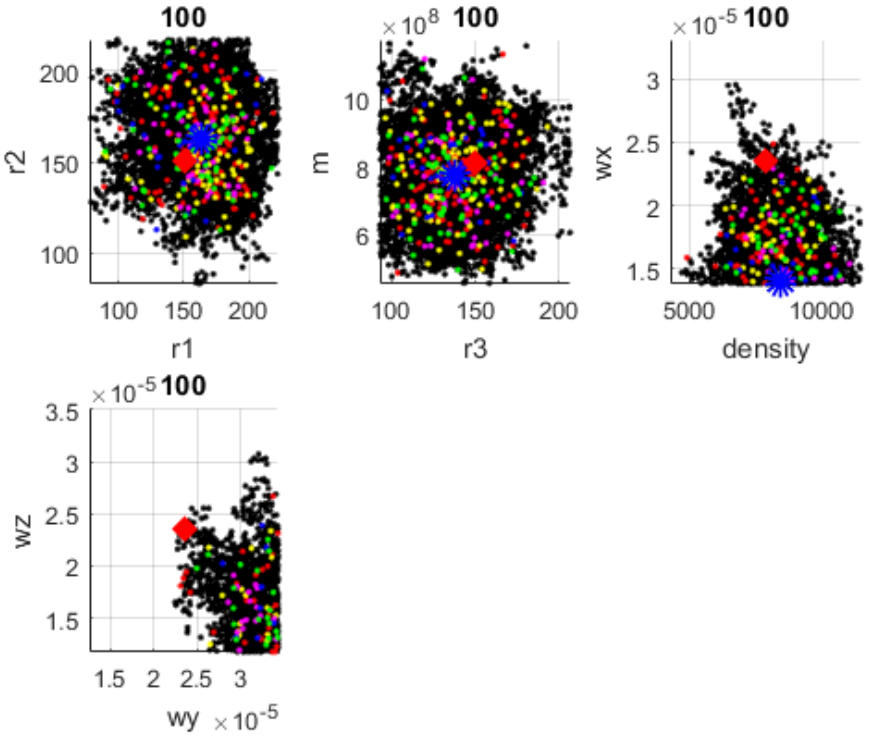


**Figure 12: Plot of all particles, the estimated state, and true values. Iteration 100.**

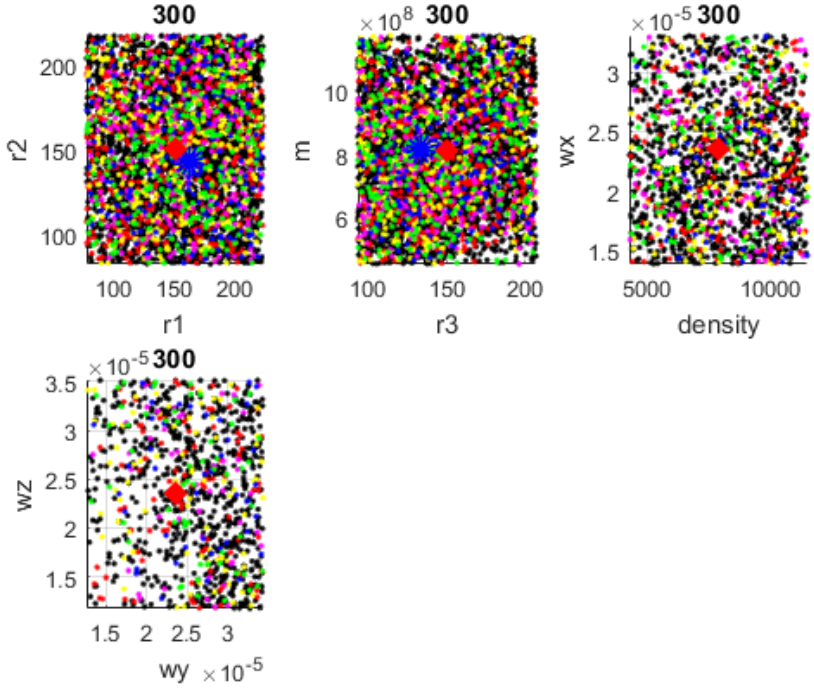


**Figure 13: Plot of all particles, the estimated state, and true values. Iteration 300.**

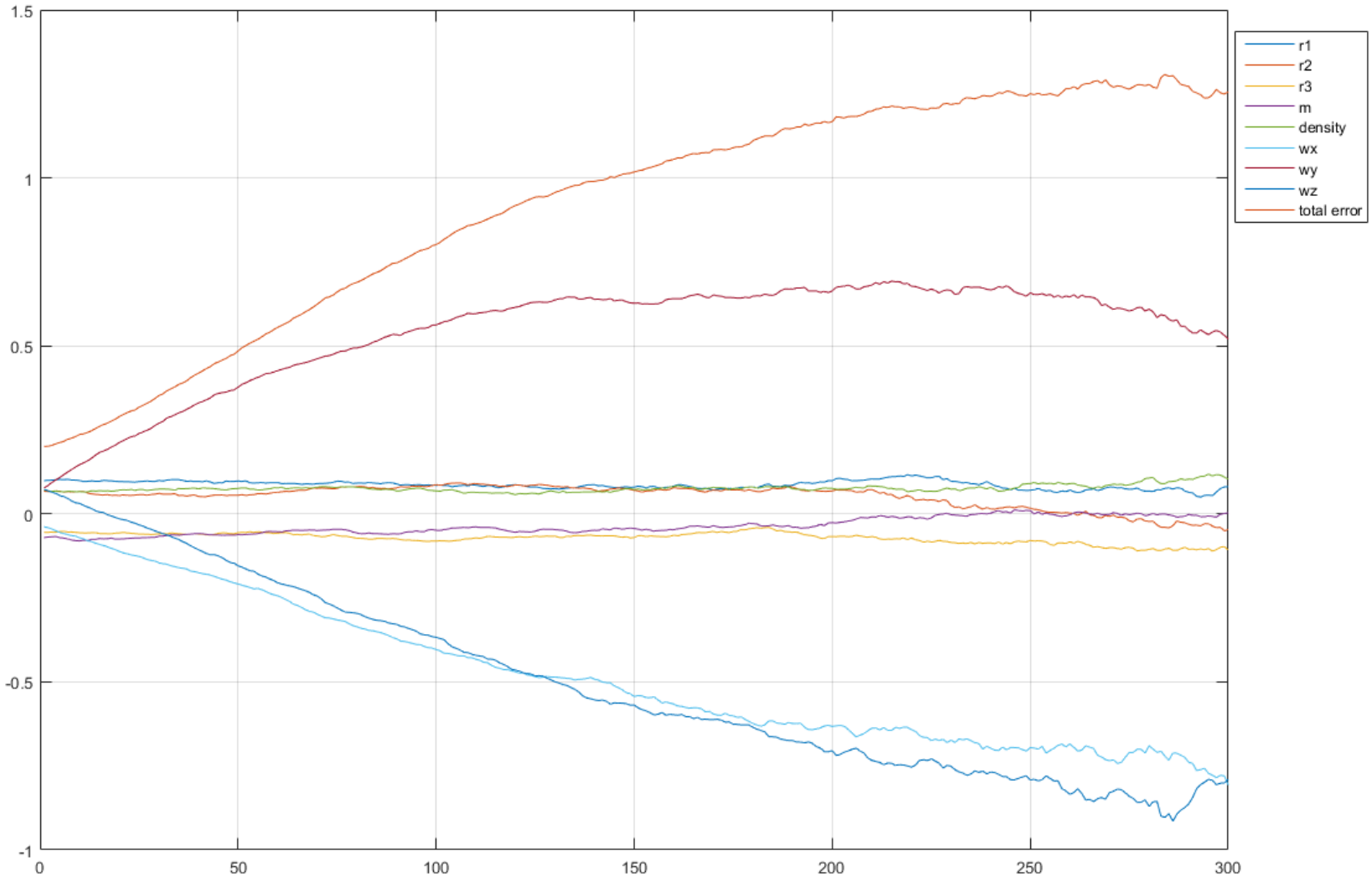


**Figure 14: Plot of the state error against the number iterations for run 2**

This approach was far less than successful than the previous approach. It can be seen from Figure 14 that the components of angular velocity were the main cause for the particle filters failure. This failure will be discussed in Section 8.2.

# 8 Discussion of Results

This section discusses the results in Section 7. It also discusses improvements that could be made to the work presented here.

## 8.1 Several Measurements Simultaneously

This section discusses the results from Section 6.1.

From the summary of results presented in Table 4, the particle filter performed remarkably well. Each run of the particle filter converged to a solution, and all solutions were under 10% error. Thus it can be concluded that the particle filter method can provide an accurate estimate for the location and mass of a single point mass concentration.

An interesting note is that the final error did not appear to be affected by the measurement bias. Intuitively, a larger measurement bias would cause the particle filter to weight particles that are further from the true state more heavily than particles that are closer to the true state. Thus, one would expect the particle filter to become more accurate as the measurement bias decreases because it is weighting particles that are closer to the true state more heavily.

The fact that the error remains essentially the same despite the measurement bias decreasing indicates that there may be a problem with how the weights are assigned in general. The particle filter assigns individual weights based on the difference between the measurement taken from the true state and the measurement taken from the individual particle. The weight is then assigned from a logarithmic normal distribution with mean of 0 and variance equal to the variance of the differences of all the measurements compared to the true state. The logarithmic normal distribution was chosen because it gave the better results compared to a normal distribution. When a normal distribution was used, much of the weight was concentrated in a few particles. When the particles were resampled only a few particles were selected This caused the particle filter to produce bad

results. A logarithmic normal distribution was also chosen because it is what Dr. Furfaro and B. Gaudet used in their particle filter. In future experiments, it would be beneficial to compare a multitude of distributions to see which one yields better results.

It is important to talk about the performance of the particle filter beyond just the final error. This includes a discussion of the speed of convergence and how the particle filter converged. Ideally the particle filter will converge as quickly as possible. Furthermore, it is ideal that the particle filter converge asymptotically to a final error and never increase.

From the summary in Table 4, it can be seen that a lower initial bias did not result in a faster convergence. A lower initial bias will result in an initial estimate that is closer to the true state. Therefore intuitively, one would expect a particle filter given an initial state to converge quicker around the solution. However, this was not the case. The reason behind this can be understood by comparing two plots with different initial biases. Figure 15 is a plot of run 3, which had an initial bias of 30% and a measurement bias of 1%. Figure 16 is a plot of run 4, which had an initial bias of 10% and a measurement bias of 1%.

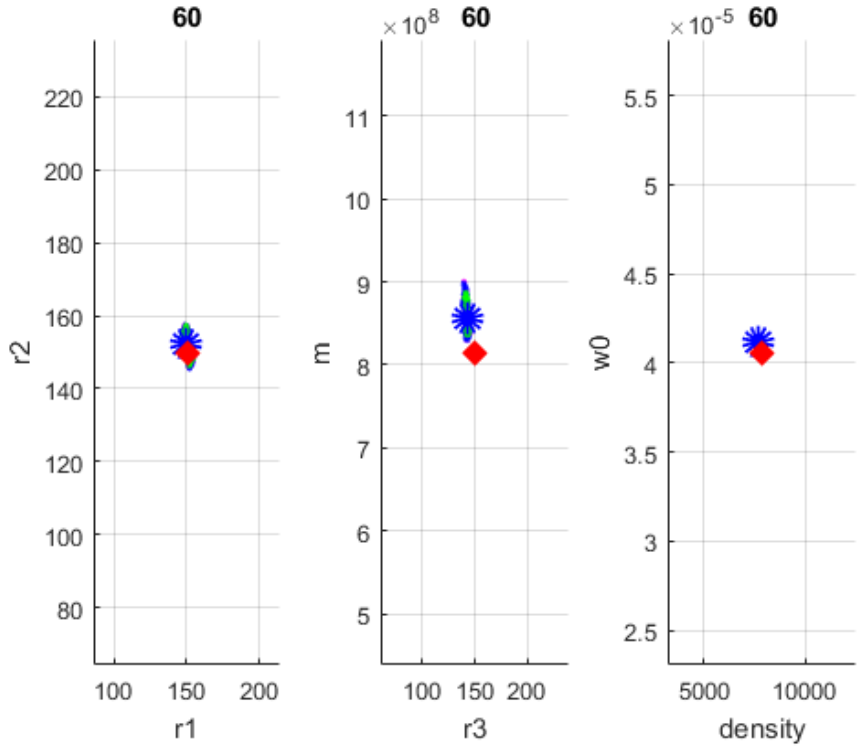


**Figure 15: Plot of run 3 at Iteration 60**

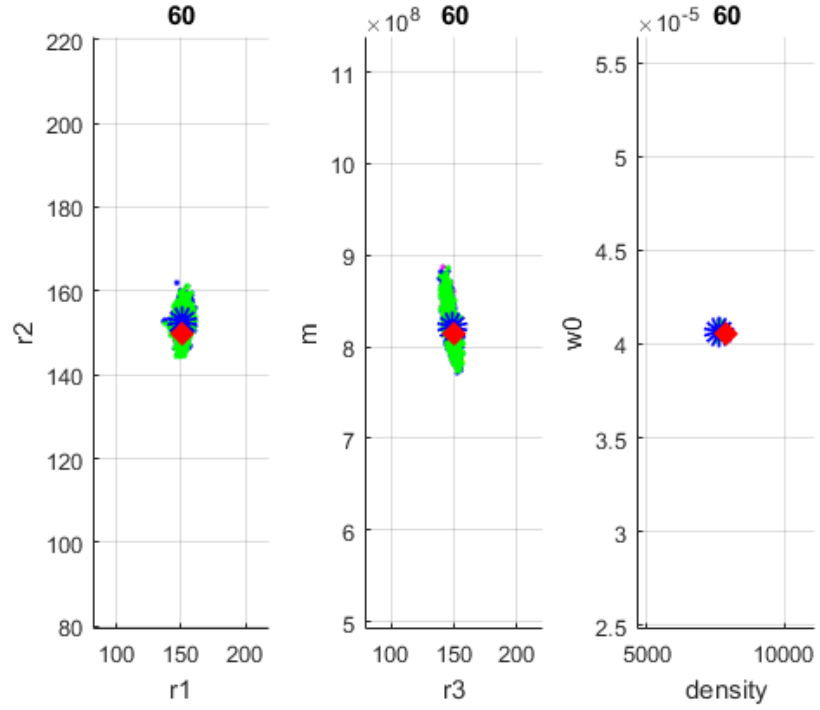


**Figure 16: Plot of run 4. Iteration 60**

It can be seen from the two figures that by iteration 60, run 3 had a much smaller spread of particles around the estimated state, whereas the spread of particles in run 4 was much larger. The effect of the initial state being further away from the true state in run 3 was that there were fewer particles that had higher weights. Thus when the particle filter resampled the particles, it was able to quickly weed out wrong answers early on in the first few iterations of the particle filter, leaving behind only a few particles which had a higher weight. Since only a few particles remained after resampling, the standard deviation of the particles dropped very quickly, which cause the particle filter to converge quicker. When several particles were very close to the true state, as in the case of run 4 with only 10% initial bias, the weight was not concentrated in a few particles. Rather several particles had a very high weight, which meant these particles were more likely to be in the new set when the particle filter resampled. Thus the standard deviation did not drop as quickly, and the particle filter required several more iterations to converge.

Another effect of the initial bias is that when the weight is not concentrated in a few particles, the estimated state is allowed to wander. As discussed before, when the initial bias is higher more particles are weeded out in the early iterations of the particle filter. When the initial bias is low, more particles have a higher weight. When several particles have a higher weight, the particle filter will tend to think that potentially less accurate solutions are correct. This allows the particle filter to wander around the area of high weighted particles, sometimes away from the true state. Figure 17 shows the error over time for run 4, which had an initial bias of 10%. Notice how the total error drops to just slightly over 3% then rises back up to just over 4%. The problem is that when all of the particles have equal amounts of weight, the particle filter cannot discern which state is more correct than another.

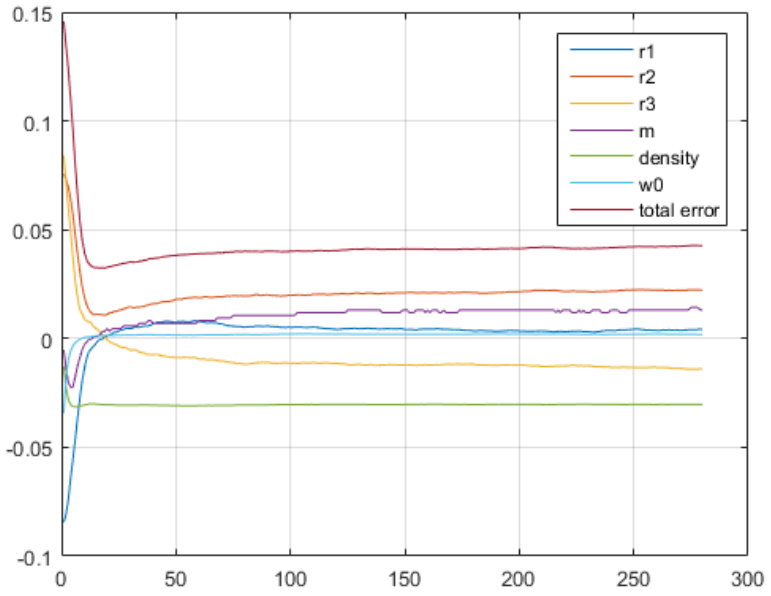


**Figure 17: Plot of % error for run 4**

For comparison, the plot of run 3, which had an initial bias of 30% is presented in Figure 18. Notice how the total error converges almost asymptotically and never rises.

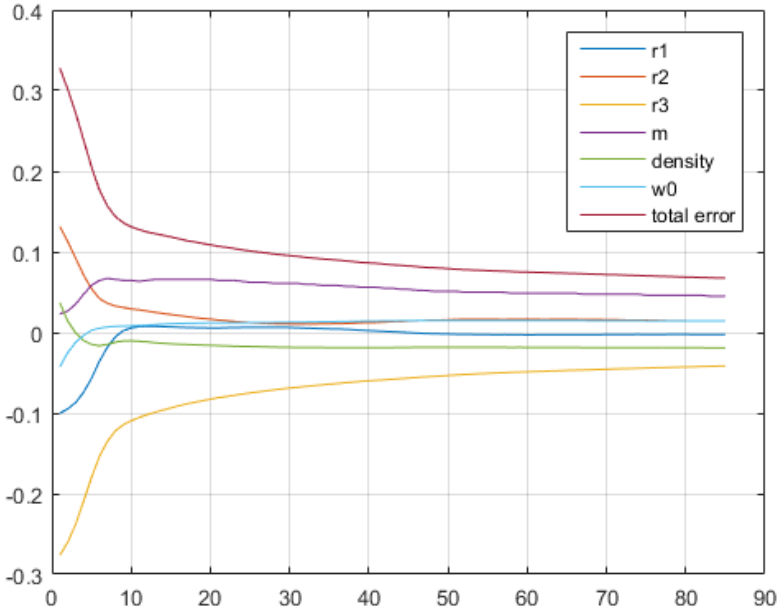


**Figure 18: Plot of % error for run 3**

## 8.2 Several Measurements Taken Over Time

This section discusses the results from Section 7.2.

It is clear from the results in Table 5 that the particle filter was not able to accurately determine the parameters of the dynamical system. Each run failed to converge, and the final error was much higher than the initial error.

By examining Figure 14, it is obvious that the components of the angular velocity vector were the primary reason the particle filter failed. The components of angular velocity diverged significantly, more than any of the other parameters. Figure 13 shows that the spread of particles in the last two plots, which contain the components of angular velocity, grew considerably. Note that the spread of particles in the last two plots is much larger than in any of the other plots. This indicates that the estimate was extremely uncertain, and the particle filter could not discern which state could be correct.

The particle filter failed in this simulation because it was not able to determine the components of angular velocity. The most likely reason for this is that the measurements taken were not unique enough to allow the particle filter to weed out particles that were more incorrect than others. This is similar to the problem that occurred in the first method described in Section 6.3.1.1. If multiple states could potentially provide the same measurement, then there is no way to determine which state is more correct.

## 8.3 Improvements to the Current Work

This project was very ambitious. Estimating several parameters simultaneously using only measurements from the gravitational forces is an extremely difficult task. To improve the estimate of the asteroid's dynamical state, it would be beneficial to estimate certain parameters separately from one another.

In practice, it is not necessary to estimate all of these parameters simultaneously from one type of measurement. By estimating parameters separately using different techniques, a more accurate estimation of the asteroid's dynamical state can be made. One suggestion is to estimate the angular velocity using optical techniques. It is possible, and much easier, to estimate angular velocity by taking multiple images of the asteroid and use an image processing technique such as stereophotoclinometry. This technique was utilized to determine the rotational state of the asteroid 21 Lutetia during a flyby on the Rosetta mission[7] .

With the rotational state of the asteroid determined using stereophotoclinometry, the state space that the particle filter must estimate shrinks from 8 to just 5. Also recall that the angular velocity components were the primary reasons the particle filter failed in the second scenario. Thus, it is more likely that the particle filter will succeed if these problem parameters are taken care of ahead of time.

## 8.4 Future Work

This work has shown that it is possible to determine the location and mass of a point mass concentration embedded in an asteroid. This was achieved by using the simple point mass gravity model described in section 6.2. This simple representation of the asteroid's gravitational model can then be replaced by more accurate models such as the polyhedron model developed by Dan Scheeres[9] . Rather than estimating a point mass, the average density of the asteroid and the components of the internal density gradient would be estimated. This will provide a more accurate description of the asteroid's density gradient, and thus a more accurate description of the gravitational field.

# 9 Conclusion

This paper has shown that it is possible to accurately determine the location and mass of a mass concentration within an asteroid given that a sufficient number of unique measurements are taken.

This work can now be expanded on by swapping out the point mass gravity model for the more accurate polyhedron model. The hope is that this work could shed a lot of light on the inner structure of asteroids and provide more accurate models of the dynamical systems of asteroids. It will hopefully be of much use to scientists and engineers who will rely on these models to safely navigate spacecraft in close proximity of asteroids.

# 10 References


[1] A. F. M. Smith and A. E. Gelfand, "Bayesian Statistics without Tears: A Sampling-Resampling Perspective", in *The American Statistician*, vol. 46, No. 2, pp. 84-88, 1992.

[2] H. Shaub and J. L. Junkins, H. Schaub and J. L. Junkins, Analytical Mechanics of Space Systems, 2nd ed., Reston, VA: AIAA Education Series, 2009.

[3] S. Sawai, D. J. Scheeres, and S. B. Broschart, in *Journal of Guidance, Control, and Dynamics*, vol. 25, No. 4, pp 786-795, 2002.

[4] N. Wheeler, Simplified Production of Dirac Delta Function Identities, Reed College Physics Department, Portland, OR, 1997.

[5] R. Eckhardt, "Stan Ulam, John von Neumann, and the Monte Carlo Method", in *Los Alamos Science Special Issue*, No. 15, pp. 131-141, 1987.

[6] A. Gelman and C. R. Shalizi, "Philosophy and the practice of Bayesian statistics", in *British Journal of Mathematical and Statistical Psychology*, vol. 66, pp. 8-38, 2013

[7] R. Pardo de Santayana, M. Lauer, P. Muñoz, and F. Castellini, "Surface Characterization and Optical Navigation at the Rosetta Flyby of Asteroid Lutetia", in *International Symposium on Space Flight Dynamics*, Laurel, MD, 2014.

[8] N. J. Gordon, D. J. Salmond, and A. F. M Smith, "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation", in *IEE Proceedings-F*, vol. 140, No. 2, pp. 107-113 1993.

[9] B. Gaudet and R. Furfaro, "Estimation of Asteroid Model Parameters Using Particle Filters", in *Advances in the Astronautical Sciences* Vol. 148, pp. 915-932., 2013

[10] L. A. McGee and S. F. Schmidt, "Discovery of the Kalman Filter as a Practical Tool for Aerospace and Industry", National Aeronautics and Space Administration Ames Research Center, Moffett Field, CA, 1985

[11] S. Ullah, "On the Use of Kalman, and Particle Filtering for Positioning, Navigation and Tracking in Autonomous Driving Vehicle", Northwestern University Electrical Engineering and Computer Science Department, Evanston, IL, 2015

[12] P. S. Maybeck, Stochastic Models, Estimation, and Control Volume 1, Academic Press, Inc., New York, NY, 1979

[13]