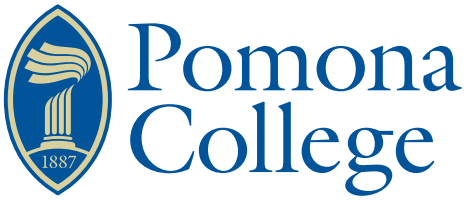


Senior Thesis in Mathematics

Kalman Filter Methods for Model Parametrization

Author:
Lindsey Tam

Advisor:
Dr. Blerta Shtylla

Submitted to Pomona College in Partial Fulfillment
of the Degree of Bachelor of Arts

April 23, 2020

Abstract

This thesis explores the theory and applications of Kalman Filters in both state and parameter estimation. Kalman Filters are able to recursively generate predictions for linear systems and these predictions become progressively more accurate because of the system's ability to correct predictions using incoming data. Nonlinear forms of the Kalman Filter exist, including the Extended Kalman Filter (EKF) and the Unscented Kalman Filter (UKF). The EKF works by linearizing the system around the mean, making it more effective for simpler and lower dimensional systems. The UKF addresses the shortcomings of the EKF by linearizing the system around multiple points, known as sigma points. After exploring the theory behind these techniques, this paper implements a few examples of the EKF and the UKF for both state and parameter estimation.

Contents

Chapter 1

Introduction

This paper explores the Kalman Filter (KF), Extended Kalman Filter (EKF), and the Unscented Kalman Filter (UKF). The KF is a data prediction method that estimates the value of states in linear dynamical systems. This is a recursive predictive-corrective process that enables the KF to continuously generate predictions about state variables without relying on large amounts of initial data. In the case of nonlinear systems, alternative forms of the Kalman Filter were developed, including the EKF and the UKF. The EKF linearizes the nonlinear system around the mean by taking the Jacobian of the nonlinear function. Navigation and signal processing are the main applications of the EKF. Instead of linearizing the system around a single point, the UKF utilizes many points, known as sigma points. The sigma points then undergo a nonlinear transformation known as the Unscented Transform. Applications of the UKF includes modeling biological systems. Beyond state estimation, these filters can also be used for parameter estimation. The method of parameter estimation explored in this thesis is known as joint parameter estimation and involves creating new states for the estimated parameters.

In addition to understanding the theory behind these algorithms, this paper also explores how they can be applied to real world systems. The first example observes a system that models the pathway of human metabolites, which are the byproducts of the metabolism. The goal of this example is to understand how the EKF works, how it can be used for multiple state correction, and how parameter estimation can be applied. A second example explored in this paper involves applying the UKF to the Van der Pol oscillator, which is a self sustaining nonconservative oscillator. The purpose of this simple example is to demonstrate how a simple version of UKF can be implemented on MATLAB while also observing how process and measurement noise influences the system. The third and final example uses the same metabolites system as the first example, but applies it in the context of the UKF. The parameter estimation results from the EKF and UKF on the metabolites system, are both promising.

Though performance of the two is similar, the UKF can be further optimized by adjusting the model's hyper-parameters, which the EKF does not have.

The ultimate goal of researching the UKF is to create a model that can forecast glucose values in real time of patients with Type 1 diabetes. This is inspired by a separate study that was able to do real-time glucose forecasting in Type 2 diabetes patients. Using an existing model of Type 1 diabetes that includes 12 states, the UKF will be most useful in extracting parameter values. Implementing this model has the potential to improve treatment methods and the quality of life of Type 1 diabetes patients.

Chapter 2

State Space Models

Applying the Kalman Filter to a system requires an understanding of state space models. A state space model represents a system's inputs, outputs, and states by describing a series of first order differential equations. A linear continuous state space system has a general form of

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + w(t) && \text{(system)}, \\ y(t) &= Hx(t) + v(t) && \text{(observation)}, \end{aligned}$$

where x is the state vector, A is a square transformation matrix, B is the input matrix, u is a system input, w is the process noise vector, y is the transformed prediction, H is the observation matrix, v is the measurement noise vector, and t is time.

For implementing Kalman Filters, the continuous state space system must be discretized into time steps, call it k. The general equation is given below as

$$\begin{aligned} x_k &= e^{At}x_{k-1} + \int_0^t e^{A(t-r))}Bu_k dr + w_k && \text{(system)}, \\ y_k &= Hx_k + v_k && \text{(observation)}, \end{aligned}$$

where e^{At} is the matrix exponential of A. Generally, the matrix exponential is very time-consuming to compute, especially if the matrix is not diagnolizable. In these cases, an alternative solution is to use Euler's method, which approximates the discretized system using small time steps, as shown by

$$\begin{aligned} x_k &= x_{k-1} + tf(x) && \text{(system)}, \\ y_k &= Hx_k + v_k && \text{(observation)}, \end{aligned}$$

where t is some small time step and $f(x)$ is the state after it undergoes a transformation. In most of the examples that are explored in this thesis, computing the matrix exponential is inefficient and results in excessive computing time. Therefore, Euler's method will be mostly used for discretization.

Consider an example of a moving mass, call it m, under a force, call it $u(t)$. The goal of this system is to predict the mass's position and velocity given its previous position and velocity. Therefore, position and velocity are the two states in this system. From differential equations, recall that the position of the mass is denoted by $x(t)$, the velocity of the mass is denoted by $\dot{x}(t)$, and the acceleration is denoted by $\ddot{x}(t)$. According to Newton's second law of motion, the force exerted on an object is given by the equation $F = ma$, or

$$u(t) = m\ddot{x}(t).$$

This is a second order differential equation. To convert it into its state space format, begin by transforming it into a first order differential equation by substituting $x_1(t)$ for $x(t)$ and $x_2(t)$ for $\dot{x}(t)$, resulting in

$$\dot{x}_1(t) = x_2(t),$$
$$\dot{x}_2(t) = \frac{u(t)}{m}.$$

For the sake of simplicity, assume this system has no process or measurement noise. The state vector, x, containing both position and velocity, can now be rewritten as

$$x = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}.$$

In order to predict the state of the system at the next time step, take the derivative of the state vector and rewrite it in vector-matrix form (which is possible because this system is linear), resulting in

$$\dot{x} = \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ \frac{u(t)}{m} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u(t).$$

Here, one can see how this form can be applied to the general linear state space model where $A = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ and $B = [0, \frac{1}{m}]^T$.

For the sake of this example, assume that only the position of the mass is measurable. Therefore, the transformed prediction, y, is given by

$$y = \begin{bmatrix} x_1(t) \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} x = x_1(t).$$

Now that the continuous state system is defined, it can be discretized. Since this example considers a simple linear system, the matrix exponential can be calculated as follows

$$e^{At} = \begin{bmatrix} e^0 & e^t \\ 0 & e^0 \end{bmatrix} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix},$$

$$\int_0^t e^{A(t-r))} e^{At} B u_k dr = \int_0^t e^{A(t-r))} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u(t)dt = \int_0^t e^{A(t-r))} \begin{bmatrix} \frac{t}{m} \\ \frac{1}{m} \end{bmatrix} u(t)dt = \begin{bmatrix} \frac{t^2}{2m} \\ \frac{t}{m} \end{bmatrix} u(t).$$

Now the discretized state space model can be written as

$$x = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} + x_{k-1} \begin{bmatrix} \frac{t^2}{2m} \\ \frac{t}{m} \end{bmatrix} u(t) \qquad \text{(system)},$$

$$y = \begin{bmatrix} 1 \\ 0 \end{bmatrix} x \qquad \text{(observation)}.$$

It is important to be able to discretize state space models because the goal of the Kalman Filter is to make some prediction about a future state. Therefore, discretization enables the Kalman Filter to make approximations about the system's state given past information. In addition, discretization makes numerical computing more efficient.

Chapter 3

Kalman Filters

The Kalman Filter (KF) recursively generates predictions for linear dynamical systems [**?**]. The basic foundations of the this algorithm include generating a prediction given some initial knowledge of the data and using actual measurements from the system to continually correct the prediction. Therefore, unlike other predictive methods like machine learning, the KF can begin generating estimates without large amounts of initial data. The KF begins by assuming the given data is noisy and Gaussian [**?**, **?**]. The first predictive step assumes knowledge of initial states and the model process. Since we know the system is linear, the model process can be denoted as a matrix and the states can be expressed as a vector. Through matrix vector multiplication, this algorithm simulates how states are transformed after undergoing some process. The next step involves calculating the covariance in order to calculate the Kalman Gain, which is a measure of how much the estimate should be changed given actual measurements of the system. The corrective step utilizes the Kalman Gain, which is a matrix representing a weight to correct the prediction. This process can be done recursively, allowing the model to become progressively more accurate as more data is added. Overtime, it is expected that the model will converge with the actual system measurements. Examples and code for implementing the KF are not explored in this thesis, but can be found in [**?**]

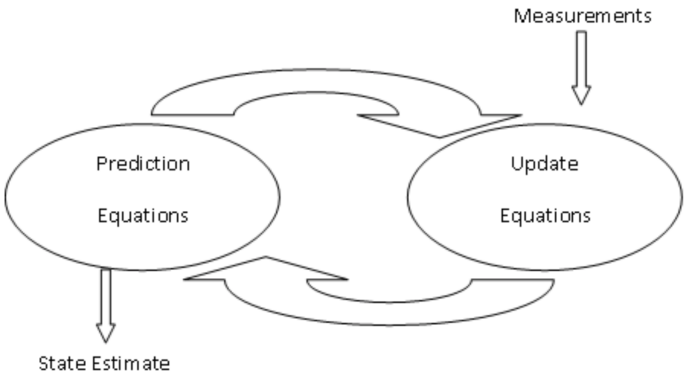


Figure 3.1: A basic diagram demonstrating the recursive nature of the Kalman Filter [**?**].

The most common application of KFs is in navigation, image processing, and finance. A relevant example is using computer vision to monitor and track vehicles in real time. This enables a traffic camera to know when to take a picture to capture a vehicle's license plate. Another example is the development of the Global Positioning Systems (GPS) [**?**]. The KF also has many aeronautical applications, which include long-distance flight and autopilot systems. In fact, the KF was initially created to help develop the navigation system in the Apollo Program [**?**].

The Kalman Filter is named after its developer, Rudolph Emil Kalman. Kalman was born on May 19, 1930 in Budapest, Hungary. After arriving in the United States, Kalman completed his undergraduate studies and masters degree in electrical engineering from Massachusetts Institute of Technology and completed his doctoral degree in Columbia University. He would spend the next years of his life teaching. In the 1960s to 1970s he became a professor at Stanford university [**?**]. In the 1970s and 1990s, Kalman spent time as a professor of engineering at the University of Florida. Kalman is most known for his work on the Kalman Filter, which was developed in the late 1950s. The Kalman Filter greatly aided the United States' military projects, resulting in former President Obama to award Kalman the National Medal of Science in 2009. In addition, in 1985, Kalman was awarded the Kyoto Prize, which is the Japanese version of the Nobel Peace Prize. Kalman passed away July 2, 2016 at the age of 86 and is survived by his wife and two children [**?**].

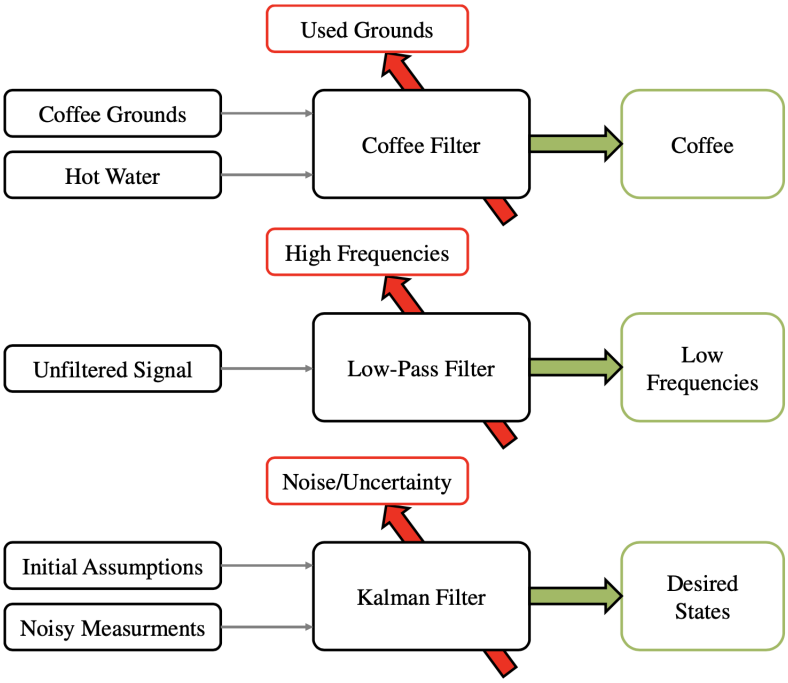


Figure 3.2: A simpler way to explore the KF is to facilitate a comparison with a coffee filter. This image is taken from a paper by Rhudy et al. comparing the Kalman Filter to a coffee filter [**?**].

3.1 Kalman Filter Algorithm

Before delving into the formal steps of the KF it is important to understand the underlying foundation of this algorithm. The goal of the Kalman Filter is to predict the new state of a system, after it undergoes some transformation. Let this be represented by

$$\frac{dx}{dt} = f(x) + \varepsilon',$$

where $\frac{dx}{dt}$ is the prediction of the states, f is a linear function that transforms the states, given by x, and ε' is internal system noise. Since the system we are considering is linear, we can represent this transformation using matrix multiplication. Let matrix M' be used to represent the transformation f in order to approximate the equation above as

$$\frac{dx}{dt} \approx M'x + \varepsilon'.$$

The Kalman Filter works by generating predictions and making correction at various time steps. Let these time steps be denoted as k, such that k is a

nonnegative integer and let x_k be the estimate of the state variables at time k with initial assumptions of the state variables beginning at x_0, such that $x_0 = \mathcal{N}(m_0, c_0)$, where m_0 and c_0 is mean and variance, respectively.

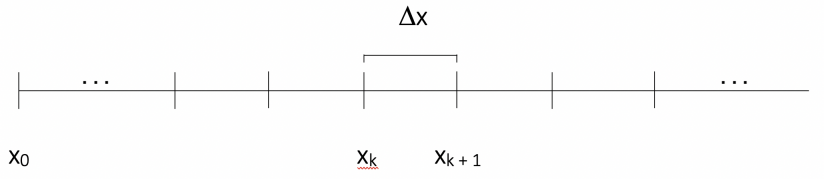


Figure 3.3: A diagram showing discretized time steps of the system. Though this diagram depicts time steps that are equal, this should not be assumed for all cases.

Next, the system can be discretized by approximating $\frac{dx}{dt}$ using the limit definition of derivative:

$$\begin{aligned}
\frac{x_{k+1} - x_k}{\Delta t} &\approx M' x_k + \varepsilon'_k \\
x_{k+1} - x_k &\approx M' x_k \Delta t + \varepsilon'_k \Delta t \\
x_{k+1} &\approx M' x_k \Delta t + \varepsilon'_k \Delta t + x_k \\
x_{k+1} &\approx (M' \Delta t + I) x_k + \varepsilon'_k \Delta t.
\end{aligned}$$

From here, substitute M for $M'\Delta t + I$ and ε_k for $\varepsilon'_k \Delta t$ to get the dynamics model:

$$x_{k+1} \approx M x_k + \varepsilon_k,$$

where $\varepsilon_k = \mathcal{N}(0, c_1)$ or Gaussian white noise. This general equation is significant to the generation of predictions and will continue to be used in different versions of the Kalman Filter.

Another important aspect of the KF is the incorporation of state measurements. We cannot assume that we will always have measurements for all states. More likely, we will only have measurements for a subset of states, Therefore, it is necessary to transform the prediction into a format that can be compared with state values. The transformed estimate of the the system measurement at the next time step, call it y_{k+1}, is given by

$$y_{k+1} \approx H x_{k+1} + v_{k+1},$$

where H is the matrix observation function and v_k is measurement noise vector at time step k such that $v_k = \mathcal{N}(0, c_2)$ or Gaussian white noise. H enables the state variables to be linearly transformed to match the outputs of the system. The dimensions of H reflects which state variables have measurable values in the system. It is not assumed that every state variable is measurable, so H

allows us to compare the measurable state variables to x_k. Simple applications of H include creating matrices with 0's and 1's, with 1's denoting that a state variable is measurable and a 0's representing non-measurable states. In other cases, H is an integer used as a scaling factor.

Given the dynamics model and data model, the algorithm of the KF can be discussed in greater depth. As a general overview, the KF algorithm consists of three major components:

1. Initialize state variables
2. Generate a prediction
3. Update prediction with measurements from the system.

The recursive component of the filter consists of repeating steps 2 and 3 repeatedly, while step 1 only needs to be done once. Details about each step are explored further below.

1. Begin by initializing the state estimate and the initial state covariance matrix. The state estimate, x_0, is a column vector containing state variables, call them $x_a, x_b, \ldots, x_n$, such that $x_0 = [x_a, x_b, \ldots, x_n]^T$, where T is the transpose. The state estimate can be found by taking the expected value of the data, which is normally distributed. If the system states are finite, the expectation is denoted by

$$\mathbb{E}[x_0] = \sum_{i=a}^{n} x_i p_i = [x_a p_a + x_b p_b + \ldots + x_n p_n]^T,$$

and if the system states are continuous case, the expectation is denoted by

$$\mathbb{E}[x_0] = \int_a^n x f(x) dx$$

where $f(x)$ is the probability density function. The state covariance matrix, P_0, is a square matrix whose contents are the covariance of the pairwise elements [1]. A state covariance matrix is a symmetrical positive semi-definite square matrix whose diagonals correspond to the variance of a variable at location i and elsewhere is the covariance of the pairwise elements. In practice, covariance matrices help us better understand the spread of data. For the case of the KF, calculating the state covariance is necessary for computing Kalman Gain, which is used in the correction step. Calculating the state covariance matrix can be done by

$$P_0 = \begin{bmatrix} var(x_a) & \ldots & cov(x_a, x_n) \\ \vdots & \ddots & \vdots \\ cov(x_n, x_a) & \ldots & var(x_n) \end{bmatrix}.$$

[1] Recall that the covariance of a variable with itself is the variance of the variable

Enough should be known about the modeled system to generate these values. In the case where one knows the true value of the initial states, the state covariance would have all 0 values. On the other hand, if one is unsure of the initial values, values in P_0 are expected to be higher. It is important to initialize the model with values close to the true value, else the system will converge at a slow rate.

2. After initializing the state estimate and state covariance, a prediction can be generated. The estimate of the system at the next time step, x_{k+1}, is given by

$$x_{k|k-1} = Fx_{k-1|k-1} + w_{k-1},$$

where F is the state transition matrix and w_k is the process noise vector. Every state variable contained in x_k is defined by a linear differential equation. These linear differential equations can be used to generate the F matrix. Therefore, F should be a square matrix whose dimension is equal to the number of states variables.

w_k is the process noise vector at time k. Process noise can be thought of as the model's accuracy. When process noise is 0, it implies that the model is perfectly accurate and does not have to correct for incoming system measurements. On the other hand, high process noise will essentially restart the system based on incoming measurements. w_k has the same dimensions as x_k, allowing us to identify whether or not to adjust the equations for the state variables.

3. Next, correct the prediction with incoming measurements from the system. Begin by calculating the state covariance matrix in order to calculate Kalman Gain. The state covariance matrix at time step k given the last time step, is

$$P_{k|k-1} = FP_{k-1}F^T + Q_{k-1},$$

where F^T is the transpose of F, and Q_k is the process noise covariance of w_k. The Kalman Gain at time step k, is given by

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R_k)^{-1},$$

where H^T is the transpose of the observation matrix and R is the measurement noise covariance matrix.

From this equation, one can see that balancing Q and R is critical for model performance. Larger values of Q indicate higher modeling error, which leads to a higher Kalman Gain and increased model correction. On

the other hand, large values of R imply high measurement error, leading to a lower Kalman Gain and less model correction.

The Kalman Gain is a measure of how much to change the model based on incoming data. Low values of the Kalman Gain imply the model is accurate while higher values indicate the model should adjust based on the incoming data.

Next, calculate the transformed prediction in order to correct the prediction. The transformed state vector, $\hat{y}_k$, is given by

$$\hat{y}_k = Hx_{k|k-1} + v_k,$$

where v_k is measurement noise, which is added to account for measurement error. The corrected prediction, $x_{k|k-1}$, is given by

$$x_{k|k} = x_{k|k-1} + K_k(y_k - \hat{y}_k),$$

where y_k is the actual measurements of the system, K_k is the Kalman Gain matrix at time step k, and x_{k-1} are the values of the state variable at the last time step. The quantity $y_k - \hat{y}_k$ is also known as measurement residual or innovation. Later on, we will see how this value is used to gauge model performance.

The final step is to update the state covariance matrix, P_k, through the equation

$$P_{k|k} = (I - K_k H)P_{k|k-1},$$

where I is the identity matrix, K_k is the Kalman Gain at time step k, H is the observation matrix, and $P_{k|k-1}$ is the state covariance at time step k given the last time step. P_k will be used in the next iteration of the filter.

Steps 2 and 3 may be repeated to progressively generate more accurate predictions.

Table 3.1: Description of all variables in the Kalman Filter

Variables in the Kalman Filter		
Variable	Description	Dimensions
x	State variables	$x \times 1$
$\hat{y}$	Transformed state vector	$y \times 1$
y	Actual system measurement(s)	$y \times 1$
v	Measurement noise vector	$y \times 1$
w	Process noise vector	$x \times 1$
F	State function	$x \times x$
H	Observation function	$y \times x$
K	Kalman Gain	$x \times y$
Q	Process noise covariance	$x \times x$
R	Measurement noise covariance	$y \times y$
P	Covariance matrix	$x \times x$

Chapter 4

Extended Kalman Filters

The Extended Kalman Filter (EKF) is one of the non-linear versions of the Kalman Filter. For the most part, the EKF algorithm is nearly identical to the KF algorithm. The critical difference is in linearizing the non-linear state and observation function. The EKF uses the Jacobian to linearly approximate the non-linear function around the mean of the Gaussian distribution. Skipping this step would result in the transformed data being non-Gaussian; though taking the Jacobian enables the transformation to remain Gaussian, it is not exact, resulting in some generalization. Linear approximation through a single point also makes the EKF inefficient when dealing with complex, higher order systems. Because of this, the model is highly subject to error, which can be somewhat reduced by setting accurate initial values. Though these flaws exist, the EKF performs strongly with applications of real time spatial fields, including navigation and positioning systems.

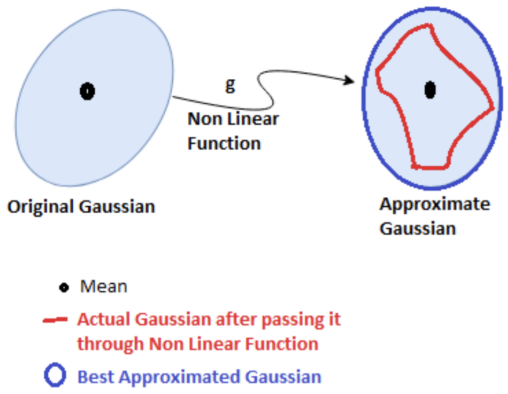


Figure 4.1: Overview of how the EKF approximates the Gaussian distribution using the mean after the system has been transformed by a nonlinear function.

4.1 Extended Kalman Filter Algorithm

Similar the the KF, the EKF has two main models:

$$\begin{aligned}&\text{state model: } x_k = f(x_{k-1}) + w_{k-1}\\&\text{data model: } \hat{y}_k = h(x_k) + v_k,\end{aligned}$$

where x_k is the state vector at time k, f is the vector-valued non-linear transformation function, w_k is the process noise at time k such that $w_k = \mathcal{N}(0, Q)$ or gaussian white noise with covariance Q, h is the vector-valued nonlinear observation function, and v_k is the measurement noise at time k such that $v_k = \mathcal{N}(0, R)$ or gaussian white noise with covariance R.

The main difference between these models and the models from the KF is that f and h are vector-valued non-linear functions. A prediction can be obtained by taking results from the last time step and transforming them using f, which is the non-linear set of ODEs used to describe the system. Similarly, the transformed prediction can be generated by taking the ODEs of states that are measurable, h, to transform the prediction and then add measurement noise.

Although using the non-linear function works for the state and data model, a non-linear function cannot be used in the correction step, such as calculating covariance of the state or the Kalman Gain. To overcome this, the EKF linearizes around the non-linear system by calculating the Jacobian matrix of f and h. Recall that the Jacobian matrix allows for the linear approximation of a non-linear system by taking first-order partial derivatives. Let F, the Jacobian of f, be given by

$$F = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix},$$

where f_i is the ODE corresponding to the ith state variable and x_i is the ith state variable. One can see that this results in F being a square matrix. In addition, let H be the Jacobian of h, given by

$$H = \frac{\partial h}{\partial x} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix},$$

where h_i is the ODE corresponding to the ith measurable state variable and x_i is the ith state variable.

Using these linearized versions of f and h, the correction step can be applied, making the algorithm similar to the KF. Recall that the algorithm consists of three major components:

1. Initialize the state variables
2. Generate a prediction
3. Update prediction with measurements from the system.

The recursive component of the filter consists of repeating steps 2 and 3 repeatedly, while step 1 only needs to be done once. For the most part, the EKF is going to be very similar to the KF, with a few exceptions. Details are explained more in depth below.

1. Begin by initializing the state estimate, x_0, and the initial covariance matrix, P_0. Similar to the KF, the initial state estimate can be found by

$$\mathbb{E}[x_0] = \sum_{i=a}^{n} x_i p_i = [x_a p_a + x_b p_b + \ldots + x_n p_n]^T,$$

   if the system is finite, but otherwise, if it is continuous, it is given by

$$\mathbb{E}[x_0] = \int_a^n x_i p_i dx,$$

   where $x_a, x_b, \ldots, x_n$ are the state variable, $p_a, p_b, \ldots, p_n$ is the probability of getting each state variable, and T is the transpose. Initialize the covariance matrix by

$$P_0 = \begin{bmatrix} var(x_a) & \ldots & cov(x_a, x_n) \\ \vdots & \ddots & \vdots \\ cov(x_n, x_a) & \ldots & var(x_n) \end{bmatrix}.$$

   Often, P_0 will be initialized as a diagonal matrix with the diagonals being the variance of each state variance and every other value being set to 0.

2. After initializing the model, a prediction can be generated. The prediction step deviates from the Kalman Filter because we can no longer represent the transformation through matrix multiplication because the system is no longer linear. Instead, the prediction is generated by directly applying the last state estimate to the non-linear transformation f and adding process noise w_k, which is given by

$$x_{k|k-1} = f(x_{k-1|k-1}) + w_{k-1}.$$

3. The correction step requires transforming the prediction and calculating the Kalman Gain. Begin by linearizing both f and h by calculating the

Jacobian. Let F be the Jacobian of f and let H be the Jacobian of h, both of which are given by

$$F = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix},$$

$$H = \frac{\partial h}{\partial x} = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix},$$

where f_i is the ODE corresponding to the ith state variable, h_i is the ODE corresponding to the ith measurable state variable, and x_i is the ith state variable.

Using F and H, one can now calculate the state covariance and the Kalman Gain. Recall that the Kalman Gain is a method to determine whether to place more weight on incoming measurements or the system's model. $P_{k|k-1}$, can be found by

$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q_{k-1},$$

where F is the Jacobian of the nonlinear state function, P_{k-1} is the state covariance at the last time step, T is the transpose, and Q is the process noise covariance. Using $P_{k|k-1}$, calculate the Kalman Gain by

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R_k)^{-1},$$

where H is the Jacbian of the nonlinear observation matrix, and R is the measurement noise covariance. Recall that H and F are linear approximations of f and h. Therefore, it is assumed that there is some amount of error when calculating the Kalman Gain.

The next component of the correction step is to transform the prediction to compare with incoming measurements from the system, y_k. The transformed prediction, $\hat{y}_k$, is given by

$$\hat{y}_k = h(x_{k|k-1}) + v_k,$$

where h is the vector-valued non-linear observation function and v_k is measurement noise. The equation can be updated by

$$x_{k|k} = x_{k|k-1} + K_k(y_k - \hat{y}_k).$$

Finally, update the covariance matrix, P_k, which will be used in the next iteration of the filter. P_k can be updated by

$$P_{k|k} = (I - K_k H) P_{k|k-1},$$

where I is the identity matrix, K_k is the Kalman Gain at time k, $P_{k|k-1}$ is the covariance matrix given the last time step, and H is the Jacobian of the nonlinear observation matrix.

Table 4.1: Description of all variables in the Extended Kalman Filter

Variables in the Extended Kalman Filter		
Variable	Description	Dimensions
x	State variables	$x \times 1$
$\hat{y}$	Transformed state vector	$y \times 1$
y	Actual system measurement(s)	$y \times 1$
v	Measurement noise	$y \times 1$
w	Process noise	$x \times 1$
f	Non linear state function	$x \times 1$
F	Jacobian of f	$x \times x$
h	Non linear observation function	$y \times 1$
H	Jacobian of h	$y \times x$
K	Kalman Gain	$x \times y$
Q	Process noise covariance	$x \times x$
R	Measurement noise covariance	$y \times y$
P	Covariance matrix	$x \times x$

4.2 Modeling Metabolites using EKF

This example is inspired by research that models the biological pathway of metabolites, which are molecules that are the byproduct of the body's metabolism. This model contains four different states, or metabolites, which contain 18 unknown parameters [**?**]. Unlike the original research, which had datasources of their own, this example will simulate data by using a built-in ODE solver on Matlab. The dataset used to generate these results are included along with the code in the appendix. Ultimately, the goals of this example is to demonstrate how the EKF can be applied, how the EKF can correct for multiple variables or states, and how the EKF can be used for parameter fitting.

In this example, the four metabolites, or states, have the following differential equations:

$$\begin{aligned}
\dot{x}_1 &= \alpha_1 x_3^{g_{13}} - \beta_1 x_1^{h_{11}}, \\
\dot{x}_2 &= \alpha_2 x_1^{g_{21}} - \beta_2 x_2^{h_{22}}, \\
\dot{x}_3 &= \alpha_3 x_2^{g_{32}} - \beta_3 x_3^{h_{33}} x_4^{h_{34}}, \\
\dot{x}_4 &= \alpha_4 x_1^{g_{41}} - \beta_4 x_4^{h_{44}},
\end{aligned}$$

where there are 18 unknown parameters ($\alpha_1, \ldots, \alpha_4, \beta_1, \ldots, \beta_4, g_{13}, g_{15}, g_{21}$, $g_{32}, g_{41}, h_{11}, h_{22}, h_{33}, h_{34}, h_{44}$) and the state vector is given by $x = [x_1, x_2, x_3, x_4]^T$. For now, we will use the true value of these parameters (which can be found in [**?**]), but later on the EKF will be applied for parameter estimation. In both the original example as well as this one, sampling time will be 0.1 seconds for 5 seconds, totaling 50 UKF estimates.

Implementing EKF

Recall that implementing the EKF requires initializing the model, generating a prediction, and correcting the prediction. Below outlines the process of how the EKF can be applied for a single time step. Applying the EKF to multiple time steps is simulated on Matlab, and the code for doing so can be found in the appendix.

1. The model is initialized by setting the state variable to $x_0 = [4, 1, 3, 4]^T$, which is very close to the true values of the system. Since the initialized state has values near the true value, the state covariance can also be set to a lower value of $P_0 = .01I$.

2. The next step is to generate a prediction. To do so, the system must be discretized into time steps. The system is given in a continuous state space in the form of $\dot{x} = Ax$, and can be discretized by converting it into the form $\dot{x} = e^{At}x$. However, in this system, calculating the matrix exponential is computationally intense (taking days to run on a single iteration on Matlab). Therefore, the Euler method will be used with a time step of 0.05 seconds, resulting in

$$\begin{aligned} x_{1|0} &= x_{0|0} + 0.05 * f(x) \\ &= \begin{bmatrix} 4 \\ 1 \\ 3 \\ 4 \end{bmatrix} + 0.05 \begin{bmatrix} \alpha_1 3^{g_{13}} - \beta_1 4^{h_{11}} \\ \alpha_2 4^{g_{21}} - \beta_2 1^{h_{22}} \\ \alpha_3 1^{g_{32}} - \beta_3 3^{h_{33}} 4^{h_{34}} \\ \alpha_4 4^{g_{41}} - \beta_4 4^{h_{44}} \end{bmatrix} \\ &= \begin{bmatrix} 3.4152 \\ 1.6500 \\ 2.5786 \\ 3.2906 \end{bmatrix}. \end{aligned}$$

3. Finally, the prediction can be corrected. Calculate F, the jacobian of the nonlinear transformation $f = [\dot{x}_1, \dot{x}_2, \dot{x}_3, \dot{x}_4]^T$, and H, the Jacobian of the nonlinear observation function $h = [\dot{x}_1, \dot{x}_2, \dot{x}_3, \dot{x}_4]^T$, as follows

$$F = H = \begin{bmatrix} -\beta_1 h_{11} x_1^{h_{11}-1} & 0 & x_5 x_3 g_{13}^{g_{13}-1} & 0 \\ \alpha_2 g_{21} x_1^{g_{21}-1} & -\beta_2 h_{22} x_2^{h_{22}-1} & 0 & 0 \\ 0 & \alpha_3 g_{32} x_2^{g_{32}-1} & -\beta_3 h_{33} x_3^{h_{33}-1} x_4^{h_{34}} & -\beta_3 h_{34} x_4^{h_{34}-1} x_3^{h_{33}} \\ \alpha_4 g_{41} x_1^{g_{41}-1} & 0 & 0 & -\beta_4 h_{44} x_4^{h_{44}-1} \end{bmatrix}.$$

H should be adjusted according to how many states are being corrected. For instance, if correction is being done for all states, H should remain as it is above. However, if only the first state is being corrected, then

$h = [\dot{x}_1]$ and $H = [-\beta_1 h_{11} x_1^{h_{11}-1}, 0, x_3 g_{13}^{g_{13}-1}, 0]$.

Assuming that the first state is the only state being corrected, F and H are used to calculate the Kalman Gain, which is

$$K_1 = \begin{bmatrix} .05402 & -.0370 & .0465 & .0149 \end{bmatrix}^T.$$

Additionally, the prediction can be transformed in order to compare it with incoming system measurement. In the case where only the first state is being predicted, the prediction is transformed by applying the nonlinear observation function to the prediction. Similar to generating the prediction, this can be done by using Euler's method to discretize the system, resulting in

$$\begin{aligned} y_1 &= x_{1,1|0} + 0.05 * h(x_{1|0}) \\ &= 3.4152 + 0.05(\alpha_1 2.5786^{g_{13}} - \beta_1 3.4152^{h_{11}}) = 2.9600. \end{aligned}$$

This transformed value is quite close to the measured value of 3.1126, meaning that in the first time step, the residual is approximately .15. Recall that the measured value was simulated on Matlab and can be found in the appendix.

Finally, the prediction can be corrected, resulting in

$$x_{1|1} = x_{1|0} + K_1(y_1 - \hat{y}_1) = \begin{bmatrix} 3.5751 \\ 1.5404 \\ 2.7162 \\ 3.3346 \end{bmatrix}.$$

Applying the EKF for 50 time steps by repeating steps 2 and 3 on Matlab results in figure **??**. Multiple state correction is similar to the implementation above. The only difference is in the correction step. Notice the improvement in performance when all states are being corrected, as seen in fig **??**.

For more examples on applying the EKF and MATLAB code, the following can be referenced [**?**, **?**].

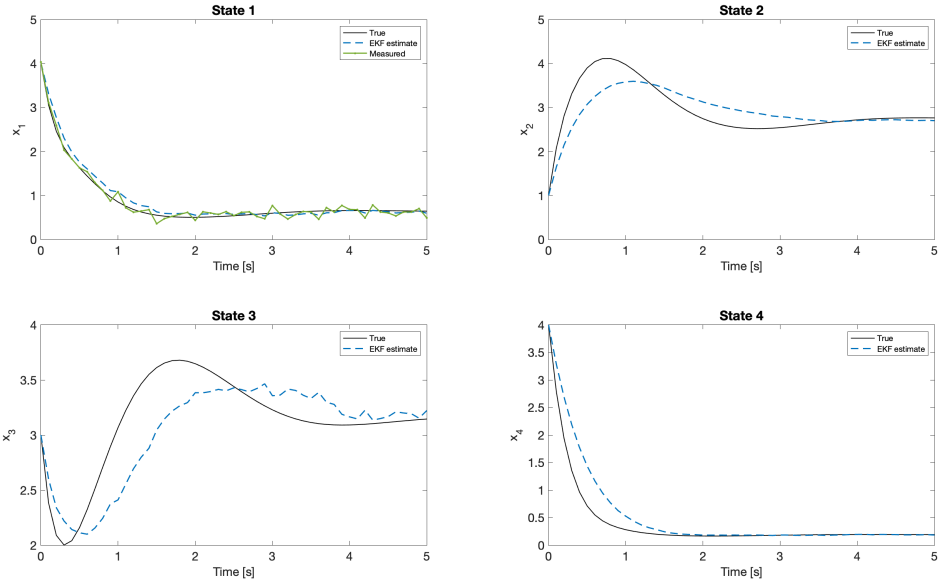


Figure 4.2: Implementing EKF with correction applied to State 1 and initialized with states $[4, 1, 3, 4]^T$.

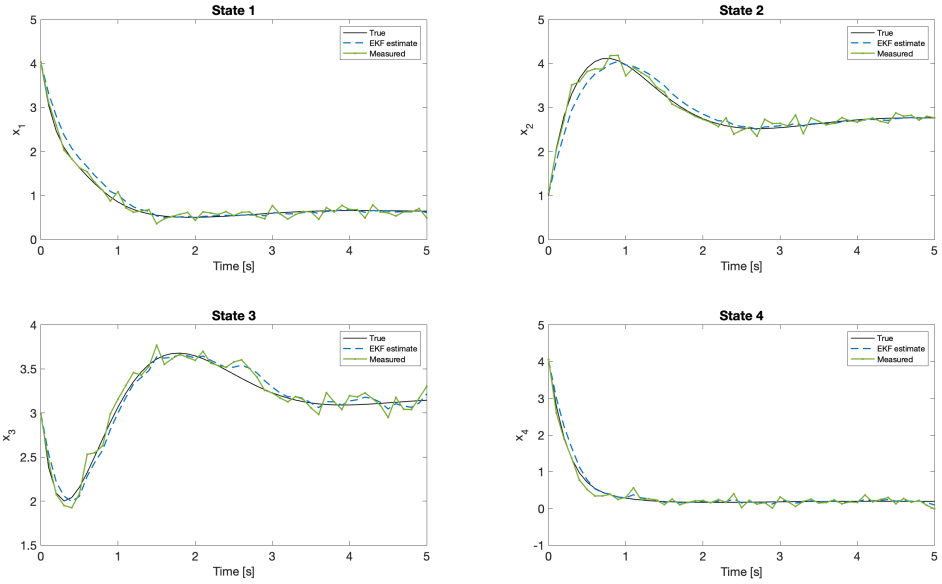


Figure 4.3: Using EKF with correction applied to all states using a model initialized with states $[4, 1, 3, 4]^T$.

EKF Parameter Estimation

The EKF (as well as other versions of the Kalman Filter), can be utilized for parameter estimation. There are many methods of parameter estimation, which include using dual techniques. In fact, there is a nonlinear version of the Kalman Filter, called the Dual Unscented Kalman Filter, that was designed in order to simultaneously monitor state and parameter values. Though this thesis does not explore these methods, more information can be found in [**?**, **?**]. In particular, this thesis explores the use of joint parameter estimation. This technique involves creating new states for each of the unknown parameters.

In this example, the goal is to use the EKF to find parameter values that best fit this model to its true states. This example continues to use the metabolites system with the goal of estimating parameter α_1. Begin by declaring α_1 as a separate state. The new set of differential equations for this 5 state system is given by

$$\begin{aligned}
\dot{x}_1 &= x_5 x_3^{g_{13}} - \beta_1 x_1^{h_{11}}, \\
\dot{x}_2 &= \alpha_2 x_1^{g_{21}} - \beta_2 x_2^{h_{22}}, \\
\dot{x}_3 &= \alpha_3 x_2^{g_{32}} - \beta_3 x_3^{h_{33}} x_4^{h_{34}}, \\
\dot{x}_4 &= \alpha_4 x_1^{g_{41}} - \beta_4 x_4^{h_{44}} \\
\dot{x}_5 &= 0.
\end{aligned}$$

The ODE for x_5 is 0, because x_5 is a parameter, or constant, and does not change overtime. In this 5 state system, the nonlinear state transition function is equal to $f = [\dot{x}_1, \dot{x}_2, \dot{x}_3, \dot{x}_4, \dot{x}_5]^T$. By applying the EKF to this system while correcting for all 5 states, the EKF performs well, as shown in **??**.

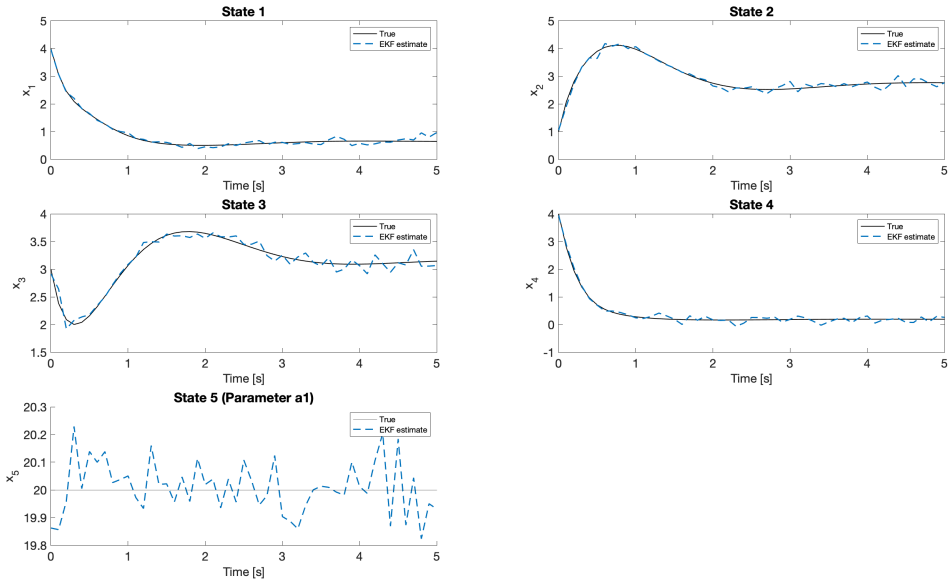


Figure 4.4: Results obtained by implementing the EKF with parameter estimation of α_1. The model is initialized with $[4, 1, 3, 4, 20]^T$, which is close to the true initial value of the system. While not depicted in this visual, the model still performs well when initialized with values further from the true system values.

Parameter estimation can be applied to more than one parameter. In this example, consider parameter estimation for four states, $\alpha_1, \alpha_2, \alpha_3, \alpha_4$. This introduces four new states into the original four state system, resulting in

$$\begin{aligned}
\dot{x}_1 &= x_5 x_3^{g_{13}} x_5^{g_{15}} - \beta_1 x_1^{h_{11}}, \\
\dot{x}_2 &= x_6 x_1^{g_{21}} - \beta_2 x_2^{h_{22}}, \\
\dot{x}_3 &= x_7 x_2^{g_{32}} - \beta_3 x_3^{h_{33}} x_4^{h_{34}}, \\
\dot{x}_4 &= x_8 x_1^{g_{41}} - \beta_4 x_4^{h_{44}} \\
\dot{x}_5 &= \dot{x}_6 = \dot{x}_7 = \dot{x}_8 = 0.
\end{aligned}$$

The results of the EKF when applied to this 8 state system is shown in figure **??**. After 50 time-steps, the EKF prediction of these four parameter is quite close to their true values, as shown in Table **??**.

Parameter Values			
Parameters	True	EKF estimate	IUKF Estimate
α_1	20.0	20.0125	19.9744
α_2	8.0	7.9819	8.0017
α_3	3.0	2.9818	3.0015
α_4	2.0	1.9614	2.0016

Table 4.2: This table shows the true values of the parameters, the final EKF prediction of the parameters, and the final IUKF prediction of the parameters. Here, the term final is being used to denote the performance of the filter after 50 time-steps.

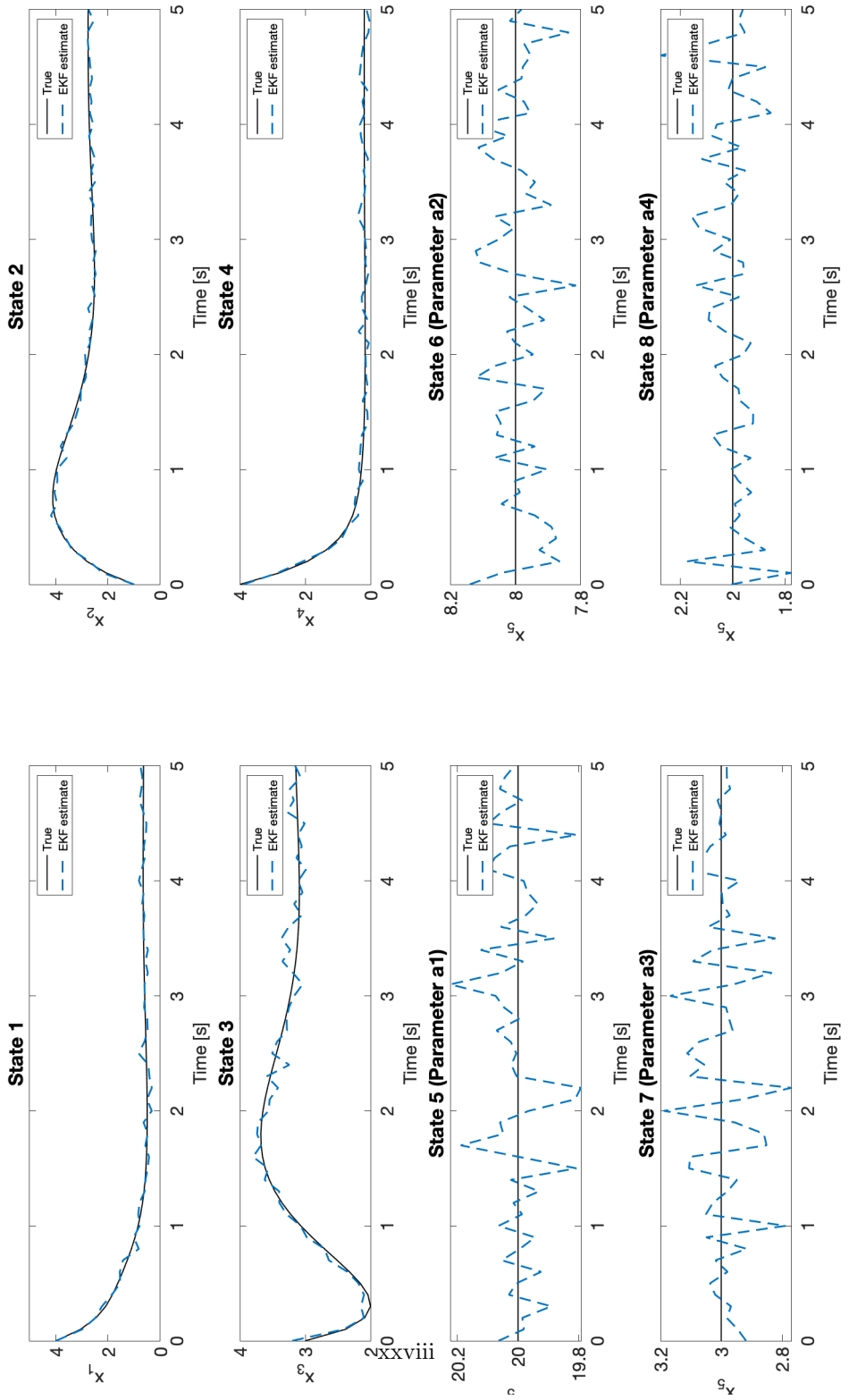


Figure 4.5: EKF performance of an 8 state system with 4 parameters. This model was initialized with $[4; 1; 3; 4; 20; 8; 3; 2]^T$, which is almost exactly the system's true values.

Chapter 5

Unscented Kalman Filters

The Unscented Kalman Filter (UKF) is another nonlinear version of the Kalman Filter and was developed to address the shortcomings of the EKF. As opposed to using the Jacobian to linearly approximate around a single point, the UKF uses the Unscented Transform (UT) to approximate around multiple points, known as sigma points. The UT is a method of approximating probability distributions that have undergone a non linear transformation using limited statistics. The UT uses these sigma points, which are represented in a Sigma Point Matrix, to represent the normal distribution of the data. Then, the sigma points undergo a non-linear transformation, resulting in a posterior distribution that is not normal [**?**, **?**] . We are able to approximate the normal distribution of the posterior distribution using the weights and covariance that were calculated prior to the transformation. This process enables the Kalman Filter to be applied to more complex non linear problems.

Unlike the Kalman Filter and the Extended Kalman Filter, the UKF also has a set of parameters. Explanations of each parameter and their default values can be found in Table 4.2. Here, parameters are necessary for controlling the spread of sigma points. This was not needed for the EKF, since the EKF was only linearizing around the mean. The thesis will not be exploring how to tune these parameters, but one method to do so is through ad hoc testing.

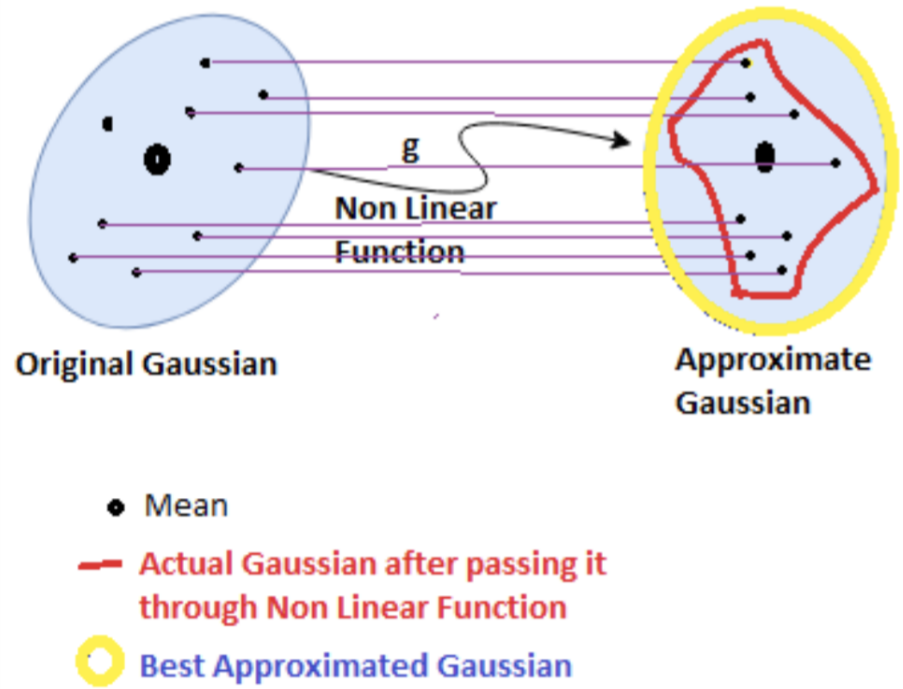


Figure 5.1: A depiction of the general overview of the UKF. Here we see that sigma points are sampled from a Gaussian distribution and are propagated through a non-linear function, called g. Though the output is non-Gaussian, an approximation of the Gaussian distribution can be obtained.

The term 'unscented' was arbitrarily coined by the developer of the UKF, Jeffrey Uhlmann. In an interview, he shares:

> "Initially I only referred to it as the "new filter." Needing a more specific name, people in my lab began referring to it as the "Uhlmann filter," which obviously isn't a name that I could use, so I had to come up with an official term. One evening everyone else in the lab was at the Royal Opera House, and as I was working I noticed someone's deodorant on a desk. The word "unscented" caught my eye as the perfect technical term. At first people in the lab thought it was absurd—which is okay because absurdity is my guiding principle—and that it wouldn't catch on. My claim was that people simply accept technical terms as technical terms: for example, does anyone think about why a tree is called a tree?"

5.1 Unscented Kalman Filter Algorithm

Before delving into the details of the algorithm, we will explore a high-level overview of the UKF process. As with the KF and the EKF, the UKF follows the three major components of:

1. Initializing the model's states
2. Generating a prediction
3. Updating prediction with measurements from the system.

In the first step, initializing the model requires calculation of sigma points. These sigma points characterize the normal distribution of the data.

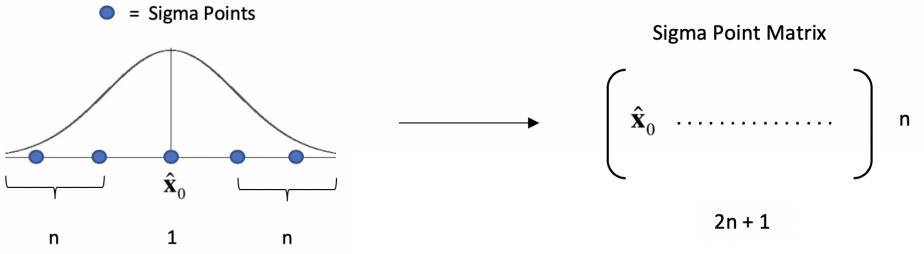


Figure 5.2: A brief illustration on how sigma points are used to characterize the Gaussian distribution of the data and how these sigma points are used to generate the Sigma Point Matrix.

After the sigma points undergo a nonlinear transformation, the resulting transformed sigma point matrix can be used to approximate the normal distribution of the data, which we can use to generate a prediction.

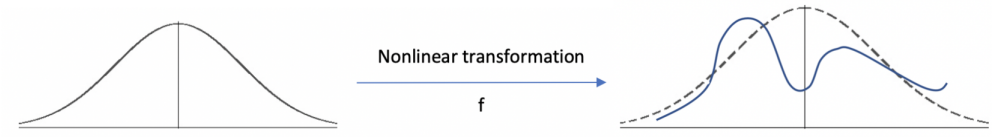


Figure 5.3: After the Gaussian-distributed data undergoes a nonlinear transformation, call it f, the result is no longer linear

One can see that this follows the basic outline of the previous versions of the KF, but the process of doing each step varies. More details about each step are explained below.

1. The first step is to initialize state vector, x_0, and covariance, P_{x_0}, which is exactly similar to both the KF and the EKF. Since we know that the x_0 is normally distributed, the expectation can be calculated by

$$x_0 = \mathbb{E}[x_0] = \sum_{i=a}^{n} x_i p_i = [x_a p_a + x_b p_b + \ldots + x_n p_n]^T,$$

where $x_0 = [x_a, x_b, \ldots, x_n]^T$, $p_a, p_b, \ldots, p_n$ is the respective probability of obtaining each state variable and T is the transpose. Next initialize the state covariance, by

$$P_{x_0} = \mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T].$$

Unlike the KF and EKF, the UKF requires the calculation of sigma points, which is a way characterizing the distribution of the data. The number of sigma points is deterministic and depends on the dimensions of the system. In general, a UFK will have $2x + 1$ sigma points, where x represents the dimension of the state vector [**?**, **?**, **?**]. All sigma points are stored in a sigma point matrix, called χ. To get a general idea about the distribution of the data, we use λ, which is a scalar value that determines how spread out the sigma points are from the mean. λ can be calculated by

$$\lambda = \alpha^2(\mathrm{x} + \kappa) - x,$$

where x is a scalar representing the number of states in the system, and α and κ are both parameters that control for the spread of sigma points around the mean value of the state. The spread of the sigma points is proportional to α. For both α and κ, the smaller the values are, the closer the sigma points are to the mean.

Another parameter of the UKF is β, which uses information regarding state distribution to adjust sigma points. β has a default value of 2 if the data is Gaussian (which is an assumption we will make throughout this paper).

Since the goal of this step is to characterize the distribution, set one of these sigma points to the mean, $\chi_{\ 0,k-1}$ which can be expressed as

$$\chi_{\ 0,k-1} = x_{k-1|k-1},$$

where 0 is a row on χ and $k-1$ is the time step. Then allocate half of the remaining points will be smaller than the mean and the other half will be larger than the mean by

$$\chi_{\ i,k-1} = x_{k-1|k-1} + \left(\sqrt{(x+\lambda)P_{x,k-1}}\right)_i \qquad i = 1, \ldots, x,$$

$$\chi_{\ i,k-1} = x_{k-1|k-1} - \left(\sqrt{(x+\lambda)P_{x_{k-1}}}\right)_{i-n} \qquad i = x+1,\ldots,2x+1,$$

where x is the dimension of the state vector, $(\sqrt{(d_x+\lambda)P_{x_{k-1}}})$ is a matrix, and the i subscript is the i^{th} column of the matrix. Recall that the square root of a matrix, satisfies the following condition: $A = B^2$, where A and B are both matrices.

Next, we must calculate a weight for each sigma point. Weights are scalars used to calculate posterior sigma points after they have undergone a nonlinear transformation. One set of weights, $W^{(m)}$, will be used to calculate the posterior mean while another set of weights $W^{(c)}$ will be used to calculate the posterior covariance. Weights can have positive or negative values, but will ultimately sum to 1 [**?**].

The initialized weight for the mean, $W_0^{(m)}$ can be found by

$$W_0^{(m)} = \frac{\lambda}{x+\lambda},$$

where x is the number of states in the system. Similarly, the weight for the covariance at the initial time step, $W_0^{(c)}$, is given by

$$W_0^{(c)} = \frac{\lambda}{x+\lambda} + (1-\alpha^2+\beta).$$

In later time steps, $W_i^{(m)}$ and $W_i^{(c)}$ follow the same equation, given by

$$W_i^{(m)} = W_i^{(c)} = \frac{\lambda}{2(x+\lambda)} \qquad i = 1,\ldots,2x.$$

2. A prediction can be generated by performing a nonlinear transformation on the sigma point matrix, χ, in order to generate a prediction and provide an update on the covariance matrix. After χ undergoes a nonlinear transformation, f, the result is a transformed sigma point matrix at time step k given the last time step, $k-1$, which will is called $\chi_{k|k-1}$ and is given by

$$\chi_{k|k-1} = f(\chi).$$

Though χ has a Gaussian distribution, $\chi_{k|k-1}$ does not because it has been transformed by the nonlinear state function f. The sum of the columns in $\chi_{k|k-1}$ and the weights calculated in step 3 will then be used to generate a prediction of the state variables, x_k, by

$$x_{k|k-1} = \sum_{i=0}^{2x} W_i^{(m)} \chi_{i,k|k-1}.$$

Adding the weights help approximate the Gaussian distribution of the state variables after they undergo a transformation.

3. Now that the prediction component of the filter is completed, we move on to the correction step, which begins by calculting the transformed covariance matrix and then transforming our predictions into a format that can be compared with system measurements. Calculation of the posterior (also called augmented) sigma points is necessary for converting system measurements into a format that can be compared with the state variables. System measurements must undergo a non linear transformation, h, resulting in a non Gaussian distribution. Therefore, the Unscented Transform is used again. Begin by calculating the posterior covariance matrix for the state variable, which is necessary for updating the state covariance later on by

$$P_{x,k|k-1} = \sum_{i=0}^{2x} W_i^{(c)} (\chi_{i,k|k-1} - x_{k|k-1})(\chi_{i,k|k-1} - x_{k|k-1})^T + Q,$$

where Q is process noise that provides the error in our model f, T is the transpose, and $W_i^{(c)}$ are the weights calculated earlier.

Then calculate augmented sigma points, $\chi^{(aug)}$, by

$$\chi_{0,k|k-1}^{(aug)} = x_{k|k-1}$$

$$\chi_{i,k|k-1}^{(aug)} = x_{k|k-1} \pm \left(\sqrt{(x+\lambda)P_{x_k}} \right)_i \qquad i = 1, \ldots, 2x$$

Recall that λ was calculated in step 2. Since λ is not time dependent, we can use the same value used earlier.

Now we calculate a sigma point matrix that represents the transformation of the prediction so that it can be compared with the states. This is necessary, especially in cases where measurements are not being obtained for all state variables. This sigma point matrix, $\mathcal{Y}_{k|k-1}$, can be obtained by having the $\chi_{k|k-1}$ undergo nonlinear transformation h, by

$$\mathcal{Y}_{k|k-1} = h(\chi_{k|k-1}^{(aug)}).$$

$\mathcal{Y}_{k|k-1}$ is a transformed sigma point matrix. While in this format, it cannot be compared with the state variables. However, $\mathcal{Y}_{k|k-1}$ can be used to convert the system measurements into a format, y_k, which can be found by

$$y_k = \sum_{i=0}^{2x} W_i^{(m)} \mathcal{Y}_{i,k|k-1}.$$

Now, the prediction can be compared with actual system measurements.

Then, we are able to calculate the Kalman Gain in order to determine how much to correct the model.

Unlike previous versions of the KF, in addition to calculating the covariance of the state variables, calculations is also done for the covariance of observations, P_y, and for state variables with observations, P_{xy}. When generating the covariance matrix for P_y the covariance of measurement noise is added. P_y can be found by

$$P_{y,k|k-1} = \sum_{i=0}^{2x} W_i^{(c)} (\mathcal{Y}_{i,k|k-1} - y_k)(\mathcal{Y}_{i,k|k-1} - y_k)^T + R,$$

where R is the covariance of measurement noise and $W_i^{(c)}$ are the weights calculated in step 3. Next, calculate P_{xy} by

$$P_{xy,k|k-1} = \sum_{i=0}^{2x} W_i^{(c)} (\chi_{i,k|k-1}^{(aug)} - x_k)(\mathcal{Y}_{i,k|k-1} - y_k)^T.$$

Finally, using P_y and P_{xy}, calculate the Kalman Gain, by

$$K_k = P_{xy,k|k-1}(P_{y,k|k-1})^{-1}.$$

Finally, we are able to correct the prediction and updating the covariance matrix, which will both be used in the next iteration. Similar to the UKF and the KF, the prediction of the state variables at the next time step, x_{k+1}, is given by

$$x_{k|k} = x_{k|k-1} + K_k(\hat{y}_k - y_k),$$

where $\hat{y}_k$ is system measurements. Conclude this iteration of the filter by preparing $P_{x,k}$ for the next iteration. $P_{x,k}$ can be found by

$$P_{x,k|k} = P_{x,k|k-1} - K_k(P_{y,k|k-1}){K_k}^T.$$

Table 5.1: Description of all variables in the Unscented Kalman Filter

Variables in the Unscented Kalman Filter		
Variable	Description	Dimensions
x	State variables	$x \times 1$
y	Transformed prediction	$y \times 1$
$\hat{y}$	Sytem measurements	$y \times 1$
χ	Sigma point matrix for states	$x \times (2x + 1)$
$\mathcal{Y}$	Sigma point matrix for obs	$y \times (2x + 1)$
f	Nonlinear state function	$x \times 1$
h	Nonlinear observation function	$y \times 1$
P_x	Covariance of states	$x \times x$
P_y	Covariance of observations	$y \times y$
P_{xy}	Covariance of states/obs	$x \times y$
Q	Process noise covariance	$x \times x$
R	Measurement noise covariance	$y \times y$
$W^{(m)}$	Weight for posterior mean	scalar
$W^{(c)}$	Weight for posterior covariance	scalar

Table 5.2: Description of all parameters in the Unscented Kalman Filter

Parameters in the Unscented Kalman Filter			
	Description	Bounds	Default
α	Controls spread of sigma points	$0 < \alpha \leq 1$	.001
β	Adjust sigma point weight	$\beta \geq 0$	2
κ	Sigma point weighting constant	$0 \leq \kappa \leq 3^*$	0

* Some use $\kappa = 3 - x$

5.2 Van der Pol Example

Applying the UKF to the Van der Pol oscillator will be used as a simple example to demonstrate the impacts of measurement and process noise. The Van der Pol equation, so named after its developer Balthasar Van der Pol, describes a self sustaining oscillator that create energy at small amplitudes and remove energy from large amplitudes. The Van der Pol equation describes a nonconservative oscillator (also known as a relaxation oscillator). Applications of the Van der Pol oscillators include circuits, vacuums, and modeling biological systems [**?**]. The Van der Pol oscillator is represented by a nonlinear second order differential equation:

$$\frac{d^2y}{dt^2} + \mu(y^2 - 1)\frac{dy}{dt} = 0$$

where μ is a damping coefficient, $\frac{d^2y}{dt^2}$ is acceleration, $\frac{dy}{dt}$ is velocity, and y is position. Therefore, for all $\mu < 0$, dampening occurs and the system tends to 0. The rate at which the system converges to zero is dependent on the size of μ, with larger values taking longer to converge and smaller values converging faster. If $\mu = 0$, the system becomes a simple harmonic oscillator, where motion is periodic. Lastly, if $\mu > 0$, the system enters a limit cycle, which is an isolated closed trajectory. [**?**].

The UKF can be applied to this nonlinear system to determine where the system will be at a point in time. To do so, relevant state variables include position and velocity. [1]

$$x_k = \begin{bmatrix} y \\ v \end{bmatrix}$$

Transforming the Van der Pol equation from a second order differential equation to a first order differential equation makes it easier to define f, the state function of the system. By substituting v for $\frac{dy}{dt}$, and $\frac{dv}{dt}$ for $\frac{d^2y}{dt^2}$ one can rewrite the Van der Pol equation as

$$\frac{dv}{dt} + \mu(y^2 - 1)v + y = 0.$$

From this, we get the differential equations associated with the state variables to generate the nonlinear transformation function f. For the sake of simplicity, assuming $\mu = 1$, we get

$$\dot{x}_k = \begin{bmatrix} \frac{dy}{dt} \\ \\ \frac{d^2y}{dt^2} \end{bmatrix} = \begin{bmatrix} v \\ \\ \frac{dv}{dt} \end{bmatrix} = \begin{bmatrix} v \\ \\ -1(y^2 - 1)v - y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \\ -1 & 1 - y^2 - \end{bmatrix} x_k = f.$$

[1] In Matlab, x(1) and x(2) represent position and velocity, respectively

In this particular example, the only measurement received from the system is position. Therefore, this filter is continually correcting for the position state state variable through measurement function h. Here, measurements of the system are simulated by adding noise to the position state variable. Even though there are only measurements for one state variable, we can still generate estimates of both state variables. In terms of parameters α, κ, and β, default values were used.

To simulate a Van der Pol Oscillator, an ordinary differential equation solver can be applied to state function f to generate true values of the system. Of course, systems do not perform perfectly; variations in model performance can be captured by randomly adding noise to the system.

All of this can be modeled on Matlab; all source code is from Matlab and can be referenced in Appendix A [**?**]

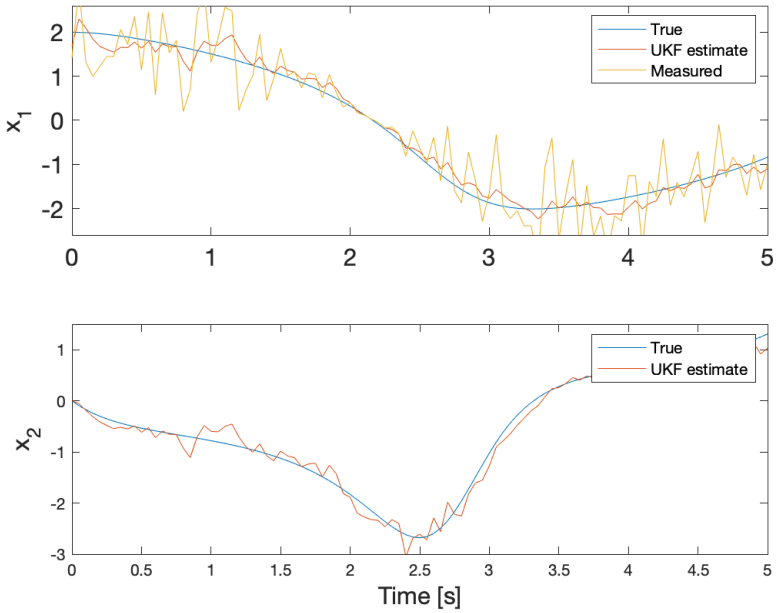


Figure 5.4: Performance of the UKF with R = 0.2, and Q = (0.02 0.1). As expected, the model converges on the true values of the system for both state variables. In this case, the only measurements we are receiving from the system are position, which is why the second state variable has no measured values.

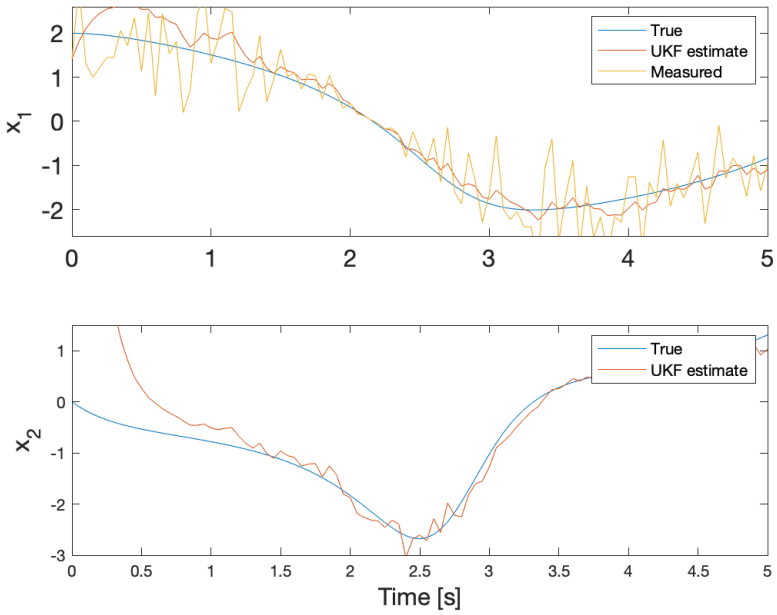
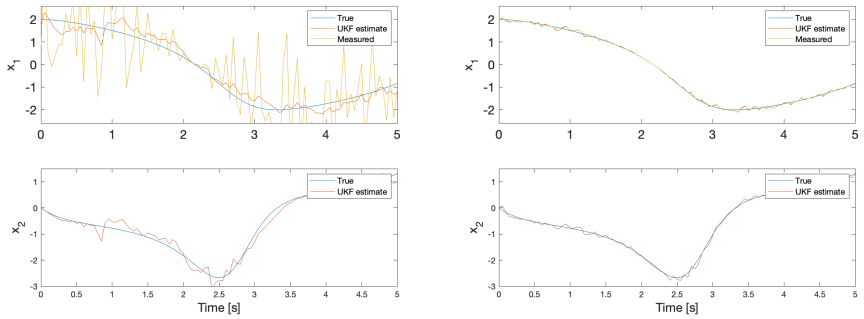
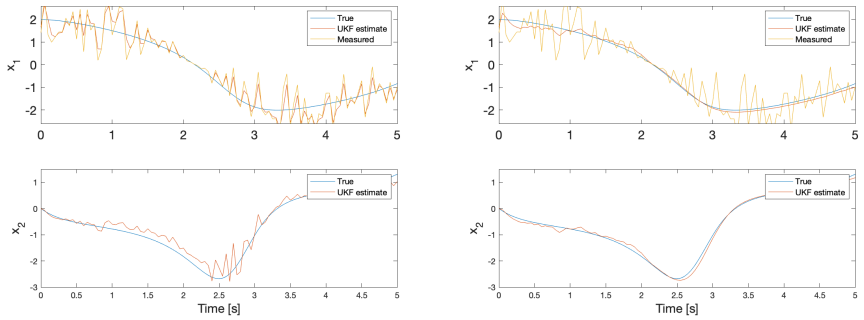


Figure 5.5: Performance of the UKF with poor initial conditions: (1, 7) instead of (2,0). Recall that inaccurate intial conditions cause convergence to take place more slowly. This seems to be the case, especially in the state variable that is not being corrected for.



(a) UKF with high measurement noise (0.9) (b) UKF with low measurement noise (0.002)

Figure 5.6: UKF on VDP oscillator with difference values of measurement noise. The model's behavior changes in response to the different values of measurement noise. Even when measurement noise is high, the UKF continues to perform well. Though, the rate of convergence appears to be slower. On the other hand, when measurement noise is low, the UKF seems to converge instantly with the measured values. The velocity state variable also quickly converges with the true value of the system.



(a) UKF with high process noise (0.9 0.8) (b) UKF with low process noise (0.0001 0.0001)

Figure 5.7: UKF on VDP oscillator with different values of process noise. Recall that process noise measures errors in the model. For the purpose of this example, process error was set to extreme high and low values. In theory, the Van der Pol equation should have small process error because it is a well used equation.

5.3 Modeling Metabolites using UKF

This is the same system explored in chapter **??**, but explored in the context of using an UKF as opposed to an EKF. The original research used an adaptation of the UKF, called the Iterative Unscented Kalman Filter (IUKF), to model the biological pathway of metabolites. Recall that this model contains four different states and 18 unknown parameters. These researchers utilized the IUKF for parameter fitting and was useful in enabling the model converge faster by resetting the covariance to re-excite the model. By not resetting the covariance at this step (as is done in the UKF), the three state variables without measurements converge significantly slower in this model. Utilizing the IUKF on this model was effective, because data regarding metabolites is highly influenced by noise, which is a factor that makes other approaches, such as regression and annealing, fail [**?**].

In the paper, researchers had access to their own data sources and used an approach that was adapted from the UKF. Though we are not using their exact dataset, we will be simulating data using the same approach as the previous example. However, this example will be following UKF algorithm, as opposed to the IUKF algorithm. By doing so, the state variables without incoming measurements converge significantly slower than the measurable states. Ultimately, the goal of this example is to demonstrate how the UKF works on higher dimensional and more complex systems and how the UKF can be utilized in parameter estimation.

Recall that the four metabolites have the following differential equations:

$$\begin{aligned}
\dot{x}_1 &= \alpha_1 x_3^{g_{13}} x_5^{g_{15}} - \beta_1 x_1^{h_{11}}, \\
\dot{x}_2 &= \alpha_2 x_1^{g_{21}} - \beta_2 x_2^{h_{22}}, \\
\dot{x}_3 &= \alpha_3 x_2^{g_{32}} - \beta_3 x_3^{h_{33}} x_4^{h_{34}}, \\
\dot{x}_4 &= \alpha_4 x_1^{g_{41}} - \beta_4 x_4^{h_{44}},
\end{aligned}$$

with 18 parameters $(\alpha_1, \ldots, \alpha_4, \beta_1, \ldots, \beta_4, g_{13}, g_{15}, g_{21}, g_{32}, g_{41}, h_{11}, h_{22}, h_{33}, h_{34}, h_{44})$. In both the original example as well as this one, sampling time will be 0.1 seconds for 5 seconds, totaling 50 UKF estimates. Recall that data is simulated on MATLAB and the model is initialized with state variable $x_0 = [4, 1, 3, 4]^T$ and the state covariance to $P_0 = .01I$.

In the original example, researchers used the following hyper-parameter values: $\epsilon = 1, \kappa = -14$. In theory, values of κ can be negative, but negative hyper-parameter values cannot be inputted into MATLAB. For this example, the hyper-parameter values used in this example are set to Matlab default values ($\alpha = 1e-3, \beta = 2, \kappa = 0$). The performance of the UKF on this system with these default hyper parameter values is shown in Figure **??**. Future work includes looking into ways to tune this hyper-parameters.

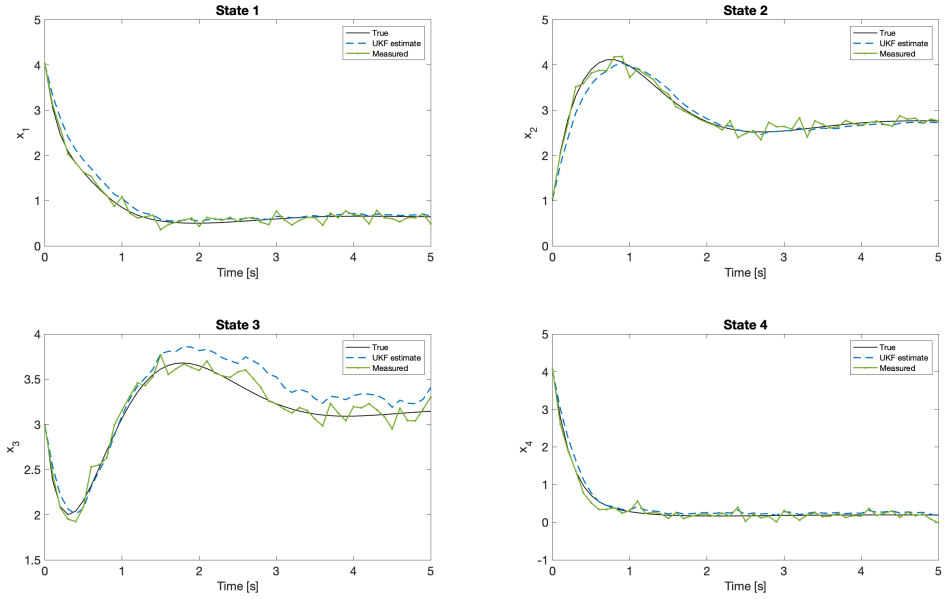


Figure 5.8: Performance of all four state variables with all four states being corrected. The model is initialized with $x_0 = [4, 1, 3, 4]^T$, which is close to the system's true values and is why the system quickly converges with the true values. The measured and true values are very close together because the system has low values of measurement noise, $R = 0.01$, and process noise, $Q = diag([0.20.10.3.4.2.3.2.1])$. The values for Q and R were not taken from the original example [**?**], which is a factor to consider when comparing results.

Parameter estimation

Similar to chapter chapter **??**, joint parameter estimation can also be applied to the UKF. This example will be using the exact same dataset for measured and true system values as the one used in the EKF version. In order to estimate parameters, $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, declare them as new states. Recall from the earlier EKF example that the system is as follows

$$\begin{aligned}
\dot{x}_1 &= x_5 x_3^{g_{13}} x_5^{g_{15}} - \beta_1 x_1^{h_{11}}, \\
\dot{x}_2 &= x_6 x_1^{g_{21}} - \beta_2 x_2^{h_{22}}, \\
\dot{x}_3 &= x_7 x_2^{g_{32}} - \beta_3 x_3^{h_{33}} x_4^{h_{34}}, \\
\dot{x}_4 &= x_8 x_1^{g_{41}} - \beta_4 x_4^{h_{44}} \\
\dot{x}_5 &= \dot{x}_6 = \dot{x}_7 = \dot{x}_8 = 0.
\end{aligned}$$

The system is initialized with the same values as the EKF four parameter example, with the initial state being $x_0 = [4, 1, 3, 4, 20, 8, 3, 2]^T$ and the state covariance being $P_0 = .01I$. Also, the values of Q and R also remain the same. The results of the UKF on this 8 state system is illustrated in **??**. Compared with the EKF results, the UKF seems to do poorer for State 3. However, in terms of parameter estimation, the EKF and UKF version seem to have the exact same performance, as shown in **??**. These results may possibly be adjusted by changing hyper parameter values and can be explored in future work.

Parameter Values				
Parameters	True	EKF estimate	UKF estimate	IUKF estimate
α_1	20.0	20.0125	20.0125	19.9744
α_2	8.0	7.9819	7.9819	8.0017
α_3	3.0	2.9818	2.9818	3.0015
α_4	2.0	1.9614	1.9614	2.0016

Table 5.3: This table shows the true values of the parameters, the final EKF prediction of the parameters (as a reference to compare filter performance), the final UKF prediction of parameters, and the final IUKF prediction of the parameters. Here, the term final is being used to denote the performance of the filter after 50 time-steps. Interestingly, the performance of the EKF is the exact same as the performance of the UKF. By changing the hyper-parameters for the UKF, the model may be able to provide more accurate results.

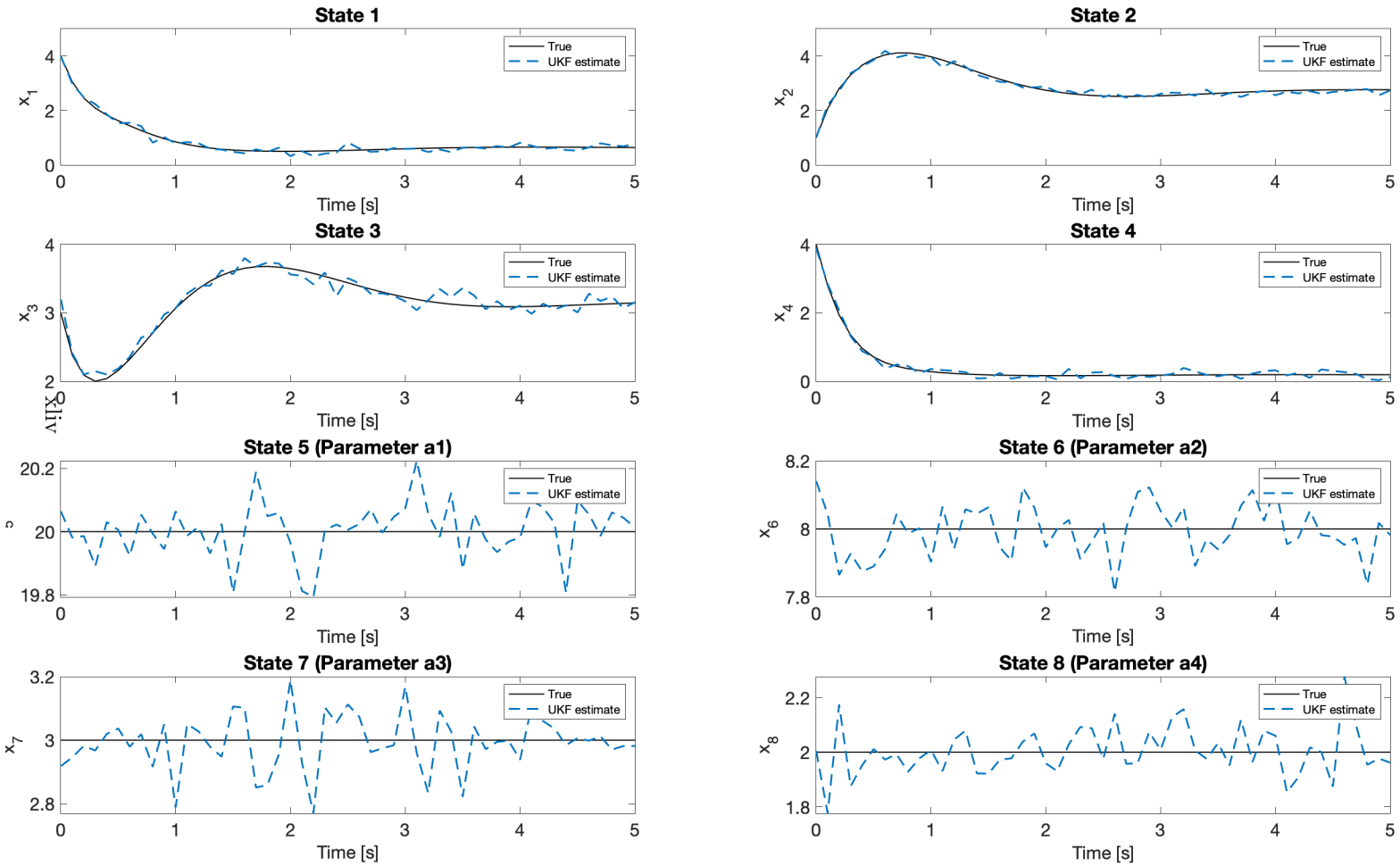

State 1
State 2
State 3
State 4
State 5 (Parameter a1)
State 6 (Parameter a2)
State 7 (Parameter a3)
State 8 (Parameter a4)
True
UKF estimate
Time [s]

Figure 5.9: ds

Residuals, also known as the innovation, are one way to access the model's performance. Recall that the residual is the difference between the actual measurement values and the predicted measurement values. Generally, a strong residual graph has

- a symmetrical distribution that is clustered toward the center,
- values that are close to 0,
- a random or unclear pattern.

Figure **??** is the residual graph of the system where the EKF is applied to the four states and four parameters, $\alpha_1, \ldots, \alpha_4$. While this residual graph has values close to 0 and a symmetrical distribution, the pattern is similar to the behavior of the states. Therefore, though UKF can be applied to this nonlinear system, performance can be further optimized.

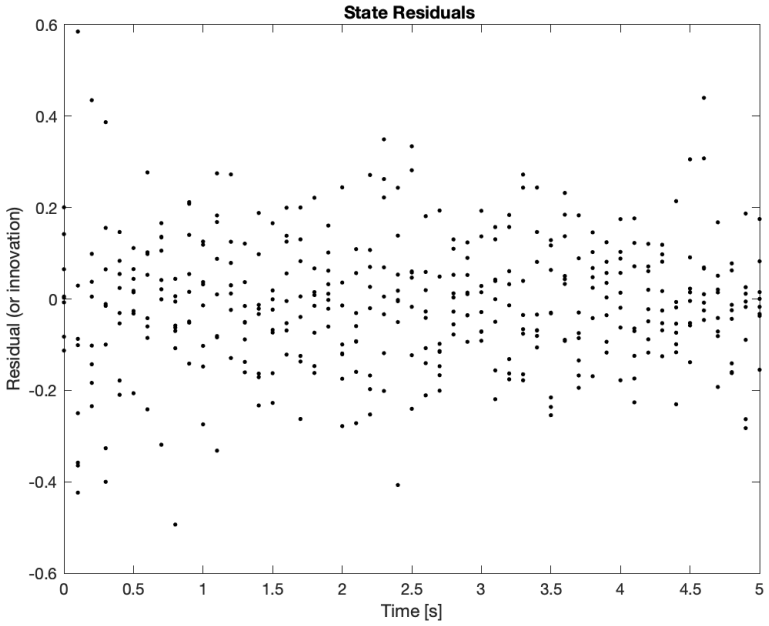


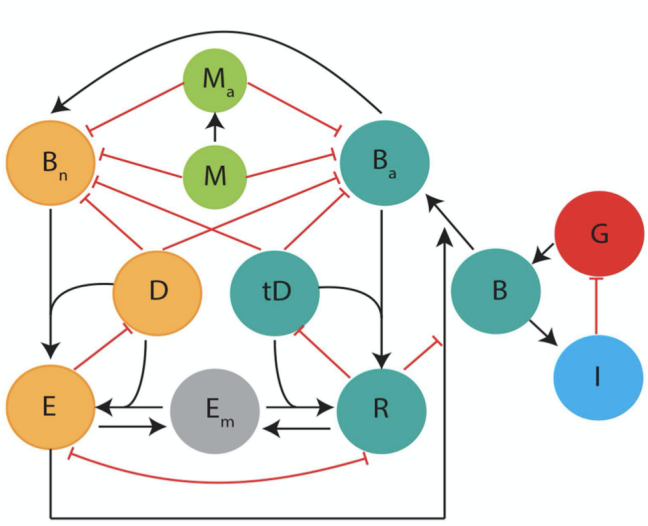
Figure 5.10: Built in matlab, four state correction and 4 param

Future work includes using the UKF for parameter fitting. One approach would be to use a technique similar to the one employed in chapter 3.2. Alternatively, another non-linear Kalman Filter technique, known as the Dual Unscented Kalman Filter, may also be a promising solution. Either of these two methods will likely be better than the method used in chapter 3.2, because calculating the exponential matrix and jacobian of the system is very computationally expensive and can be avoided with the UKF.

Chapter 6

Future Work

This thesis was inspired by previous work that involved modeling a system that forecasted glucose in real time in patients with Type 2 Diabetes. Researchers utilized the Dual Unscented Kalman Filter for simultaneous state and parameter estimation [**?**]. The goal of this thesis was to create a similar model for Type 1 Diabetes (T1D). In previous work by [**?**], a 12 state mathematical model for T1D was developed. The relationship between these 12 states are shown in Figure **??**



Variable	Description
B	Healthy β-cell population
Ba	Apoptotic β-cell population
Bn	Necrotic β-cell population
D	Immunogenic DC population
tD	Tolerogenic DC population
E	Effector T cell population
Em	Memory T cell population
R	Regulatory T cell population
Ma	Active macrophage population
M	Macrophage populations
G	Glucose levels
I	Insulin levels

Figure 6.1: This image is taken from [**?**] and provides a visualization on the relationships between the 12 states in this Type 1 Diabetes system.

The 12 state system is given as

$$\frac{d}{dt}G = R_0 - (G_0 + S_I I)G$$
$$\frac{d}{dt}I = \sigma_1 K_1(G, G_1)B - \delta_1 I$$
$$\frac{d}{dt}M = J + (k+b)M_a - cM - f_M M B_a - f_M M B_n - e_1 M(M + M_a)$$
$$\frac{d}{dt}M_a = f_M M B_a + f_M M B_n - kM_a - e_2 M_a(M + M_a)$$
$$\frac{d}{dt}B = \alpha_B K_1(G)B - \delta_B B - \eta_e(t)K_2(E,R)B - W(B,t)$$
$$\frac{d}{dt}B_a = \tilde{\delta_B} + \tilde{\eta_e}(t)K_2(E,R)B + \tilde{W}(B,t)$$
$$- dB_a - f_M M B_a - f_{M_a} B_a - f^{tD}(D_{SS} - D)B_a - f_D D B_a$$
$$\frac{d}{dt}B_n = dB_a - f_M M B_n - f_{M_a} M_a B_n - F_{Td}(D_{SS} - D)B_n - f_D D B_n$$
$$\frac{d}{dt}D = f_{tD} B_n (D_{SS} - D - tD) + f_{tD} B_n tD - b_{DE} E D - \mu_D D$$
$$\frac{d}{dt}tD = f_{tD} B_a (D_{SS} - D - tD) - f_{tD} B_n tD - b_{IR} R tD - \mu_D tD$$
$$\frac{d}{dt}E = a_E(T_{naive} - E) + b_p \frac{DE}{\theta_D + D} - r_{am} E + b_E D E_m - \mu_E E R$$
$$\frac{d}{dt}R = a_R(T_{naive} - R) + b_p \frac{tDR}{\theta_D + tD} - r_{am} R + b_R tD E m - \mu_R E R,$$
$$\frac{d}{dt}Em = r_{am}(E + R) - (a_{Em} + b_E D + b_R tD)eM$$

Future direction includes utilizing parameter estimation techniques from this thesis and applying them to this system. Since this system has several more dimensions and is more complex than other systems considered in this thesis, results using the UKF will likely be better than that of the EKF. In addition, further study of alternative parameter estimation techniques, especially the Dual Unscented Kalman Filter, can be beneficial.

Bibliography

[1] David Albers, Matthew Levine, Andrew Stuart, Lena Mamykina, Bruce Gluckman, and George Hripcsak. Mechanistic machine learning: How data assimilation leverages physiologic knowledge using bayesian inference to forecast the future, infer the present, and phenotype. *Journal of the American Medical Informatics Association : JAMIA*, 25:1392–1401, 10 2018.

[2] Yi Cao. Learning the extended kalman filter - file exchange - matlab central, Jan 2008.

[3] J. Gove and DY Hollinger. Application of a dual unscented kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. *Journal of Geophysical Research*, 111:21 PP.–21 PP., 04 2006.

[4] Remembering rudolf e. kalman (1930 – 2016), Jul 2016.

[5] Professor rudolf emil kalman.

[6] Shuichi Kinoshita. Van der pol oscillator, 2013.

[7] David Kohanbash. Kalman filtering – a practical implementation guide (with code!), Jan 2014.

[8] Chot Hun Lim, Lee Yeng Ong, Tien Sze Lim, and Voon Chet Koo. Kalman filtering and its real-time applications. *Real-time Systems*, Jun 2016.

[9] Nader Meskin, Hazem Nounou, Mohamed Nounou, and Aniruddha Datta. Parameter estimation of biological phenomena: An unscented kalman filter approach. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, 10:537–43, 03 2013.

[10] Matthew Rhudy, Roger Salguero, and Keaton Holappa. A kalman filtering tutorial for undergraduate students. *International Journal of Computer Science & Engineering Survey*, 08:01–18, 02 2017.

[11] Esra Saatci and Aydin Akan. *Dual Unscented Kalman Filter and Its Applications to Respiratory System Modeling*, page 4. intechopen, 04 2009.

[12] Blerta Shtylla, Marissa Gee, An Do, Shahrokh Shabahang, Leif Eldevik, and Lisette de Pillis. A mathematical model for dc vaccine treatment of type i diabetes. *Frontiers in Physiology*, 10:1107, 2019.

[13] Eric Wan and Ronell Merwe. The unscented kalman filter for nonlinear estimation. In *The Unscented Kalman Filter for Nonlinear Estimation*, volume 153-158, pages 153 – 158, 02 2000.

[14] Eric A. Wan and Rudolph Van Der Merwe. The unscented kalman filter. In *Kalman Filtering and Neural Networks*, pages 221–280. Wiley, 2001.

[15] Eric W Weisstein. Van der pol equation, Dec 2019.

Appendix A

Van der Pol Example

Matlab Code for VDP example

Appendix B

EKF Metabolites Example

EKF Matlab code

Appendix C

UKF Metabolites Example code

UKF Matlab code