AN EXPLORATION OF STOCHASTIC MODELS

by

JOSHUA GROSS

B.S., Kansas State University, 2010

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Major Professor
Nathan Albin

Copyright

Abstract

The term stochastic is defined as having a random probability distribution or pattern that may be analyzed statistically but may not be predicted precisely. A stochastic model attempts to estimate outcomes while allowing a random variation in one or more inputs over time. These models are used across a number of fields from gene expression in biology, to stock, asset, and insurance analysis in finance. In this thesis, we will build up the basic probability theory required to make an "optimal estimate", as well as construct the stochastic integral. This information will then allow us to introduce stochastic differential equations, along with our overall model. We will conclude with the "optimal estimator", the Kalman Filter, along with an example of its application.

Table of Contents

Table of Contents iv

List of Figures v

Acknowledgements v

Dedication vi

1 Introduction 1

1.1 Linear Dynamics . . . 2

1.2 Linear Control . . . 4

1.3 Indirect Observation . . . 6

1.4 Incorporating Uncertainty . . . 7

2 Probability Theory 9

2.1 Random Variables and Their Properties . . . 9

2.2 Gaussian Random Variables . . . 13

2.3 Stochastic Processes . . . 16

2.4 White Noise and Brownian Motion . . . 19

2.5 Stochastic Integrals . . . 26

2.6 Stochastic Differential Equations . . . 33

2.7 Discretizing the System . . . 41

3 Optimal Estimation 43

3.1 The Overall Model 43
3.2 The Kalman Filter Algorithm 58

4 A Oscillatory Example **61**

5 Conclusion **65**

Bibliography **67**

A The Fundamental Theorem for Linear Systems **68**

B Central Limit Theorem **70**

List of Figures

2.1 Three Samples of a Brownian Process 22

3.1 The Kalman Filter Algorithm (c) Petteri Aimonen (http://en.wikipedia.org/wiki/File:Basic_concept_of_Kalman_filtering.svg). Used under Creative Commons license CC0 1.0. 59

4.1 Unmeasured Estimate vs. Kalman Filter 62

4.2 Unmeasured Estimate vs. Kalman Filter Zoomed 63

4.3 Noisy Orbit of the System 64

Acknowledgments

I would like to thank my major professor Dr. Nathan Albin for all of his time, patience, enthusiasm, and guidance throughout my time at Kansas State, and the preparation of this report. I also want to thank Dr. Hrant Hakobyan and Dr. Virginia Naibo who were on my committee, and expanded my love and knowledge of mathematics throughout my undergraduate and graduate careers.

I'd also like to thank my friends and family for their love and support. My years here at Kansas State would have lacked in excitement and adventure if not for their presence. Finally, a special thanks to Katie Adams. I couldn't have done this without your love, support, inspiration, and patience. Thank you for everything.

Dedication

TO MY FAMILY

WITH LOVE

Chapter 1

Introduction

A cornerstone in the development of modern technology and reasoning has been the ability to reduce real world systems into mathematical models. Through deterministic system theory engineers, scientists and analysts have developed an abundance of material to study these models and gain insight into the behavior of the real world systems. However, there are several shortcomings present in deterministic system theory that help motivate the development of stochastic models.

The first, is that no mathematical model is perfect. Through data collection, empirical evidence, and physical laws, every model must attempt to reduce the real world system down to the key components that drive it. While this may still leave a fair representation of the system, it results in a number of components unaccounted for. While they might not make large contributions in a visible way, they are still present, and thus result in numerous sources of uncertainty in the model.

A second issue is that while controls can be incorporated into the model, they are subject to disturbances which can neither be controlled or modeled deterministically. Throughout Chapter 1, a model of interest will be a simple predator-prey model, where one method of control is the number of hunting licenses provided by the Fish and Wildlife Service. This number is based upon the number of deer beyond the carrying capacity of the ecosystem.

However, there is no guarantee that hunters will kill that number of deer for a multitude of reasons. As such, the control is just another guess at what is going to benefit the system best.

Finally, measurements of a system cannot provide perfect and complete data. If a computer sensor is used, it is subject to interference from physical conditions, or it might simply be too expensive to measure to a certain degree of specificity. Similarly, referencing the predator-prey model again, measurements on populations are approximations found by counting the number of sightings of that animal in a particular area, and extrapolating to account for the entire region.

It is clear from these reasons that other methods are required to take into account the issues of uncertainty. However, before presenting the model that will be the focus of this report, it is important to see how it was reached and what questions *it* will pose.

Lastly, it should be noted that throughout this report the main reference source was "Stochastic Models, Estimation and Control" by John Maybeck[3].

1.1 Linear Dynamics

When modeling real world systems, regardless of field, it is inevitable that some quantities of interest will vary with respect to changes in others. Differential equations are ideal for modeling such situations, and any student who seeks to pursue a major in engineering, physics or mathematics is guaranteed to spend *some* time studying the applications and solutions of linear systems of ordinary differential equations. Systems of linear ODE's are the foundation of the stochastic model built in this report. This section seeks to present the basic form of such equations and the general solutions to those systems.

The first form that is most commonly worked with is given by $\dot{\mathbf{x}}(t) = \mathbf{F}(t)\,\mathbf{x}(t)$ where $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{F}(t)$ is an $n \times n$ matrix of piecewise continuous functions, and $\dot{\mathbf{x}}(t) = [\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_n}{dt}]^{\mathrm{T}}$ along with some initial condition $\mathbf{x}(0) = \mathbf{x}_0$. If $\mathbf{F}(t)$ is a time-

invariant matrix, then it is easily shown that the solution to the system is

$$\mathbf{x}(t) = e^{\mathbf{F}\,t}\,\mathbf{x}_0$$

where $e^{\mathbf{F}\,t}$ is an $n \times n$ matrix function defined by its Taylor series. The matrix function $e^{\mathbf{F}\,t}$ is known as the *state transition matrix*.

If $\mathbf{F}(t)$ is a time-varying matrix of piecewise continuous functions, then given the same assumptions as before, the solution to the system is of the form

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\,\mathbf{x}_0$$

where $\mathbf{\Phi}(t, t_0)$ is the state transition matrix, and has the following properties

1. $\dot{\mathbf{\Phi}}(t, t_0) = \mathbf{F}(t)\,\mathbf{\Phi}(t, t_0)$
2. $\mathbf{\Phi}(t_0, t_0) = \mathbf{I}$

It is quickly confirmed that this solution satisfies the system through direct computation.

$$\begin{aligned}
\frac{d}{dt}[\mathbf{x}(t)] &= \frac{d}{dt}[\mathbf{\Phi}(t, t_0)\,\mathbf{x}_0] \\
\dot{\mathbf{x}}(t) &= \mathbf{F}(t)[\mathbf{\Phi}(t, t_0)\,\mathbf{x}_0] \\
\dot{\mathbf{x}}(t) &= \mathbf{F}(t)\,\mathbf{x}(t)
\end{aligned}$$

The existence and uniqueness of the solution, and thus the state transition matrix, is guaranteed by the Fundamental Theorem of Linear Systems of Ordinary Differential Equations (found in the appendix) [1].

A good example for this model is a natural ecosystem. Each of the variables in $\mathbf{x}(t)$ represent the population of some plant or animal, where each has some level of dependence on at least one other with no outside involvement from mankind. For example, in the Midwest,

a common predator-prey relationship of importance is that between rabbits, coyotes, deer, and (prior to the arrival of man) mountain lions. In this case both rabbits and deer are prey for coyotes and mountain lions, though at varying degrees. This system and solution allow the modeler to conduct a full analysis of how it will behave from the beginning of time to the end, but in a way, is the equivalent of winding a clock and letting it go. It will continue on and on, but lacks any further input or control from the outside world, which is impractical for the present day, as there is constant involvement from the presence of mankind.

1.2 Linear Control

Consider again the model describing the ecosystem between rabbits, coyotes, deer, and mountain lions. As European settlers arrived in the Midwest, mountain lions were pushed out due to over hunting, and deer lost their natural predator. As the understanding of ecosystems evolved, it was clear that without a natural predator the deer population would overeat its food source and begin wiping itself out through starvation. To combat this, a certain number of deer are allowed to be hunted each year to keep population levels in check. To model the effects of this with the present model, a control vector $\mathbf{u}(t)$ and a design matrix specific to the model $B(t)$ are included in the previous model. In the context of this model, it can be thought of as a periodic function which applies some negative effect upon the deer population during certain seasons. With all of the same definitions and assumptions from Section 1.1, the model is then the sum of the two vectors

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\,\mathbf{x}(t) + B(t)\,\mathbf{u}(t) \qquad \mathbf{x}(0) = \mathbf{x}_0$$

where $[B(t)\,\mathbf{u}(t)]$ is a piecewise-continuous vector valued function.

The linear time-varying system of differential equations defined above is commonly called a *non-homogeneous linear system*. The solution to the system is very similar to that of

Section 1.1, and is written

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\,\mathbf{x}_0 + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau) B(\tau)\,\mathbf{u}(\tau) d\tau$$

It can be shown through direct computation similar to that of Section 1.1 that this solution satisfies the system (using Lebniz' Rule on the integral)

$$\begin{aligned}
\frac{d}{dt}[\mathbf{x}(t)] &= \frac{d}{dt}[\mathbf{\Phi}(t, t_0)\,\mathbf{x}_0 + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau) B(\tau)\,\mathbf{u}(\tau) d\tau] \\
&= \mathbf{F}(t)[\mathbf{\Phi}(t, t_0)\,\mathbf{x}_0] + B(t)\,\mathbf{u}(t) + \int_{t_0}^{t} \mathbf{F}(t)\,\mathbf{\Phi}(t, \tau) B(\tau)\,\mathbf{u}(\tau) d\tau] \\
&= \mathbf{F}(t)[\mathbf{\Phi}(t, t_0)\,\mathbf{x}_0 + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau) B(\tau)\,\mathbf{u}(\tau) d\tau] + B(t)\,\mathbf{u}(t) \\
&= \mathbf{F}(t)\,\mathbf{x}(t) + B(t)\,\mathbf{u}(t)
\end{aligned}$$

A common question asked at this point, is "Given an initial condition $\mathbf{x}(0) = \mathbf{x}_0$ and a vector $\mathbf{x}_f \in \mathbb{R}^n$ can $\mathbf{u}(t)$ be chosen such that $\mathbf{x}(t) = \mathbf{x}_f$ for some finite $t \in T$"? If so, then the system is called *controllable*. Relating back to the predator-prey model, it is clear why this would be important. If the Kansas Fish and Wildlife Service wishes to keep several populations of different species within acceptable ranges, a controllable system could allow them to analyze $\mathbf{u}(t)$ to determine the actions required to keep things running smoothly.

The biggest issue with the present model is that no measurement occurs throughout the process to determine if the model is still accurately representing the system at a given time t. In the real world, biologists will trek out into the field to get an approximate count of the current population of different plants an animals, and can update their procedures from there. The next section seeks to correct this.

1.3 Indirect Observation

To maintain an accurate model of a system that changes with time, it is important to be able to make observations and measurements with which to compare the current state. In the present model, these observations are incorporated through a discrete-time measurement equation $\mathbf{z}(t_i)$.

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\,\mathbf{x}(t) + B(t)\,\mathbf{u}(t) \qquad \mathbf{x}(0) = \mathbf{x}_0$$
$$\mathbf{z}(t_i) = \mathbf{H}(t_i)\,\mathbf{x}(t_i)$$

Here, $\mathbf{H}(t_i)$ can be thought of as interference between what can be accurately measured, and the true state of the system. In the predator-prey model, it is infeasible to attempt to count every single animal in an ecosystem. However, approximations can be made by counting the number of animals seen in a specific area, then scaling that to the size of the region in question. One application of discrete-time measurements is what is commonly referred to as a *predictor-corrector* model, where the model is allowed to run up until a measurement time t_i, then a weighted correction is made using the incomplete data from $\mathbf{z}(t_i)$.

A second question that is commonly asked in control models with indirect observation is about *observability*, which asks "If we have complete data on the measurements $\mathbf{z}(t)$ and the control function $\mathbf{u}(t)$ for some period of time T, is it possible to deduce the behavior of $\mathbf{x}(t)$ throughout T?". It is easy to see why this is important to a group like the Fish and Wildlife Service, since aside from the assumed relations in the model between the animals in the ecosystem, the only information provided are the measurements, and the control provided through exterior forces (hunting or releasing additional animals to the system).

Both controllability and observability are very important questions that are studied in depth throughout control and optimization theory. There are extensions of these ideas to stochastic models, though additional work is needed than will be presented in this report,

and is thus, a topic for future research.

1.4 Incorporating Uncertainty

As mentioned before, the objective of a mathematical model would be to generate an adequate, tractable representation of the behavior of all outputs of interest from the real physical system. The term adequate is used here, because not all aspects of a real world system can be included in a model, while still obtaining something workable. A number of small factors play into little changes within the system that are just unreasonable to make variables for. In the predator-prey model, the populations are in constant flux from weather changes, natural disasters, unexpected interference from man, etc. One way all of these effects can be modeled is through the central limit theorem (found in the appendix), which says that the sum of numerous small random effects can be modeled as one large effect with the properties of a normal distribution (to be discussed in chapter 2). The model from Section 1.3 can be extended to include such uncertainty, and is formally written

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\,\mathbf{x}(t) + B(t)\,\mathbf{u}(t) + \mathbf{n}_1(t) \qquad \mathbf{x}(0) = \mathbf{x}_0$$

$$\mathbf{z}(t_i) = \mathbf{H}(t_i)\,\mathbf{x}(t_i) + \mathbf{n}_2(t_i)$$

Focusing on the predator-prey model, the noise in the differential equation can represent the number of small effects which cause populations to fluctuate, while the noise in the measurement equation represents human error when approximating the population.

As mentioned at the end of Section 1.3, random noise significantly impacts the modeler's ability to determine controllability and observability, and as such, different definitions and assumptions are required for systems incorporating these random effects. While this is not the focus of the report, it is something worth thinking about as ideas are presented.

With noise now present within the model, there are several important questions to ask

that this report seeks to answer.

1. How does one model noise?

   The function $\mathbf{n}(t)$ will appear as instantaneous random perturbations. This presents problems with the current model which is using the usual Riemann integrals. Chapter 2 will present methods for understanding what the object $\mathbf{n}(t)$ is, as well as how to integrate such a function.

2. What type of object is $\mathbf{x}(t)$?

   With the presence of a random variable, the system $\mathbf{x}(t)$ is no longer deterministic. Well known measurements in probability theory will be extended to the stochastic processes. These new methods will be used to analyze the system through the mean, variance, etc. of these combinations of random variables and deterministic functions.

3. How is this new type of problem solved?

   As mentioned in question 1, Riemann integrals are no longer feasible for all portions of the differential equation. In fact, it's not even completely truthful to simply call $\dot{\mathbf{x}}(t)$ a differential equation. After determining a way to integrate stochastic processes, it will be possible to rewrite the model in Section 1.3 in a more workable form as a stochastic differential equation.

4. How can measurements be incorporated to this new problem?

   The final goal of this report is to present an "optimal guess" similar to the predictor-corrector models of linear control theory. This filter will attempt to correct for the random error within the model, while improving the guess based on the measurements available at the present time, conditioned on all previous information gathered.

Chapter 2

Probability Theory

2.1 Random Variables and Their Properties

While there are many probability spaces that the following definitions may be applied to, this report focuses primarily on the applications of Gaussian variables. Thus, the space that will be used in all applications is the triplet (Ω, $\mathcal{F}$, P), where Ω is a sample space, $\mathcal{F}$ is a σ-algebra, but more specifically the Borel σ-algebra, and P is a probability measure, mapping $\mathcal{F}$ into the interval [0,1].

Definition 1. A *scalar random variable* $x(\cdot)$ is a measurable function $x : \Omega \to [0, 1]$.

This definition can be extended to a *vector random variable* $\mathbf{x}(\cdot) = [x_1(\cdot), x_2(\cdot), \ldots, x_n(\cdot)]$, which is simply a vector where each entry is a scalar random variable.

From this definition, it is clear that a topic of interest for random variables is "When will a random variable $\mathbf{x}(\omega)$ realize a value within some set A?". One way this is answered is through the probability measure P. However, additional information can be gained by constructing a function for the particular random variable in question.

Definition 2. The *probability distribution function* $F_{\mathbf{x}} : \mathbb{R}^n \to [0, 1]$ determines the probability that a realization of the random variable $\mathbf{x}$ will belong to a set determined by its

dummy variable ξ, and is related to the probability measure by

$$\begin{aligned} F_{\mathbf{x}}(\xi) &= F_{x_1,x_2,\ldots,x_n}(\xi_1,\xi_2,\ldots,\xi_n) \\ &= \mathrm{P}(\omega : x_1(\omega) \leq \xi_1, x_2(\omega) \leq \xi_2, \ldots, x_n(\omega) \leq \xi_n) \end{aligned}$$

With the current definition of the distribution function, the probability of a realization of $\mathbf{x}$ can be constructed through sums and differences of the distribution function. A second probability function that will be applied throughout the remainder of the report can be derived from the probability distribution function by attempting to differentiate it.

Definition 3. If a scalar function $f_{\mathbf{x}}(\cdot)$ exists such that

$$F_{\mathbf{x}}(\xi_1,\xi_2,\ldots,\xi_n) = \int_{-\infty}^{\xi_1}\int_{-\infty}^{\xi_2}\cdots\int_{-\infty}^{\xi_n} f_{\mathbf{x}}(\rho_1,\rho_2,\ldots,\rho_n)d\rho_1,d\rho_2,\ldots,d\rho_n$$

Then the function $f_{\mathbf{x}}(\cdot)$ is called the *probability density function* of the random variable $\mathbf{x}(\cdot)$. It is a vector-valued function from the realizations of random variables in $\mathbb{R}^n$ to the interval $[0,1]$ indicating the probability of that event occurring.
It should also be noted that in general, the probability density function is not guaranteed to exist (especially a continuous one). However, for a Gaussian random variable it is guaranteed to exist *and* be continuous. Therefore, as Gaussian random variables are a major focus of this report, the probability density function will always be assumed to exist.

From the above definition of the probability density function, it can be seen that the integral of the probability density function over a region $A \in \mathbb{R}^n$ results in the probability of the realization $\mathbf{x}(\omega)$ belonging to A, and is written

$$\begin{aligned} \mathrm{P}(\{\omega : \mathbf{x}(\omega) \in A\}) &= \int_A f_{\mathbf{x}}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ &= \int_A \cdots \int_A f_{x_1,x_2,\ldots,x_n}(\xi_1,\ldots,\xi_n)d\xi_1 d\xi_2 \ldots d\xi_n \end{aligned}$$

The middle term in the above equation is comfortable for scalar random variables or to simplify notation with vector random variables. However, the right term is used to address each individual piece in the vector random variable. The integrand on the right hand side is called the *joint density function* of the variables $x_1(\cdot), x_2(\cdot), \ldots, x_n(\cdot)$. (It can be noted that $\int_{\mathbb{R}^n} f_{\mathbf{x}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = 1$).

Definition 4. Two random variables $\mathbf{x}$ and $\mathbf{y}$ are *independent* if their joint distribution function is equal to the product of their distribution functions

$$F_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\gamma}) = F_{\mathbf{x}}(\boldsymbol{\xi}) F_{\mathbf{y}}(\boldsymbol{\gamma})$$

Similarily, if they exist, the definition extends to the joint density function

$$f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\gamma}) = f_{\mathbf{x}}(\boldsymbol{\xi}) f_{\mathbf{y}}(\boldsymbol{\gamma})$$

Given a random variable, one often likes to know which values it will most likely assume. In particular, if there are two random variables $\mathbf{x}$ and $\mathbf{y}$ which are connected in some way, it is possible to discern additional information about one variable given some knowledge of the other. For example, if it is known that the random variable $\mathbf{y} : \Omega \longmapsto \mathbb{R}^m$, will be mapped into a particular set $B \in \mathbb{R}^m$ (due to some outside knowledge), then the estimate of $\mathbf{x}$ can be conditioned on what $\mathbf{y}$ must be.

Definition 5. The *conditional density* for $\mathbf{x}$, conditioned on the fact that $\mathbf{y}(\omega) \in B$ (using Bayes' Rule) is written

$$f_{\mathbf{x}|\mathbf{y}}(\xi \mid \mathbf{y} \in B) = \frac{\int_B f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\gamma}) d\boldsymbol{\gamma}}{\int_B f_{\mathbf{y}}(\boldsymbol{\rho}) d\boldsymbol{\rho}}$$

It can be shown that the conditional density is indeed a probability density function through Fubini's Theorem. The conditional density can then be used to determine the probability that $\mathbf{x} \in A$ given $\mathbf{y} \in B$ by integrating the above function with respect to $\boldsymbol{\xi}$ over the interval A. If it is known that the random variable $\mathbf{y}$ will assume a particular value $\mathbf{y}(\omega) = y$, then

the above definition would appear to result in a zero in the denominator, due to $B = \{y\}$. To remedy this, consider an ϵ-ball around the realization y, $B_\epsilon = (y_1 - \epsilon, y_1 + \epsilon) \times (y_2 - \epsilon, y_2 + \epsilon) \times \cdots \times (y_m - \epsilon, y_m + \epsilon)$. Then the conditional density function for a particular realization can be defined as

$$\begin{aligned} f_{\mathbf{x}|\mathbf{y}}(\xi \mid \mathbf{y} = y) &= \lim_{\epsilon \to 0} f_{\mathbf{x}|\mathbf{y}}(\xi \mid \mathbf{y} \in B_\epsilon) \\ &= \lim_{\epsilon \to 0} \frac{\int_B f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\gamma}) d\boldsymbol{\gamma}}{\int_B f_{\mathbf{y}}(\boldsymbol{\rho}) d\boldsymbol{\rho}} \end{aligned}$$

Next, some of the measurements of random variables will be examined. Two in particular are going to be important to the model. This first is the *expectation*.

Definition 6. Given a random variable $\mathbf{x}(\cdot)$, the *expectation* (or *mean*) $\mathbf{E}[\mathbf{x}]$ of $\mathbf{x}$ is the value you would expect to get if you were to generate an infinite number of realizations and then averaged them. It is given by the equation

$$\mathbf{E}[\mathbf{x}] = \int_{-\infty}^{\infty} \rho f_{\mathbf{x}}(\rho) d\rho = \hat{\mathbf{x}}$$

Definition 7. If the random variable $\mathbf{y}$ has realized some particular value $y \in \mathbb{R}^n$, one can define the *conditional expectation* or (*conditional mean*) of $\mathbf{x}$ through the conditional density function for a particular realization $\mathbf{y}(\omega) = y$.

$$\mathbf{E}[\mathbf{x} \mid \mathbf{y}(\omega) = y] = \int_{-\infty}^{\infty} \boldsymbol{\xi} f_{\mathbf{x}|\mathbf{y}}(\boldsymbol{\xi} \mid \mathbf{y} = y) d\boldsymbol{\xi}$$

Now that it is possible to find the mean of a random variable, it is often useful to know just how much it tends to vary from that mean, and if two variables might vary together. For a scalar random variable $x(\cdot)$, the *variance* ($\mathrm{Var}(x)$) is the movement away from the mean that can generally be expected from realizations of x, and is computed by $\mathbf{E}([x - \hat{x}]^2)$. Additionally, one may wish to know how two scalar variables vary together. The *covariance* of x and y ($\mathrm{Cov}(x, y)$) can be found by $\mathbf{E}[(x - \hat{x})(y - \hat{y})]$. If $\mathrm{Cov}(x, y) = 0$, then it is said

that x and y are *uncorrelated*.

With both of these ideas defined, they can now be extended to the vector random variable case to define the second important statistic, the *covariance matrix*.

Definition 8. Given the usual vector random variable $\mathbf{x}$, the *covariance matrix* is defined as $\mathbf{P} = \mathbf{E}[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T]$. This results in the main diagonal $\mathbf{P}_{i,i} = \mathrm{Var}(x_i)$ and the off-diagonal terms $\mathbf{P}_{i,j} = \mathrm{Cov}(x_i, x_j)$. Note that if the off-diagonal terms of the covariance matrix are all zero ($\mathbf{P}$ is diagonal), then the variables are all uncorrelated. In practice, $\mathbf{P}$ is generally assumed to be positive definite. From its definition, it can be seen to be symmetric, therefore $\mathbf{P}$ will always be assumed to be symmetric positive definite.

Finally, a tool that will be useful in the following section for the analysis of Gaussian random variables, is the *characteristic function* of a random variable.

Definition 9. If $\mathbf{x}$ is a n-vector-valued random variable, its *characteristic function* $\phi_{\mathbf{x}}(\cdot)$ is defined as the scalar function of the dummy vector μ as

$$\begin{aligned}\phi_{\mathbf{x}}(\mu) &= \mathbf{E}_{\mathbf{x}}[\exp\{i\mu^{\mathrm{T}}\mathbf{x}\}] \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\{i\mu^{\mathrm{T}}\boldsymbol{\xi}\} f_{\mathbf{x}}(\boldsymbol{\xi}) d\xi_1 d\xi_2 \dots d\xi_n\end{aligned}$$

where $i = \sqrt{-1}$. There are several useful properties of the characteristic function that will be discussed in the next section after Gaussian random variables have been introduced.

2.2 Gaussian Random Variables

Due to its natural occurrence in numerous real world systems, as well as its simplicity to model, the Gaussian (or normal) random variable is of particular use to this report (along with countless other models and theorems).

Definition 10. An n-dimensional vector random variable $\mathbf{x}$ is said to be a *Gaussian (normal) random vector* if it can be described through a probability density function of the

form

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

where $\mathbf{P}$ is the symmetric positive definite covariance matrix and $|\, \mathbf{P} \,|$ is the determinant of $\mathbf{P}$. The probability density function for a scalar Gaussian random variable is given by

$$f_x(\xi) = \frac{1}{\sqrt{2\pi P}} \exp\{\frac{-1}{2P}(\xi - \hat{x})^2\}$$

where $P = \mathrm{Var}(x)$.

Definition 11. Two random variables x and y are called *jointly Gaussian* if their joint density function can be written in the form of a Gaussian density function.

One reason Gaussian random variables are so valuable, is that the probability density function is completely characterized by only the covariance matrix $\mathbf{P}$ and the mean (or expectation) of $\mathbf{x}$ (as can be seen above).

The characteristic function for a Gaussian random variable $\mathbf{x}$ with the density function shown above can be derived through direct computation from the definition, and is found to be

$$\phi_{\mathbf{x}}(\mu) = \exp\{i\boldsymbol{\mu}^{\mathrm{T}}\hat{\mathbf{x}} - \frac{1}{2}\boldsymbol{\mu}^{\mathrm{T}}\, \mathbf{P}\, \boldsymbol{\mu}\}$$

The properties of the characteristic function which are of interest are that the expectation, variance and covariance can all be determined from the partial derivatives of the characteristic function. The first is that the partial derivative of the characteristic function

with respect to the k^{th} variable results in the expectation of that random variable

$$\frac{\partial \phi_{\mathbf{x}}(\mu)}{\partial \mu_k} = i \int_{-\infty}^{\infty} \xi_k e^{i\mu^{\mathrm{T}}\boldsymbol{\xi}} f_{\mathbf{x}}(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

$$\text{Let } \mu = 0$$

$$\frac{1}{i}\frac{\partial \phi_{\mathbf{x}}(\mu)}{\partial \mu_k}\bigg|_{\mu=0} = \int_{-\infty}^{\infty} \xi_k f_{\mathbf{x}}(\boldsymbol{\xi}) d\boldsymbol{\xi}$$

$$= \mathbf{E}\{x_k\}$$

The second useful computation is that the covariance between the k^{th} and j^{th} variables can be found from the equation above by now taking the partial derivative with respect to μ_j. The computation is nearly identical.

As was stated above, this is very useful when working with Gaussian random variables, as the probability density function is completely determined by these two values. These facts make it much easier to model and seek insight into systems as well as prove some relations of Gaussian random variables. Below, the characteristic function will be utilized to prove a fact that will be used throughout the remainder of the report.

Theorem 1 (Linear transformation of Gaussian vectors are Gaussian)**.** Let $\mathbf{x}$ be a Gaussian n-vector with mean $\hat{\mathbf{x}}$ and $\mathrm{Cov}(\mathbf{x}) = \mathbf{P}_{\mathbf{xx}}$ and $\mathbf{A}$ a known $(m \times n)$ matrix with $\mathrm{null}(\mathbf{A}^{\mathrm{T}}) = \{0\}$. Then the random m-vector $\mathbf{y}$ defined by $\mathbf{y} = \mathbf{A}\mathbf{x}$ is Gaussian.

Proof. The characteristic function for $\mathbf{y}$ is given by

$$\begin{aligned}
\phi_{\mathbf{y}}(\mu) &= \mathbf{E}[e^{i\boldsymbol{u}^{\mathrm{T}}\mathbf{y}}] \\
&= \mathbf{E}[e^{i\boldsymbol{u}^{\mathrm{T}}\mathbf{A}\mathbf{x}}] \\
&= \mathbf{E}[e^{i[\mathbf{A}^{\mathrm{T}}\boldsymbol{u}]^{\mathrm{T}}\mathbf{x}}] \\
&= \phi_{\mathbf{x}}(\mathbf{A}^{\mathrm{T}}\mu)
\end{aligned}$$

because $\mathbf{x}$ is Gaussian, $\phi_{\mathbf{x}}(\mathbf{A}^{\mathrm{T}}\mu)$ can be written explicitly as

$$\begin{aligned}\phi_{\mathbf{x}}(\mathbf{A}^{\mathrm{T}}\mu) &= \exp\{i(\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu})^{\mathrm{T}}\hat{\mathbf{x}} - \tfrac{1}{2}(\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu})^{\mathrm{T}}\mathbf{P}_{\mathbf{xx}}(\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu})\} \\ &= \exp\{i\boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}\hat{\mathbf{x}} - \tfrac{1}{2}(\boldsymbol{\mu})^{\mathrm{T}}\mathbf{A}\mathbf{P}_{\mathbf{xx}}(\mathbf{A}^{\mathrm{T}}\boldsymbol{\mu})\} \\ &= \exp\{i\boldsymbol{\mu}^{\mathrm{T}}(\mathbf{A}\hat{\mathbf{x}}) - \tfrac{1}{2}\boldsymbol{\mu}^{\mathrm{T}}(\mathbf{A}\mathbf{P}_{\mathbf{xx}}\mathbf{A}^{\mathrm{T}})\boldsymbol{\mu})\}\end{aligned}$$

Here it can be seen that this is the characteristic function of a Gaussian random variable with mean $\mathbf{A}\hat{\mathbf{x}}$ and covariance $\mathbf{A}\mathbf{P}_{\mathbf{xx}}\mathbf{A}^{\mathrm{T}}$. Thus $\mathbf{y}$ is indeed Gaussian. □

Theorem 2 (Linear combinations of jointly Gaussian vectors are Gaussian)**.** Let $\mathbf{x}$ and $\mathbf{y}$ be Gaussian n and m dimension jointly Gaussian vectors with means $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ and covariances $\mathbf{P}_{\mathbf{xx}}$ and $\mathbf{P}_{\mathbf{yy}}$, respectively. Let $\mathbf{A}$ and $\mathbf{B}$ be known $(p \times n)$ and $(p \times m)$ matrices such that the matrix C defined by

$$\mathbf{C} = \left[\begin{array}{c} \mathbf{A} \\ \hline \mathbf{B} \end{array}\right]$$

has the property that $\mathrm{null}(\mathbf{C}^{\mathrm{T}}) = \{0\}$. Then the random variable $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}$ is Gaussian.

Proof. Define the random variable

$$\mathbf{w} = \left[\begin{array}{c} \mathbf{x} \\ \hline \mathbf{y} \end{array}\right]$$

which is Gaussian, and thus, characterized by mean and covariance

$$\hat{\mathbf{w}} = \left[\begin{array}{c} \hat{\mathbf{x}} \\ \hline \hat{\mathbf{y}} \end{array}\right] \qquad \mathbf{P}_{\mathbf{ww}} = \left[\begin{array}{c|c} \mathbf{P}_{\mathbf{xx}} & \mathbf{P}_{\mathbf{xy}} \\ \hline \mathbf{P}_{\mathbf{yx}} & \mathbf{P}_{\mathbf{yy}} \end{array}\right]$$

It can now be seen that $\mathbf{z} = \mathbf{C}\mathbf{w}$, and Theorem 1 can be applied to see that $\mathbf{z}$ is indeed a Gaussian random variable. □

2.3 Stochastic Processes

To study real-world phenomena in which systems evolve randomly, probabilistic models are needed rather than deterministic ones. Now that a foundation of basic probability theory

has been presented, this section will begin to look at the pieces that form these models. A stochastic process is a function of both a random variable, as well as time. A common example to keep in mind as the chapter progresses is a model of a single company's stock prices. While there are a number of major factors that will affect the overall rise and fall of the price, countless smaller factors affect it on a minute-to-minute basis that result in a large amount of uncertainty as to how it will behave next. A second example one might consider is the movement of a gas particle as it is pushed around by the countless other molecules around it. These example will be expanded upon later, but as definitions and analysis are presented in the following section, such examples are very useful for gaining insight into the definitions and measurements.

Definition 12. Let Ω be the fundamental sample space used in previous sections, and T be a subset of the real line denoting a time set of interest. Then a *stochastic process* can be defined as a real-valued function $\mathbf{x}(\cdot, \cdot)$, defined on the product space $T \times \Omega$ such that for any fixed $t \in T$, $\mathbf{x}(t, \cdot)$ is a random variable.

Recalling the definition of a random variable, this says that $\mathbf{x}(\cdot, \cdot)$ is a stochastic process if $\mathbf{x}(t, \cdot)$ is a P-measurable function $\forall\, t \in T$. Additionally, if the second argument is fixed for some $\omega \in \Omega$, $\mathbf{x}(\cdot, \omega)$ is a time function, and is called a *sample* from the stochastic process. If $T \subset \mathbb{R}$ is a sequence of the form $\{t_1, t_2, \dots\}$, then the *discrete-time* stochastic process evaluated at these times becomes a sequence of random variables $\mathbf{x}(t_1, \cdot), \mathbf{x}(t_2, \cdot), \dots$. For simplicity, this notation will often be reduced from $\mathbf{x}(t, \cdot)$ to $\mathbf{x}(t)$.

Given the definition of a stochastic process as well as the above sequence of random variables, many of the ideas and definitions from the previous section can be easily extended to the stochastic case. One idea in particular that will continue to be of great importance is the Gaussian process, which will again be completely characterized through its density function by its mean and covariance matrix.

For a finite set of the random variables obtained from the sequence shown above, the usual definition of a joint density function can be utilized and measurements from Chapter

1 can be extended to the stochastic processes.

Definition 13. The *mean value function* (or *mean*) of the stochastic process $\mathbf{x}(\cdot)$ is defined for all $t \in T$ by

$$\hat{\mathbf{x}}(t) = \mathbf{E}\{\mathbf{x}(t)\}$$

Here, the average is taken over all samples from the process.

Definition 14. To determine the spread of the values about the mean at time t, consider the *covariance matrix*, $\mathbf{P}_{xx}(\cdot)$, defined by

$$\mathbf{P}_{xx}(t) = \mathbf{E}\{[\mathbf{x}(t) - \hat{\mathbf{x}}(t)][\mathbf{x}(t) - \hat{\mathbf{x}}(t)]^{\mathrm{T}}\}$$

Similarly, given that a stochastic process evaluated at two separate times t_1, $t_2 \in T$ may produce two different random variables, one may consider their covariance through the *covariance kernel*

$$\mathbf{P}_{xx}(t_1, t_2) = \mathbf{E}\{[\mathbf{x}(t_1) - \hat{\mathbf{x}}(t_1)][\mathbf{x}(t_2) - \hat{\mathbf{x}}(t_2)]^{\mathrm{T}}\}$$

Next, several important distinctions must be made regarding the terms independent, and uncorrelated when referring to stochastic processes. A process $\mathbf{x}(\cdot, \cdot)$ is *independent (in time)* or *white* if for any choice of $N \in \mathbb{N}$ and any $t_1, t_2, \ldots, t_N \in T$, the set of random vectors $\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_N)$ are independent from one and other. That is,

$$f_{\mathbf{x}(t_1),\ldots,\mathbf{x}(t_N)}(\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_N}) = \prod_{i=1}^{N} f_{\mathbf{x}(t_i)}(\boldsymbol{\xi_i})$$

However, two processes $\mathbf{x}(\cdot, \cdot)$ and $\mathbf{y}(\cdot, \cdot)$ are said to be *independent (of each other)* if for any N $t_1, t_2, \ldots, t_N \in T$

$$\begin{aligned} &\mathrm{P}(\{\omega : \mathbf{x}(t_1, \omega) \leq \boldsymbol{\xi_1}, \ldots, \mathbf{x}(t_N, \omega) \leq \boldsymbol{\xi_N}, \mathbf{y}(t_1, \omega) \leq \boldsymbol{\rho_1}, \ldots, \mathbf{y}(t_N, \omega) \leq \boldsymbol{\rho_N}\}) \\ &= \mathrm{P}(\{\omega : \mathbf{x}(t_1, \omega) \leq \boldsymbol{\xi_1}, \ldots, \mathbf{x}(t_N, \omega) \leq \boldsymbol{\xi_N}\})\, \mathrm{P}(\{\mathbf{y}(t_1, \omega) \leq \boldsymbol{\rho_1}, \ldots, \mathbf{y}(t_N, \omega) \leq \boldsymbol{\rho_N}\}) \end{aligned}$$

A process $\mathbf{x}(\cdot, \cdot)$ is *uncorrelated (in time)* if for any choices of $t_1, t_2 \in T$ such that $(t_1 \neq t_2)$,

$$\mathbf{E}[\mathbf{x}(t_1)\,\mathbf{x}^{\mathrm{T}}(t_2)] = \mathbf{E}[\mathbf{x}(t_1)]\,\mathbf{E}[\mathbf{x}^{\mathrm{T}}(t_2)]$$
$$\text{or } \mathbf{P}_{xx}(t_1, t_2) = 0$$

While two processes $\mathbf{x}(\cdot,\cdot)$ and $\mathbf{y}(\cdot,\cdot)$ are *uncorrelated* if for all $t_1, t_2 \in T$

$$\mathbf{E}[\mathbf{x}(t_1)\,\mathbf{y}^{\mathrm{T}}(t_2)] = \mathbf{E}[\mathbf{x}(t_1)]\,\mathbf{E}[\mathbf{y}^{\mathrm{T}}(t_2)]$$
$$\text{or } \mathbf{P}_{xy}(t_1, t_2) = 0$$

As was mentioned before, Gaussian processes are going to be of particular interest because again, for any given fixed $t \in T$, the density function will be completely determined by the mean and covariance matrix.

Definition 15. A process $\mathbf{x}(\cdot,\cdot)$ is a *Gaussian process* if for any choice of $N \in \mathbb{N}$ and $t_1, t_2, \ldots, t_N \in T$ the finite joint distribution function for $\mathbf{x}(t_1,\cdot), \mathbf{x}(t_2,\cdot), \ldots, \mathbf{x}(t_N,\cdot)$ is Gaussian.

2.4 White Noise and Brownian Motion

White Gaussian noise and Brownian Motion are two of the fundamental pieces used to mimic random processes found in nature. Due to the central limit theorem, which states that a large number of random variables added together tend to a Gaussian distribution, Gaussian white noise is ideal for modeling a number of tiny effects which may play a part within the noise of a system. Brownian Motion was first founded upon the jiggling motion exhibited by a particle suspended in a fluid, which was eventually attributed to the countless collisions of molecules in the fluid. As analysis continued, it was found that Brownian motion could describe all kinds of events that have *diffusion* aspects. This section seeks to define a number of properties required to describe and define Brownian motion.

As was discussed in Chapter 1, the overall goal is to present a model which incorporates the noise added by the numerous small effects present in real world systems, and then a method to accurately make a prediction with this noise present. However, as mentioned

before, the central limit theorem allows these effects to be summed into one *white Gaussian process*, as is shown below. Recall the model which first incorporated noise:

$$\dot{\mathbf{x}}(t) = F(t)\,\mathbf{x}(t) + B(t)\,\mathbf{u}(t) + G(t)\,\mathbf{n}_1(t)$$

Given that the noise is a white Gaussian process, one could apply the previous theorems regarding Gaussian random variables with the hope of showing the density function for the model itself is Gaussian as well. If this is accomplished, then the system can be described completely by its mean and covariance matrix, and analysis will be greatly simplified. By applying the insights from deterministic system theory from Chapter 1, a proposed solution to the system could be written

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\,\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau) B(\tau)\,\mathbf{u}(\tau) d\tau + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau) G(\tau)\,\mathbf{n}_1(\tau) d\tau$$

However, there is currently no way of interpreting the final integral (which contains a random process with infinite variance). Interpreting this will be another goal for the remainder of this chapter. First, several definitions are needed.

Definition 16. A process $\mathbf{x}(\cdot, \cdot)$ is a *white Gaussian process* if for any choice of $t_1, t_2, \ldots, t_N \in T$, the N random vectors $\mathbf{x}(t_1, \cdot), \mathbf{x}(t_2, \cdot), \ldots, \mathbf{x}(t_N, \cdot)$ are independent Gaussian random vectors. This implies $\mathbf{P}_{xx}(t_i, t_j) = 0 \;\forall\; i \neq j$.

Consider now a white Gaussian process $\mathbf{w}(t, \cdot)$. While a process modeling white noise should be independent at any two separate times, the infinite variance of the white Gaussian process is going to cause issues with integrability in the Riemann sense. Thus, while $\mathbf{w}(t, \cdot)$ is nice for formally describing the model incorporating noise, additional work is needed to present a *solvable* differential equation.

Definition 17. A process $\mathbf{x}(t, \cdot)$ is said to be a *process with independent increments* if for

any $N \in \mathbb{N}$ and any partition $t_0 < t_1 < \cdots < t_N \in T$ the N random variables

$$\begin{aligned}
\delta_1(\cdot) &= [\mathbf{x}(t_1, \cdot) - \mathbf{x}(t_0, \cdot)] \\
\delta_2(\cdot) &= [\mathbf{x}(t_2, \cdot) - \mathbf{x}(t_1, \cdot)] \\
&\vdots \\
\delta_N(\cdot) &= [\mathbf{x}(t_N, \cdot) - \mathbf{x}(t_{N-1}, \cdot)]
\end{aligned}$$

are mutually independent.

Definition 18. The process $\beta(\cdot, \cdot)$ is called a *scalar constant-diffusion Brownian motion process* if

1. it is a process with independent increments,

2. the increments are Gaussian random variables such that for any $t_1, t_2 \in T$,

$$\begin{aligned}
\mathbf{E}\{\beta(t_2) - \beta(t_1)\} &= 0 \\
\mathbf{E}\{[\beta(t_2) - \beta(t_1)]^2\} &= q \mid t_2 - t_1 \mid
\end{aligned}$$

$$\beta(t_0, \omega_i) = 0 \ \forall \ \omega_i \in \Omega \text{ except possibly a set of probability zero.}$$

After listing all of these requirements for a Brownian Process, an important question to ask is "Does such a process even exist?". Though a proof will not be included in this report, an enjoyable quote regarding Brownian Motion that should be shared is from Professor David Freedman, who said "One of the leading results on Brownian Motion is that it exists!". It is good to know that the ideas and analysis done throughout the remainder of this report will not be simply make believe. Below is a figure of several samples of a Brownian Process to assist in the intuition that will be built in the remainder of the section.

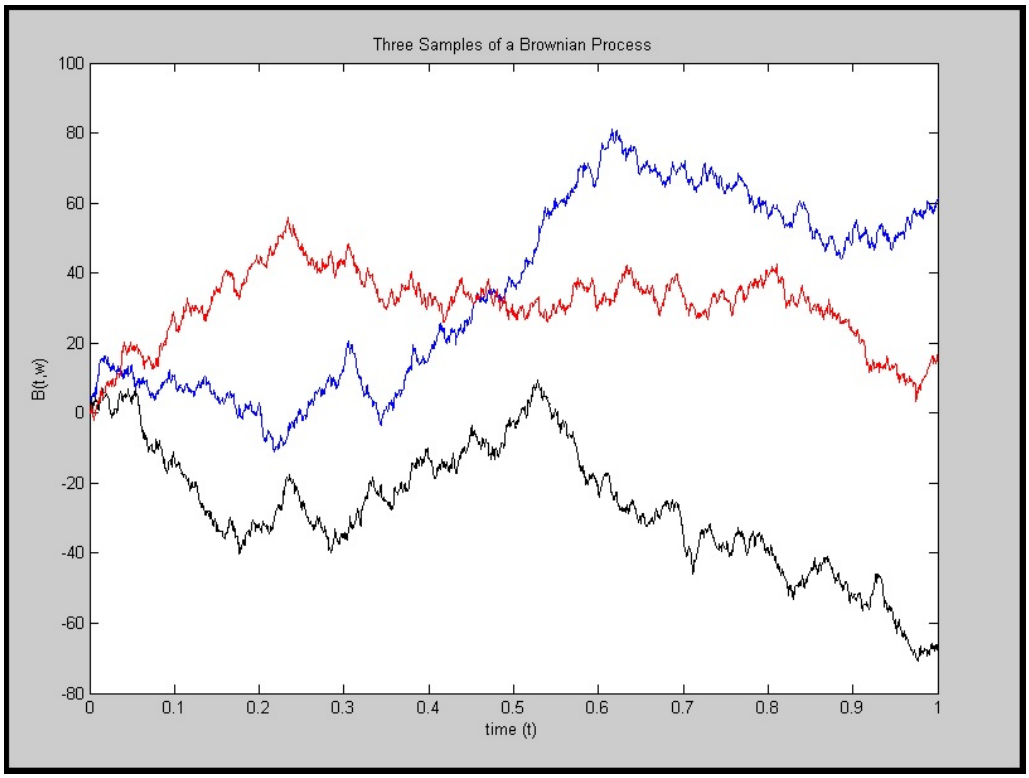

Figure 2.1: Three Samples of a Brownian Process

It is convenient at this moment to look back to the examples stated at the beginning of Section 1.3. Specifically, the gas molecule being pushed around by the countless others in its environment. From one moment to the next, it is nearly impossible to tell which direction the molecule will be bumped next. As one looks at the expected change in position from t_1 to t_2 it could move in *any* direction, therefore one can expect that from one moment to the next the overall effects balance out in a way to *expect* it to remain where it was. At the same time, the number of molecules and the rate at which they're moving can be expected to affect how quickly the molecule bounces around. This is summed up through the *diffusion rate* q. Over some period of time, the molecule is expected to vary from its initial starting point (zero) at the given diffusion rate.

It should be noted that a model which is just as easy to work with is the scalar (non-constant) diffusion Brownian Process. Here, the only difference is that the diffusion rate q is actually a function of time $q(t)$. The only change this presents to the above model, is

that the variance of an increment is now found to be

$$\mathbf{E}\{[\beta(t_2) - \beta(t_1)]^2\} = \int_{t_1}^{t_2} q(\tau) d\tau$$

With both white Gaussian noise and Brownian Motion defined, a link between the two can begin to form. While it can be shown that Brownian motion is continuous, but non-differentiable in the usual sense, it is possible to show that the "derivative" of a Brownian Motion process is Gaussian white noise in a slightly different sense. The goal now is to gain a better understanding of a Brownian process, as well motivate the need for a way to integrate stochastic processes by seeing exactly why $\mathbf{w}(t)$ cannot be integrated.

First, an interesting fact about Brownian motion will be shown [4]. Let $t_0, s, t \in T$ with $t_0 = 0 < s < t$, then consider a scalar Brownian motion process $\boldsymbol{\beta}(t, \cdot)$with constant diffusion $q = 1$. Then consider the covariance of the process at the two times.

$$\begin{aligned}
\mathbf{E}[\boldsymbol{\beta}(s)\,\boldsymbol{\beta}(t)] &= \mathbf{E}[\boldsymbol{\beta}(t)\,\boldsymbol{\beta}(s) - \boldsymbol{\beta}(s)^2 + \boldsymbol{\beta}(s)^2] \\
&= \mathbf{E}[(\boldsymbol{\beta}(t) - \boldsymbol{\beta}(s))\,\boldsymbol{\beta}(s)] + \mathbf{E}[\boldsymbol{\beta}(s)^2] \\
&= \mathbf{E}[(\boldsymbol{\beta}(t) - \boldsymbol{\beta}(s))(\boldsymbol{\beta}(s) - \boldsymbol{\beta}(0))] + \mathbf{E}[\boldsymbol{\beta}(s)^2] \\
&= \mathbf{E}[(\boldsymbol{\beta}(t) - \boldsymbol{\beta}(s))]\,\mathbf{E}[\boldsymbol{\beta}(s) - \boldsymbol{\beta}(0)] + \mathbf{E}[\boldsymbol{\beta}(s)^2] \\
&= \mathbf{E}[\boldsymbol{\beta}(s)^2] = 1|s - 0| = s
\end{aligned} \tag{2.1}$$

This says that the covariance of the Brownian process at *any* two times is simply the lesser of those two values. One way to understand this intuitively is by imagining that the Brownian process has realized some value at time s. From that point, it is equally probable to move in any direction from there out to time t. Thus, the best idea for the covariance of the two random variables is simply the variance of the lesser one.

Now, since Brownian motion is non-differentiable in the usual sense, consider a process

$\mathbf{w}_{\triangle t}(t,\cdot)$ defined by

$$\mathbf{w}_{\triangle t}(t) = \frac{\boldsymbol{\beta}(t+\triangle t) - \boldsymbol{\beta}(t)}{\triangle t} \tag{2.2}$$

for some fixed time step $\triangle t$. Because Brownian motion is zero mean, it is clear that $\mathbf{w}_{\triangle t}(t)$ is also zero mean. Now consider the covariance at the two times $s, t \in T$ with $s < t$. The following computations will make use of the fact shown above for the covariance of a Brownian process.

$$\begin{aligned}
\mathbf{E}[\mathbf{w}_{\triangle t}(t)\,\mathbf{w}_{\triangle t}(s)] &= \mathbf{E}\left[\frac{\boldsymbol{\beta}(t+\triangle t) - \boldsymbol{\beta}(t)}{\triangle t}\frac{\boldsymbol{\beta}(s+\triangle t) - \boldsymbol{\beta}(s)}{\triangle t}\right] \\
&= \left(\frac{1}{\triangle t}\right)^2 [\mathbf{E}\{\boldsymbol{\beta}(t+\triangle t)\,\boldsymbol{\beta}(s+\triangle t)\} - \mathbf{E}\{\boldsymbol{\beta}(t+\triangle t)\,\boldsymbol{\beta}(s)\} \\
&\quad - \mathbf{E}\{\boldsymbol{\beta}(t)\,\boldsymbol{\beta}(s+\triangle t)\} + \mathbf{E}\{\boldsymbol{\beta}(t)\,\boldsymbol{\beta}(s)\}] \\
&= \left(\frac{1}{\triangle t}\right)^2 [s + \triangle t - s - \mathbf{E}\{\boldsymbol{\beta}(t)\,\boldsymbol{\beta}(s+\triangle t)\} + s] \\
&= \left(\frac{1}{\triangle t}\right)^2 [s + \triangle t - \mathbf{E}\{\boldsymbol{\beta}(t)\,\boldsymbol{\beta}(s+\triangle t)\}]
\end{aligned}$$

At this point, the value of $\mathbf{E}\{\boldsymbol{\beta}(t)\,\boldsymbol{\beta}(s+\triangle t)\}$ is dependent on the size of $\triangle t$. If $s + \triangle t < t$, then by fact proved earlier, $\mathbf{E}\{\boldsymbol{\beta}(t)\,\boldsymbol{\beta}(s+\triangle t)\} = s + \triangle t$, and thus $\mathbf{E}[\mathbf{w}_{\triangle t}(t)\,\mathbf{w}_{\triangle t}(s)] = 0$. However, if $s + \triangle t > t$, then the following value is obtained

$$\begin{aligned}
\mathbf{E}[\mathbf{w}_{\triangle t}(t)\,\mathbf{w}_{\triangle t}(s)] &= \left(\frac{1}{\triangle t}\right)^2 [s + \triangle t - \mathbf{E}\{\boldsymbol{\beta}(t)\,\boldsymbol{\beta}(s+\triangle t)\}] \\
&= \left(\frac{1}{\triangle t}\right)^2 [s + \triangle t - t] \\
&= \frac{1}{\triangle t}\left[1 - \frac{t-s}{\triangle t}\right]
\end{aligned}$$

Define the function $\delta_{\triangle t}(t-s)$ to be

$$\delta_{\triangle t}(t-s) = \begin{cases} \frac{1}{\triangle t}\left[1 - \frac{t-s}{\triangle t}\right] & t - s < \triangle t \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

From here it is easy to see that $\mathbf{w}_{\triangle t}(t)$ appears to be fairly close to a zero mean white Gaussian process. However, the variance existing over a small interval $\triangle t$ prevents it from being a true white process. To turn this into a white Gaussian process, $\triangle t$ must be allowed to go to zero. However, to fully understand the next step, a significant amount of analysis is needed beyond that which is required to have a good understanding of stochastic integrals and differential equations. Therefore, it will simply be stated that as $\triangle t \to 0$, the function $\delta_{\triangle t}(t-s)$ converges in *distribution* to the δ-distribution, and $\mathbf{w}_{\triangle t}(t)$ converges to $\mathbf{w}(t)$.

$$\mathbf{E}[\mathbf{w}(t)\,\mathbf{w}(s)] = \delta(t-s) \tag{2.4}$$

This shows why it is not possible for the model involving white noise to simply be integrated, and why the stochastic integral involving a Brownian process is necessary (because $\mathbf{w}(t)$ is not even a function!). Therefore, the relation that will be used for the purposes of this report can be written

$$\mathbf{w}(t)dt = d\,\boldsymbol{\beta}(t)$$

Thus, the integral solution for the model incorporating noise in Section 2.4 can be rewritten

$$\int_{t_0}^{t} \mathbf{\Phi}(t,\tau)G(\tau)\,\mathbf{w}(\tau)d\tau \to \int_{t_0}^{t} \mathbf{\Phi}(t,\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)$$

The work in this section indicates that this is an equivalent (if informal) equality, and sets the stage for meaning to be given to the previously incomputable integral.

The statistics computed throughout this section can be extended to *vector Brownian*

motion with time-varying diffusion $\mathbf{Q}(t)$ as follows

$$\begin{aligned}
\mathbf{E}\{\boldsymbol{\beta}(t)\} &= 0 \\
\mathbf{E}\{[\boldsymbol{\beta}(t_2) - \boldsymbol{\beta}(t_1)][\boldsymbol{\beta}(t_2) - \boldsymbol{\beta}(t_1)]^{\mathrm{T}}\} &= \int_{t_1}^{t_2} \mathbf{Q}(t)dt \\
\mathbf{E}\{\mathbf{w}(t)\} &= 0 \\
\mathbf{E}\{\mathbf{w}(t)\,\mathbf{w}^{\mathrm{T}}(t')\} &= \mathbf{Q}(t)\delta(t-t')
\end{aligned}$$

where $t_2 \geq t_1$, $\mathbf{Q}(t)$ is symmetric positive definite $\forall\ t \in T$, and $\mathbf{w}(t)$ is the vector white Gaussian noise corresponding to the vector Brownian motion $\boldsymbol{\beta}(t)$.

2.5 Stochastic Integrals

To develop the idea of an integral of a stochastic process, one may begin as usual by looking at approximating area under the curve.

Let $a(\cdot)$ be a known, piecewise continuous scalar function of time and $\beta(\cdot,\cdot)$ a scalar Brownian process of diffusion $q(t)$ for all $t \in T = [0,\infty]$. Partition the time interval $[t_0, t]$ into N steps with no requirements on step size. Order the partitioning as $t_0 < t_1 < ... < t_N$ and let

$$\triangle t_N = \max_{i=1,...,N}[t_i - t_{i-1}]$$

Then define the *simple function* $a_N(\cdot)$ by taking the value of $a(t_{i-1})$ on each interval $[t_{i-1}, t_i]$.

$$a_N(t) = \begin{cases} a(t_0) & t \in [t_0, t_1) \\ a(t_1) & t \in [t_1, t_2) \\ \vdots & \vdots \\ a(t_{N-1}) & t \in [t_{N-1}, t_N) \end{cases}$$

Using this approximation to the function $a(\cdot)$, define what will be an approximation to the stochastic integral as follows:

$$\begin{aligned} l_N(t,\cdot) &= \sum_{i=0}^{N-1} a_N(t_i)[\beta(t_{i+1},\cdot) - \beta(t_i,\cdot)] \\ &= \int_{t_0}^{t} a_N(\tau) d\beta(\tau,\cdot) \end{aligned}$$

It is visible from the previous equation, the stochastic integral is similar to the Riemann-Stieltjes integral. However, it is complicated both by its dependence on $\omega \in \Omega$, and by the fact that Brownian Motion is not of bounded variation.

At this point, it is worth recalling the example of the stochastic process modeling the rise and fall of stock prices to gain an intuition of what the above integral is describing. If $a_N(t_i)$ represents the number of shares of stock X at time t_i, and $\beta(t_i)$ modeled the price of stock X at time t_i, then the product $a_N(t_i)\beta(t_i)$ is the value of all the shares held at time t_i, and $a_N(t_i)[\beta(t_{i+1}) - \beta t_i]$ is the change in value from time t_i to t_{i+1}. Finally, the integral can then be interpreted as the overall change in the value of shares held from time t_0 to t.

Recall now that a scalar Brownian Motion Process' increments are independent Gaussian random variables. Therefore, $l_N(t,\cdot)$ is the sum of independent Gaussian increments and thus, by Theorem 2, is Gaussian itself. Therefore, it is completely characterized by its mean and variance.

$$\begin{aligned}
\mathbf{E}[l_N(t)] &= \mathbf{E}\{[\sum_{i=0}^{N-1} a_N(t_i)[\beta(t_{i+1},\cdot) - \beta(t_i,\cdot)]\} \\
&= \sum_{i=0}^{N-1} a_N(t_i)\,\mathbf{E}\{[\beta(t_{i+1},\cdot) - \beta(t_i,\cdot)]\} \\
&= 0 \text{ (by the definition of Brownian motion)} \\
\mathbf{E}[l_N(t)^2] &= \mathbf{E}\{[\sum_{i=0}^{N-1} a_N(t_i)[\beta(t_{i+1},\cdot) - \beta(t_i,\cdot)]]^2\}
\end{aligned}$$

Consider expanding the above product. For $i \neq j$ it can be seen that a single term results in.

$$\begin{aligned}
&\mathbf{E}\{a_N(t_i)a_N(t_j)[\beta(t_{i+1}) - \beta(t_i)][\beta(t_{j+1}) - \beta(t_j)]\} \\
&\quad = a_N(t_i)a_N(t_j)\,\mathbf{E}\{[\beta(t_{i+1}) - \beta(t_i)][\beta(t_{j+1}) - \beta(t_j)]\} \\
&\quad = 0 \text{ (by the independent increments of the Brownian motion process)}
\end{aligned}$$

Thus, the second moment of $l_N(t,\cdot)$ simplifies to

$$\begin{aligned}
\mathbf{E}[l_N(t)^2] &= \mathbf{E}\left[\sum_{i=0}^{N-1} a_N(t_i)^2[\beta(t_{i+1}) - \beta(t_i)]^2\right] \\
&= \sum_{i=0}^{N-1} a_N(t_i)^2\,\mathbf{E}[(\beta(t_{i+1}) - \beta(t_i))^2]
\end{aligned}$$

Recall from Section 2.4

$$\mathbf{E}[(\beta(t_{i+1}) - \beta(t_i))^2] = \int_{t_i}^{t_{i+1}} q(\tau)d\tau$$

Therefore, the mean square value becomes

$$\mathbf{E}[l_N(t)^2] = \sum_{i=0}^{N-1} a_N(t_i)^2 \int_{t_i}^{t_{i+1}} q(\tau)d\tau = \int_{t_0}^{t_N} a_N(\tau)^2 q(\tau)d\tau$$

This gives one of the first assumptions for the current definition of a stochastic integral. That is, $a(\cdot)$ must be a known, piecewise continuous scalar function of time and must also satisfy that $l_N(t,\cdot)$ has finite variance. Specifically, it is required that $\mathbf{E}[l_N(t,\cdot)^2] < \infty$. It is desirable to extend the current definition from *simple functions* $a_N(t)$ to piecewise continuous $a(\cdot)$. By letting $\triangle t_N \to 0$ it will be seen that the approximations $l_N(t,\cdot)$ converge to a process which will then be defined as the *stochastic integral*.

Before continuing on, it is important to first determine what type of convergence is being discussed. There are two types that exist in this report. The first has been present throughout the chapter, though not discussed directly as it was not critical to paint the broad strokes required to reach the present topic. This convergence is called *convergence in probability*. For proving the convergence of the stochastic integral approximations, *convergence in mean square* will be used, and will be shown to imply convergence in probability as well.

Definition 19. A sequence of random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is said to *converge in probability* to $\mathbf{x}$ if $\forall\ \epsilon > 0$

$$\lim_{K\to\infty} \mathrm{P}(\{\omega :\mid \mathbf{x}_k(\omega) - \mathbf{x}(\omega) \mid \geq \epsilon\}) \to 0$$

Definition 20. A sequence of random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is said to *converge in mean square* (or in the mean) to a random variable $\mathbf{x}$ if $\mathbf{E}[\mathbf{x}_k^2]$ is finite for all k, $\mathbf{E}[\mathbf{x}^2]$ is finite, and

$$\mathbf{E}[(\mathbf{x}_k - \mathbf{x})^2] = 0.$$

The fact that convergence in the mean implies convergence in probability follows directly

from the Chebyshev inequality:

$$\mathrm{P}(\{\omega :\mid \mathbf{x}_k(\omega) - \mathbf{x}(\omega) \mid \geq \epsilon\} \leq \mathbf{E}[(\mathbf{x}_k(\omega) - \mathbf{x}(\omega))^2]/\epsilon^2$$

Consider the *Hilbert space* composed of deterministic functions of time $a_N(\cdot)$ defined on the set $[t_0, t]$ such that $\int_{t_0}^{t} a_N(\tau)d\beta(\tau, \cdot)$ is finite for piecewise continuous $q(\cdot)$ and $\forall N \in \mathbb{N}$.

$$\text{The norm here is given by } \|a_N\| = \sqrt{\int_{t_0}^{t} a_N(\tau)^2 q(\tau)d\tau}$$

To look at the convergence of the approximate stochastic integrals, it is necessary to define another integral like $l_N(t, \cdot)$, however the new integral will have $P > N$ partitions. Thus, it will be called $l_P(t, \cdot)$ and will have the form:

$$l_P(t, \cdot) = \sum_{i=0}^{P-1} a_P(t_i)[\beta(t_{i+1}, \cdot) - \beta(t_i, \cdot)]$$

Because a_N and a_P are both piecewise constant, their difference must be piecewise constant, with at most N+P points of discontinuity. Thus, let the partition $K = P \cup N$, then the difference between the two integrals can be found by

$$\begin{aligned}
[l_N(t) - l_P(t)] &= \sum_{k=0}^{K-1} [a_N(t_k) - a_P(t_k)][\beta(t_{k+1}) - \beta(t_k)] \\
&= \int_{t_0}^{t} [a_N(\tau) - a_P(\tau)]d\beta(\tau) \qquad (\text{ as } K \to \infty \text{ and } \triangle t_N \to 0)
\end{aligned}$$

Since $a_N(\cdot) - a_P(\cdot)$ is piecewise constant the integral can be de-constructed by looking at it as a limit of a sum, as was done when computing the variance of l_N,

$$\mathbf{E}\{[l_N(t) - l_P(t)]^2\} = \mathbf{E}\left\{\left[\int_{t_0}^{t} [a_N(\tau) - a_P(\tau)]d\beta(\tau)\right]^2\right\}$$
$$= \int_{t_0}^{t} [a_N(\tau) - a_P(\tau)]^2 q(\tau)d\tau$$

Finally, because distance can be measured by

$$\|(l_N(t) - l_p(t)\|^2 = \mathbf{E}\{[l_N(t) - l_P(t)]^2\}.$$

and the random variables are zero mean with finite variance, the *Hilbert space* of random variables $l_N(t,\cdot)$ can be utilized to determine the convergence of the integrals. Consider a sequence of random variables $l_1(t,\cdot), l_2(t,\cdot), ...$ such that as the partition $|P| \to \infty$, and $\triangle t_N \to 0$ the sequence $\{l_i(t,\cdot)\}$ is a Cauchy sequence which will converge to a limit in the space, denoted $l(t,\cdot)$.

$$\lim_{k\to\infty} \|(l(t) - l_k(t))^2\| = \mathbf{E}\{(l(t) - l_k(t))^2\} = 0.$$

Thus there is now the following definition of the *scalar stochastic integral* $l(\cdot,\cdot)$:

$$\begin{aligned} l(t,\cdot) &= \int_{t_0}^{t} a(\tau)d\beta(\tau) \\ &= \lim_{N\to\infty} l_N(t,\cdot) \\ &= \lim_{N\to\infty} \int_{t_0}^{t} a_N(\tau)d\beta(\tau) \end{aligned} \tag{2.5}$$

Now, because Brownian motion is Gaussian and only linear operations on $\beta(\cdot,\cdot)$ were used in developing $l(\cdot,\cdot)$, Theorem 2 implies that $l(t,\cdot)$ is Gaussian. Recalling that $\mathbf{E}[l_N(t)] = 0 \implies \mathbf{E}[l(t)] = 0$ and $\mathbf{E}[l(t)^2] = \int_{t_0}^{t} a(\tau)^2 q(\tau)d\tau$, $l(\cdot,\cdot)$ is completely characterized by these

two values.

Additionally, one can see using the previous definitions and manipulations that stochastic integrals have the following properties. Let $a(\cdot)$ and $c(\cdot)$ be piecewise continuous functions with Brownian processes $\beta(\cdot)$ and $\kappa(\cdot)$, and r a constant, then the following equalities hold

1. $\int_{t_0}^{t_2} a(\tau)d\beta(\tau) = \int_{t_0}^{t_1} a(\tau)d\beta(\tau) + \int_{t_1}^{t_2} a(\tau)d\beta(\tau)$
2. $\int_{t_0}^{t} [a(\tau) + c(\tau)]d\beta(\tau) = \int_{t_0}^{t} a(\tau)d\beta(\tau) + \int_{t_0}^{t} c(\tau)d\beta(\tau)$
3. $\int_{t_0}^{t} a(\tau)d[\beta(\tau) + \kappa(\tau)] = \int_{t_0}^{t} a(\tau)d\beta(\tau) + \int_{t_0}^{t} a(\tau)d\kappa(\tau)$
4. $\int_{t_0}^{t} ra(\tau)d\beta(\tau) = r\int_{t_0}^{t} a(\tau)d\beta(\tau) = \int_{t_0}^{t} a(\tau)d[r\beta(\tau)]$

Finally, $l(\cdot,\cdot)$ is itself a Brownian motion process with rescaled diffusion by the following. Consider two disjoint intervals $(t_1, t_2]$ and $(t_3, t_4]$ with $[l(t_2) - l(t_1)] = \int_{t_1}^{t_2} a(\tau)d\beta(\tau)$ and $[l(t_4) - l(t_3)] = \int_{t_3}^{t_4} a(\tau)d\beta(\tau)$. Because the increments of $\beta(\cdot,\cdot)$ in $[t_1, t_2)$ are independent of the increments in $(t_3, t_4]$ it is clear that $[l(t_2) - l(t_1)]$ and $[l(t_4) - l(t_3)]$ are independent, zero-mean, Gaussian increments of $l(\cdot,\cdot)$.

The extension of vector stochastic integrals is very similar to the that of random variables in Section 2.1. Let $\boldsymbol{\beta}(\cdot,\cdot)$ be Brownian motion (Gaussian process) with statistics

$$\mathbf{E}\{\boldsymbol{\beta}(t)\} = \mathbf{0}$$
$$\mathbf{E}\{[\boldsymbol{\beta}(t_2) - \boldsymbol{\beta}(t_1)][\boldsymbol{\beta}(t_2) - \boldsymbol{\beta}(t_1)]^{\mathrm{T}}\} = \int_{t_1}^{t_2} \mathbf{Q}(\tau)d\tau$$

Where $\mathbf{Q}(\cdot)$ is a matrix of piecewise continuous functions such that $\mathbf{Q}(t)$ is symmetric positive semidefinite. If $\mathbf{A}(\cdot)$ is an n-by-s matrix of piecewise continuous time functions, then similar methods to the scalar version development of the stochastic integral yeild the n-dimensional vector-valued stochastic integral

$$\mathbf{l}(t,\cdot) = \int_{t_0}^{t} \mathbf{A}(\tau)d\,\boldsymbol{\beta}(\tau)$$

This results in a Gaussian random vector determined by the statistics $\mathbf{E}\{\mathbf{l}(t)\} = \mathbf{0}$ and $\mathbf{E}\{\mathbf{l}(t)\,\mathbf{l}(t)^{\mathrm{T}}\} = \int_{t_0}^{t} \mathbf{A}(\tau)\,\mathbf{Q}(\tau)\mathbf{A}(\tau)d\tau$. Finally, it is visible that $\mathbf{l}(\cdot,\cdot)$ is itself a Brownian motion process with diffusion: $\mathbf{E}\{[\mathbf{l}(t_2) - \mathbf{l}(t_1)][\mathbf{l}(t_2) - \mathbf{l}(t_1)]^{\mathrm{T}}\} = \int_{t_1}^{t_2} \mathbf{A}(\tau)\,\mathbf{Q}(\tau)\,\mathbf{A}^{\mathrm{T}}(\tau)$.

2.6 Stochastic Differential Equations

Before presenting a workable stochastic model and examining its solution and properties, several relations will be worth working out to simplify the process. First, consider a stochastic integral of the form

$$\mathbf{l}(t) = \mathbf{l}(t_0) + \int_{t_0}^{t} A(\tau) d\,\boldsymbol{\beta}(\tau)$$

then the *stochastic differential* of $\mathbf{l}(t)$ can be defined as

$$\mathbf{dl}(t) = A(t) d\,\boldsymbol{\beta}(t)$$

so that $\mathbf{dl}(t)$ is a differential in the sense that

$$\int_{t_0}^{t} \mathbf{dl}(t) = \mathbf{l}(t) - \mathbf{l}(t_0)$$

and can be thought of as an infinitesimal difference in $\mathbf{l}(t)$.

Next, let $\mathbf{D}(\cdot)$ be a known matrix of differentiable functions, and a random process $\mathbf{y}(\cdot,\cdot)$ defined by

$$\mathbf{y}(t) = \mathbf{D}(t)\,\mathbf{l}(t) \tag{2.6}$$

Then by partitioning the interval $[t_0, t]$ into N steps, the process $\mathbf{l}(t)$ and $\mathbf{D}(t)$ can be

rewritten

$$\mathbf{l}(t) = \mathbf{l}(t_0) + \sum_{i=0}^{N-1} [\mathbf{l}(t_{i+1}) - \mathbf{l}(t_i)]$$
$$\mathbf{D}(t) = \mathbf{D}(t_0) + \sum_{i=0}^{N-1} [\mathbf{D}(t_{i+1}) - \mathbf{D}(t_i)]$$

Substituting these equations into the definition of $\mathbf{y}(t)$, multiplying, and rearranging yields

$$\mathbf{y}(t) = \mathbf{D}(t_0)\,\mathbf{l}(t_0) + \sum_{i=0}^{N-1} [\mathbf{D}(t_{i+1}) - \mathbf{D}(t_i)]\,\mathbf{l}(t_{i+1}) + \sum_{i=0}^{N-1} \mathbf{D}(t_i)[\mathbf{l}(t_{i+1}) - \mathbf{l}(t_i)]$$

Now by the mean value theorem, $[\mathbf{D}(t_{i+1}) - \mathbf{D}(t_i)]$ can be rewritten to $\dot{\mathbf{D}}(\tau_i)[t_{i+1} - t_i]$ for some $\tau_i \in [t_i, t_{i+1}]$. Then by letting $N \to \infty$, $\mathbf{y}(t)$ becomes

$$\mathbf{y}(t) = \int_{t_0}^{t} \dot{\mathbf{D}}(\tau)\,\mathbf{l}(\tau)d\tau + \int_{t_0}^{t} \mathbf{D}(\tau)d\,\mathbf{l}(\tau) + \mathbf{D}(t_0)\,\mathbf{l}(t_0) \tag{2.7}$$

Recall that for a realization $\mathbf{l}(t, \omega_i)$ that $\mathbf{l}(\cdot, \omega_i)$ is a function of time, so the first integral can computed as a regular Riemann integral. The middle term is the stochastic integral from Section 2.5, and the final term is a known vector value. Finally, the stochastic differential of the process $\mathbf{y}(t)$ is written

$$\mathbf{dy}(t) = \dot{\mathbf{D}}(t)\,\mathbf{l}(t)dt + \mathbf{D}(t)d\,\mathbf{l}(t) \tag{2.8}$$

Recall the model obtained by adding noise to the deterministic control system

$$\dot{\mathbf{x}}(t) = F(t)\,\mathbf{x}(t) + B(t)\,\mathbf{u}(t) + G(t)\,\mathbf{n}_1(t)$$

with the proposed solution

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\,\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau) B(\tau)\,\mathbf{u}(\tau) d\tau + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau) G(\tau)\,\mathbf{n}_1(\tau) d\tau$$

the extensive work in the previous sections was required to give meaning to these equations, and will be shown now by rewriting them in a more workable form. The model will initially be solved without the control function included. This will simplify a number of computations, and will have little effect on the statistics of the model since $B(t)\,\mathbf{u}(t)$ is a deterministic function. It will be included after a solid foundation has been set. First, consider the stochastic differential equation

$$\mathbf{dx}(t) = F(t)\,\mathbf{x}(t) dt + G(t) d\,\boldsymbol{\beta}(t) \qquad\qquad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{2.9}$$

The first goal is to find a process which solves equation (2.9) by satisfying the integral equation

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} F(\tau)\,\mathbf{x}(\tau) d\tau + \int_{t_0}^{t} G(t) d\,\boldsymbol{\beta}(t) \tag{2.10}$$

where $G(\cdot)$ is a known $n \times s$ matrix of piecewise continuous functions, and $\boldsymbol{\beta}(\cdot, \cdot)$ is an s-vector values Brownian motion process of diffusion $\mathbf{Q}(t)\ \forall\ t \in T$. Additionally, the middle term is a Riemann integral for a particular realization of $\mathbf{x}(t)$, and the last term is a stochastic integral.

Consider a solution $\mathbf{x}(\cdot, \cdot)$ defined by

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\,\mathbf{y}(t) \tag{2.11}$$

where $\mathbf{\Phi}(t, t_0)$ is the state transition matrix satisfying the usual properties, and $\mathbf{y}(t)$ is the

stochastic process defined by

$$\mathbf{y}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t_0, \tau) G(\tau) d\,\boldsymbol{\beta}(\tau) \tag{2.12}$$

Note that in the state transition matrix, the order of variables are swapped to indicate a backwards transition. This can be better understood by plugging $\mathbf{y}(t)$ into (2.11). Then the state transition matrix is (in a way) reversing the noise. The goal now is to demonstrate that the proposed solution satisfies both the initial condition and equation (2.10)

$$\begin{aligned}
\mathbf{x}(t_0) &= \mathbf{\Phi}(t_0, t_0)\, \mathbf{y}(t_0) \\
&= \mathbf{x}(t_0) + \int_{t_0}^{t_0} \mathbf{\Phi}(t_0, \tau) G(\tau) d\,\boldsymbol{\beta}(\tau) \\
&= \mathbf{x}(t_0) = \mathbf{x}_0
\end{aligned}$$

Thus, the initial condition is satisfied. Now, the form of equation (2.11) is the same as (2.6), therefore it can be written just as equation (2.8)

$$\begin{aligned}
\mathbf{dx}(t) &= \frac{d\,\mathbf{\Phi}(t, t_0)}{dt}\, \mathbf{y}(t) dt + \mathbf{\Phi}(t, t_0)\, \mathbf{dy}(t) \\
&= [F(t)\, \mathbf{\Phi}(t, t_0)]\, \mathbf{y}(t) dt + \mathbf{\Phi}(t, t_0)\, \mathbf{dy}(t) \\
&= F(t)\, \mathbf{\Phi}(t, t_0)\, \mathbf{y}(t) dt + \mathbf{\Phi}(t, t_0) [\mathbf{\Phi}(t_0, \tau) G(\tau)\, \mathbf{d}\boldsymbol{\beta}(\tau)]
\end{aligned}$$

now substituting into the known equation

$$\begin{aligned}
\mathbf{x}(t) &= \mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{dx}(\tau) \\
&= \mathbf{x}(t_0) + \int_{t_0}^{t} F(\tau)\,\mathbf{\Phi}(\tau, t_0)\,\mathbf{y}(\tau)d\tau + \int_{t_0}^{t} \mathbf{\Phi}(\tau, t_0)[\mathbf{\Phi}(t_0, \tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)] \\
&= \mathbf{x}(t_0) + \int_{t_0}^{t} F(\tau)\,\mathbf{\Phi}(\tau, t_0)\,\mathbf{\Phi}(t_0, \tau)\,\mathbf{x}(\tau)d\tau + \int_{t_0}^{t} \mathbf{\Phi}(\tau, t_0)[\mathbf{\Phi}(t_0, \tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau) \\
&= \mathbf{x}(t_0) + \int_{t_0}^{t} F(\tau)\,\mathbf{x}(\tau)d\tau + \int_{t_0}^{t} G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)
\end{aligned}$$

thus, the proposed solution satisfies equation (2.10) and the initial condition. The solution to the system lacking the control function can then be written

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\,\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau) \tag{2.13}$$

Having obtained the solution to the uncontrolled system, it is now visible that because the stochastic integral in equation 2.13 is composed of independent Gaussian increments, $\mathbf{x}(\cdot, \cdot)$ will be a Gaussian process if $\mathbf{x}(t_0)$ is a Gaussian random variable, or is a known quantity. Therefore, it will again be beneficial to obtain the mean, covariance matrix, and several other measurements of equation (2.13).

First, the expectation and correlation matrix will be computed:

$$\begin{aligned}
\hat{\mathbf{x}}(t) &= \mathbf{E}\{\mathbf{x}(t)\} \\
&= \mathbf{\Phi}(t, t_0)\,\mathbf{E}\{\mathbf{x}(t_0)\} + \mathbf{E}\left\{\int_{t_0}^{t} \mathbf{\Phi}(t, \tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)\right\}
\end{aligned}$$

Recall from Section 2.3 that the stochastic integral has mean zero, so that the mean function for the process $\mathbf{x}(t)$ is then

$$\hat{\mathbf{x}}(t) = \mathbf{\Phi}(t, t_0)\hat{\mathbf{x}}(t_0) \tag{2.14}$$

The correlation matrix is computed similarly

$$\begin{aligned}\mathbf{E}\{\mathbf{x}(t)\,\mathbf{x}^{\mathrm{T}}(t)\} = \mathbf{E}\{\mathbf{\Phi}(t,t_0)\,\mathbf{x}(t_0)\,\mathbf{x}^{\mathrm{T}}(t_0)\,\mathbf{\Phi}^{\mathrm{T}}(t,t_0)\} \\ + \mathbf{E}\left\{\left[\int_{t_0}^{t}\mathbf{\Phi}(t,\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)\right][\mathbf{x}^{\mathrm{T}}(t_0)]\right\} \\ + \mathbf{E}\left\{[\mathbf{x}(t_0)]\left[\int_{t_0}^{t}\mathbf{\Phi}(t,\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)\right]^{\mathrm{T}}\right\} \\ + \int_{t_0}^{t}\mathbf{\Phi}(t,\tau)G(\tau)Q(\tau)G^{\mathrm{T}}(\tau)\,\mathbf{\Phi}^{\mathrm{T}}(t,\tau)d\tau\end{aligned}$$

Where $\mathbf{Q}(t)$ is the diffusion of the Brownian motion $\boldsymbol{\beta}(\cdot,\cdot)$. Also, recall that Brownian motion is independent of $\mathbf{x}(t_0)$ by its definition, so the two middle terms can be separated, and the zero mean of the stochastic integrals removes the two terms.

$$\mathbf{E}\{\mathbf{x}(t)\,\mathbf{x}^{\mathrm{T}}(t)\} = \mathbf{\Phi}(t,t_0)\,\mathbf{E}\{\mathbf{x}(t_0)\,\mathbf{x}^{\mathrm{T}}(t_0)\}\,\mathbf{\Phi}^{\mathrm{T}}(t,t_0) + \int_{t_0}^{t}\mathbf{\Phi}(t,\tau)G(\tau)Q(\tau)G^{\mathrm{T}}(\tau)\,\mathbf{\Phi}^{\mathrm{T}}(t,\tau)d\tau$$

recall form Section 2.2 that the covariance matrix is directly related to the correlation matrix through the following equation

$$\begin{aligned}\mathbf{E}\{\mathbf{x}(t)\,\mathbf{x}^{\mathrm{T}}(t)\} &= \mathbf{P}_{xx}(t) + \hat{\mathbf{x}}(t)\,\hat{\mathbf{x}}^{\mathrm{T}}(t) \\ \mathbf{E}\{\mathbf{x}(t_0)\,\mathbf{x}^{\mathrm{T}}(t_0)\} &= \mathbf{P}_{xx}(t_0) + \hat{\mathbf{x}}(t_0)\,\hat{\mathbf{x}}^{\mathrm{T}}(t_0)\end{aligned}$$

so by substituting in the two equations, the mean square value can reduce to a propagation

of $\mathbf{P}_{xx}(t)$ as follows

$$\begin{aligned}
\mathbf{P}_{xx}(t) &= \mathbf{E}\{\mathbf{x}(t)\,\mathbf{x}^{\mathrm{T}}(t)\} - \hat{\mathbf{x}}(t)\,\hat{\mathbf{x}}^{\mathrm{T}}(t) \\
&= \boldsymbol{\Phi}(t,t_0)\,\mathbf{E}\{\mathbf{x}(t_0)\,\mathbf{x}^{\mathrm{T}}(t_0)\}\,\boldsymbol{\Phi}^{\mathrm{T}}(t,t_0) \\
&\qquad + \int_{t_0}^{t} \boldsymbol{\Phi}(t,\tau)G(\tau)Q(\tau)G^{\mathrm{T}}(\tau)\,\boldsymbol{\Phi}^{\mathrm{T}}(t,\tau)d\tau - \hat{\mathbf{x}}(t)\,\hat{\mathbf{x}}^{\mathrm{T}}(t) \\
&= \boldsymbol{\Phi}(t,t_0)\{\mathbf{P}_{xx}(t_0) + \hat{\mathbf{x}}(t_0)\,\hat{\mathbf{x}}^{\mathrm{T}}(t_0)\}\,\boldsymbol{\Phi}^{\mathrm{T}}(t,t_0) \\
&\qquad + \int_{t_0}^{t} \boldsymbol{\Phi}(t,\tau)G(\tau)Q(\tau)G^{\mathrm{T}}(\tau)\,\boldsymbol{\Phi}^{\mathrm{T}}(t,\tau)d\tau - \hat{\mathbf{x}}(t)\,\hat{\mathbf{x}}^{\mathrm{T}}(t) \\
&= \boldsymbol{\Phi}(t,t_0)\,\mathbf{P}_{xx}(t_0)\,\boldsymbol{\Phi}^{\mathrm{T}}(t,t_0) + \hat{\mathbf{x}}(t)\,\hat{\mathbf{x}}^{\mathrm{T}}(t) \\
&\qquad + \int_{t_0}^{t} \boldsymbol{\Phi}(t,\tau)G(\tau)Q(\tau)G^{\mathrm{T}}(\tau)\,\boldsymbol{\Phi}^{\mathrm{T}}(t,\tau)d\tau - \hat{\mathbf{x}}(t)\,\hat{\mathbf{x}}^{\mathrm{T}}(t)
\end{aligned}$$

by then canceling out the $\hat{\mathbf{x}}(t)$ values, a simplified equation for the variance matrix can be obtained

$$\mathbf{P}_{xx} = \boldsymbol{\Phi}(t,t_0)\,\mathbf{P}_{xx}(t_0)\,\boldsymbol{\Phi}^{\mathrm{T}}(t,t_0) + \int_{t_0}^{t} \boldsymbol{\Phi}(t,\tau)G(\tau)Q(\tau)G^{\mathrm{T}}(\tau)\,\boldsymbol{\Phi}^{\mathrm{T}}(t,\tau)d\tau \tag{2.15}$$

Now that both the mean and covariance have been established, some interesting information can come from computing $\mathbf{x}(t_2)$ for $t_2 \geq t_1 \geq t_0$

$$\begin{aligned}
\mathbf{x}(t_2) &= \boldsymbol{\Phi}(t_2,t_0)\,\mathbf{x}(t_0) + \int_{t_0}^{t_2} \boldsymbol{\Phi}(t_2,\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau) \\
&= \boldsymbol{\Phi}(t_2,t_1)\,\boldsymbol{\Phi}(t_1,t_0)\,\mathbf{x}(t_0) \\
&\qquad + \boldsymbol{\Phi}(t_2,t_1)\int_{t_0}^{t_1} \boldsymbol{\Phi}(t_1,\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau) \\
&\qquad\qquad + \int_{t_1}^{t_2} \boldsymbol{\Phi}(t_2,\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau) \\
&= \boldsymbol{\Phi}(t_2,t_1)\,\mathbf{x}(t_1) + \int_{t_1}^{t_2} \boldsymbol{\Phi}(t_2,\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)
\end{aligned}$$

Examining the line above shows that to compute the value of $\mathbf{x}(t_2)$, the only information needed is that of $\mathbf{x}(t_1)$. To be able to compute the future value from only the present value (in addition to $\mathbf{x}(t)$ being Gaussian) makes this a *Gauss-Markov process*. The probability law that describes the process of evolution depends only on the present, and not upon the history of the process. This will be useful when incorporating measurements. Using this equation for the propagation from t_1 to t_2, the autocorrelation $\mathbf{E}\{\mathbf{x}(t_2)\,\mathbf{x}^{\mathrm{T}}(t_1)\}$ can be written

$$\begin{aligned}\mathbf{E}\{\mathbf{x}(t_2)\,\mathbf{x}^{\mathrm{T}}(t_1)\} &= \mathbf{\Phi}(t_2,t_1)\,\mathbf{E}\{\mathbf{x}(t_1)\,\mathbf{x}^{\mathrm{T}}(t_1)\} + \mathbf{E}\left\{\int_{t_1}^{t_2}\mathbf{\Phi}(t_2,\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)\,\mathbf{x}^{\mathrm{T}}(t_1)\right\}\\ &= \mathbf{\Phi}(t_2,t_1)\,\mathbf{E}\{\mathbf{x}(t_1)\,\mathbf{x}^{\mathrm{T}}(t_1)\}\end{aligned}$$

by the independence of $\boldsymbol{\beta}(\cdot,\cdot)$ over $[t_1,t_2)$ of both the intervals $[t_0,t_1)$ and $\mathbf{x}(t_0)$. Therefore, the covariance kernel for $t_2 \geq t_1$ is

$$\mathbf{P}_{xx}(t_2,t_1) = \mathbf{\Phi}(t_2,t_1)\,\mathbf{P}_{xx}(t_1)$$

which again can be computed without the history of the process, using only the state at t_1.

Now consider the stochastic differential equation formed by including the control function

$$\mathbf{dx}(t) = F(t)\,\mathbf{x}(t)dt + B(t)\,\mathbf{u}(t)dt + G(t)\,\mathbf{d}\boldsymbol{\beta}(t) \tag{2.16}$$

where $\mathbf{u}(t)$ is a vector of deterministic control inputs applied at time t, and $B(t)$ is the control input matrix. Then by combining the equation (2.13) with the solution to the linear control model, the solution to (2.16) is then

$$\mathbf{x}(t) = \mathbf{\Phi}(t,t_0)\,\mathbf{x}(t_0) + \int_{t_0}^{t}\mathbf{\Phi}(t,\tau)B(\tau)\,\mathbf{u}(\tau)d\tau + \int_{t_0}^{t}\mathbf{\Phi}(t,\tau)G(t)\,\mathbf{d}\boldsymbol{\beta}(\tau) \tag{2.17}$$

The addition of the *ordinary Riemann integral* of the deterministic function and matrices

$\int_{t_0}^{t} \mathbf{\Phi}(t,\tau)B(\tau)\,\mathbf{u}(\tau)d\tau$ contributes no uncertainty to the overall model, therefore, the only work needed to recompute the statistics will be for the mean, which is now given by

$$\hat{\mathbf{x}}(t) = \mathbf{\Phi}(t,t_0)\,\hat{\mathbf{x}}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t,\tau)B(\tau)\,\mathbf{u}(\tau)d\tau \tag{2.18}$$

$\mathbf{P}_{xx}(t)$ is still propagated by equation (2.15).

2.7 Discretizing the System

In many problems, software will be implemented to model the system and apply an optimal estimator which will use discrete-time measurements. Therefore, it is necessary to reduce the continuous-time dynamic process already described to a discrete-time model where the control function $\mathbf{u}(t)$ is held constant over each sample period from sample time t_i to t_{i+1}. Then the solution (2.17) to the system from (2.16) can be written as

$$\mathbf{x}(t_{i+1}) = \mathbf{\Phi}(t_{i+1},t_i)\,\mathbf{x}(t_i) + \left[\int_{t_i}^{t_{i+1}} \mathbf{\Phi}(t_{i+1},\tau)B(\tau)d\tau\right]\mathbf{u}(t_i) + \left[\int_{t_i}^{t_{i+1}} \mathbf{\Phi}(t_{i+1},\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)\right]$$

$$\mathbf{x}(t_{i+1}) = \mathbf{\Phi}(t_{i+1},t_i)\,\mathbf{x}(t_i) + B_d(t_i)\,\mathbf{u}(t_i) + \mathbf{w}_d(t_i) \tag{2.19}$$

where $B_d(t_i)$ is the discrete-time input matrix, $\mathbf{w}_d(\cdot,\cdot)$ is a vector-valued white Gaussian discrete-time stochastic process with statistics duplicating $\int_{t_i}^{t_{i+1}} \mathbf{\Phi}(t_{i+1},\tau)G(\tau)\,\mathbf{d}\boldsymbol{\beta}(\tau)$, with all relations given below

$$\begin{aligned}
B_d(t_i) &= \int_{t_i}^{t_{i+1}} \mathbf{\Phi}(t_{i+1},\tau)B(\tau)d\tau \\
\mathbf{E}\{\mathbf{w}_d(t_i)\} &= 0 \\
\mathbf{E}\{\mathbf{w}_d(t_i)\,\mathbf{w}_d^{\mathrm{T}}(t_i)\} &= \mathbf{Q}_d(t_i) = \int_{t_i}^{t_{i+1}} \mathbf{\Phi}(t_{i+1},\tau)G(\tau)\,\mathbf{Q}(\tau)G^{\mathrm{T}}(\tau)\,\mathbf{\Phi}^{\mathrm{T}}(t_{i+1},\tau)d\tau \\
\mathbf{E}\{\mathbf{w}_d(t_i)\,\mathbf{w}_d^{\mathrm{T}}(t_j)\} &= 0 \qquad t_i \neq t_j
\end{aligned}$$

Thus, the discrete process (2.19) is equivalent to the continuous process (2.17). The mean and covariance defined by (2.19) can be computed by

$$\hat{\mathbf{x}}(t_{i+1}) = \mathbf{\Phi}(t_{i+1}, t_i)\,\hat{\mathbf{x}}(t_i) + B_d(t_i)\,\mathbf{u}(t_i) \tag{2.20}$$

$$\mathbf{P}_{xx}(t_{i+1}) = \mathbf{\Phi}(t_{i+1}, t_i)\,\mathbf{P}_{xx}(t_i)\,\mathbf{\Phi}^{\mathrm{T}}(t_{i+1}, t_i) + G(t_i)\,\mathbf{Q}_d(t_i)G^{\mathrm{T}}(t_i) \tag{2.21}$$

Chapter 3

Optimal Estimation

3.1 The Overall Model

With both the continuous and discrete time models fully constructed, the measurements can now be included to form the overall system model. Recall that the measurement at time t_i was described through the m-dimensional random vector $\mathbf{z}(t_i)$ given by

$$\mathbf{z}(t_i) = H(t_i)\,\mathbf{x}(t_i) + \mathbf{v}(t_i) \tag{3.1}$$

where $H(\cdot)$ is an $m \times n$ measurement matrix, and $\mathbf{v}(t_i)$ is an m-dimensional vector of additive noise. For the remainder of the report, $\mathbf{v}(\cdot)$ will be assumed to be a white Gaussian discrete-time stochastic process, with

$$\mathbf{E}\{\mathbf{v}(t_i)\} = 0$$

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

Where $\mathbf{R}(t_i)$ is an $m \times m$ symmetric positive definite matrix $\forall\ t_i \in T$. Additionally, $\mathbf{v}(t_i, \cdot)$ will be assumed to be independent of both the initial condition $\mathbf{x}(t_0)$ and the dynamic driving noise $\boldsymbol{\beta}(t_j, \cdot)\ \forall\ t_i, t_j \in T$. Thus, the main system of interest is

$$\mathbf{dx}(t) = F(t)\,\mathbf{x}(t)dt + B(t)\,\mathbf{u}(t)dt + G(t)\,\mathbf{d}\boldsymbol{\beta}(t) \tag{3.2}$$

with the equivalent discrete time solution with measurements available at times $t_1, t_2, \ldots$ given by

$$\mathbf{x}(t_{i+1}) = \boldsymbol{\Phi}(t_{i+1}, t_i)\,\mathbf{x}(t_i) + B_d(t_i)\,\mathbf{u}(t_i) + G(t_i)\,\mathbf{w}_d(t_i) \tag{3.3}$$

The last thing needed before incorporating the measurements to determine a best guess, is to obtain the statistics of the measurements themselves. The mean is easily determined

$$\begin{aligned}
\hat{\mathbf{z}}(t_i) &= \mathbf{E}\{\mathbf{z}(t_i)\} \\
&= \mathbf{E}\{H(t_i)\,\mathbf{x}(t_i) + \mathbf{v}(t_i)\} \\
&= H(t_i)\,\mathbf{E}\{\mathbf{x}(t_i)\} + \mathbf{E}\{\mathbf{v}(t_i)\} \\
&= H(t_i)\,\hat{\mathbf{x}}(t_i)
\end{aligned} \tag{3.4}$$

and the autocorrelation is given by

$$\begin{aligned}
\mathbf{E}\{\mathbf{z}(t_j)\,\mathbf{z}^{\mathrm{T}}(t_i)\} &= \mathbf{E}\{H(t_j)\,\mathbf{x}(t_j)\,\mathbf{x}^{\mathrm{T}}(t_i)H^{\mathrm{T}}(t_i)\} + \mathbf{E}\{\mathbf{v}(t_j)\,\mathbf{v}^{\mathrm{T}}(t_i)\} \\
&\quad + \mathbf{E}\{H(t_j)\,\mathbf{x}(t_j)\,\mathbf{v}^{\mathrm{T}}(t_i)\} + \mathbf{E}\{\mathbf{v}(t_i)H^{\mathrm{T}}(t_j)\,\mathbf{x}^{\mathrm{T}}(t_j)\} \\
&= H(t_j)\,\mathbf{E}\{\mathbf{x}(t_j)\,\mathbf{x}^{\mathrm{T}}(t_i)\}H^{\mathrm{T}}(t_i) + \mathbf{E}\{\mathbf{v}(t_j)\,\mathbf{v}^{\mathrm{T}}(t_i)\}
\end{aligned}$$

by the independence of $\mathbf{v}(t_i)$ and $\mathbf{v}(t_j)$ from $\mathbf{x}(t_j)$ and $\mathbf{x}(t_i)$, respectively. The independence can be seen by expanding $\mathbf{x}(t_i)$ through its definition in (2.17), as each piece involved is

independent of $\mathbf{v}(\cdot)$. Thus, the autocorrelation matrix can be seen to be

$$\mathbf{E}\{\mathbf{z}(t_j)\,\mathbf{z}^{\mathrm{T}}(t_i)\} = \begin{cases} H(t_j)\,\mathbf{E}\{\mathbf{x}(t_j)\,\mathbf{x}^{\mathrm{T}}(t_i)\}H^{\mathrm{T}}(t_i) & t_i \neq t_j \\ H(t_j)\,\mathbf{E}\{\mathbf{x}(t_j)\,\mathbf{x}^{\mathrm{T}}(t_i)\}H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i) & t_i = t_j \end{cases} \tag{3.5}$$

The covariance kernel is obtained similarly to the methods used in the development of $\mathbf{P}_{xx}(\cdot)$ in equation (2.15), and results in

$$\mathbf{P}_{zz}(t_j, t_i) = \begin{cases} H(t_j)\,\mathbf{P}_{xx}\,H^{\mathrm{T}}(t_i) & t_i \neq t_j \\ H(t_j)\,\mathbf{P}_{xx}\,H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i) & t_i = t_j \end{cases} \tag{3.6}$$

Finally, it should be noted that $\mathbf{x}(t_0)$, $\boldsymbol{\beta}(\cdot,\cdot)$, $\mathbf{w}(\cdot,\cdot)$, and $\mathbf{v}(\cdot,\cdot)$ are independent of each other. The goal is to combine all measurement data starting from $\mathbf{z}(t_1)$ up to $\mathbf{z}(t_i)$ at the present time to obtain an optimal estimate of the system state. The definition of optimal can vary based on the criteria of the modeler, however by adopting a Bayesian reasoning, optimal will come from the conditional expectation and conditional probability density functions. Under the assumptions that will be stated again prior to developing this estimate, it will be shown that the conditional density will remain Gaussian at all times, and thus all measurement estimates of importance such as mean, mode, median and most any other reasonable choice for a guess will converge to the same estimate value.

Notation that will be convenient when deriving the estimate in regard to the measurements $\mathbf{z}(t_i)$ are as follows. The vector composed of the measurement history random variables from $\mathbf{z}(t_1)$ to $\mathbf{z}(t_i)$ will be given by

$$\mathbf{Z}_i = \mathbf{Z}(t_i) = \begin{bmatrix} \mathbf{z}(t_1) \\ \mathbf{z}(t_2) \\ \vdots \\ \mathbf{z}(t_i) \end{bmatrix} \tag{3.7}$$

while the vector of *realization* of measurement history will be written

$$Z_i = Z(t_i) = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_i \end{bmatrix} \tag{3.8}$$

Finally, the dummy variables that will be associated with $\mathbf{z}(t_i)$ and $\mathbf{Z}(t_i)$ are ζ_i and $\mathscr{Z}_i$, respectively.

The predictor-corrector form of the estimation will be seen by considering two measurement times, t_{i-1} and t_i. It will first consider an estimate from just after the measurement at time t_{i-1} which will be denoted t_{i-1}^+ to the time just before the measurement at t_i is incorporated (denoted t_i^-), then a measurement update from t_i^- to t_i^+ will make a correction for the noise. As before (under the usual assumptions) the deterministic control function will not affect the statistics (aside from the location of the density function), and will therefore be excluded from the initial computations for simplicity. Finally, because all statistics are assumed to be Gaussian at the start, a goal of the computations is to also show that at each stage t_{i-1}^+, t_i^- and t_i^+ the conditional density function is Gaussian. Then by inductive reasoning it will be clear that for all t_i, the conditional density can be completely described by its mean and covariance matrix.

The initial conditions are assumed to be Gaussian and the conditional density at time t_0 is equal to the density function at t_0 because no measurement has been taken yet, and therefore the conditional density is simply the density function for $\mathbf{x}(t_0)$ which was assumed to be Gaussian.

Given that a measurement $\mathbf{z}(t_{i-1}, \omega_i)$ just occurred at time t_i, then the first function of interest is the density function of $\mathbf{x}(t_{i-1})$ conditioned on $\mathbf{Z}_{i-1} = Z_{i-1}$. This will be the "$i-1$" step of the inductive proof, and thus will be assumed Gaussian, and can then be

written

$$f_{\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})}(\xi \mid Z_{i-1}) = \frac{1}{[(2\pi)^{n/2}|\mathbf{P}(t_{i-1}^+)|^{1/2}]} e^{-\frac{1}{2}[\xi-\hat{\mathbf{x}}(t_{i-1}^+)]^{\mathrm{T}}\,\mathbf{P}(t_{i-1}^+)^{-1}[\xi-\hat{\mathbf{x}}(t_{i-1}^+)]} \tag{3.9}$$

where $\hat{\mathbf{x}}(t_{i-1}^+)$ and $\mathbf{P}(t_{i-1}^+)$ are defined to be

$$\hat{\mathbf{x}}(t_{i-1}^+) = \mathbf{E}\{\mathbf{x}(t_{i-1}) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} \tag{3.10}$$

$$\mathbf{P}(t_{i-1}^+) = \mathbf{E}\{[\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)]^{\mathrm{T}}[\mathbf{x}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1}^+)] \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} \tag{3.11}$$

Then by the model, the state at time t_i is given by $\mathbf{x}(t_i) = \mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{x}(t_{i-1}) + \mathbf{w}_d(t_{i-1})$, where $\mathbf{w}_d(t_{i-1})$ is the equivalent discrete-time model

$$\mathbf{w}_d(t_{i-1}) = \int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i, \tau) G(\tau)\, \mathbf{d}\boldsymbol{\beta}(\tau)$$

therefore if it can be shown that the joint conditional density function of $\mathbf{x}(t_{i-1})$ and $\mathbf{w}_d(t_{i-1})$ is Gaussian, then by Theorem 2 the conditional density function of $\mathbf{x}(t_i)$ is Gaussian. Now Bayes' rule is applied to obtain

$$\begin{aligned} f_{\mathbf{x}(t_{i-1}),\mathbf{w}_d(t_{i-1})|\mathbf{Z}(t_{i-1})}(\xi, \rho \mid \mathscr{Z}_{i-1}) &= \frac{f_{\mathbf{x}(t_{i-1}),\mathbf{w}_d(t_{i-1}),\mathbf{Z}(t_{i-1})}(\xi, \rho, \mathscr{Z}_{i-1})}{f_{\mathbf{Z}(t_{i-1})}(\mathscr{Z}_{i-1})} \\ &= \frac{f_{\mathbf{x}(t_{i-1}),\mathbf{Z}(t_{i-1})}(\xi, \mathscr{Z}_{i-1}) f_{\mathbf{w}_d(t_{i-1})}(\rho)}{f_{\mathbf{Z}(t_{i-1})}(\mathscr{Z}_{i-1})} \\ &= f_{\mathbf{x}(t_{i-1})|\mathbf{Z}(t_{i-1})}(\xi \mid \mathscr{Z}_{i-1}) f_{\mathbf{w}_d(t_{i-1})}(\rho) \end{aligned}$$

The second equality comes from $\mathbf{w}_d(\cdot)$ being independent of $\mathbf{x}(t_i)$ and $\mathbf{Z}(t_i)$ by the independent increments of the Brownian process, and therefore can be peeled off of the joint density function. In the final equality, the conditional density function was assumed Gaussian, the process $\mathbf{w}_d(\cdot)$ is a white Gaussian process, and the product of two Gaussian density functions is Gaussian. Therefore, $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}(\xi \mid Z_{i-1})$ is Gaussian. Now its mean and co-

variance must be computed to write out the function explicitly. The following computations are made clear by recalling that $\mathbf{\Phi}(t_i, t_{i-1})$ is non random, and $\mathbf{w}_d(\cdot)$ is independent of any measurements, therefore its conditional expectation is simply its expectation.

$$\begin{aligned}
\mathbf{E}\{\mathbf{x}(t_i) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} &= \mathbf{E}\{\mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{x}(t_{i-1}) + \mathbf{w}_d(t_{i-1}) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} \\
&= \mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{E}\{\mathbf{x}(t_{i-1}) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} + \mathbf{E}\{\mathbf{w}_d(t_{i-1}) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} \\
&= \mathbf{\Phi}(t_i, t_{i-1})\,\hat{\mathbf{x}}(t_{i-1}) + \mathbf{E}\{\mathbf{w}_d(t_{i-1})\} \\
&= \mathbf{\Phi}(t_i, t_{i-1})\,\hat{\mathbf{x}}(t_{i-1})
\end{aligned}$$

Where the final equality comes from $\mathbf{w}_d(\cdot)$ being a zero mean white Gaussian process and thus independent of any measurements, and $\hat{\mathbf{x}}(t_{i-1}^+)$ from the equation (3.10). Thus, the optimal estimate for the mean prior to the measurement at t_i is defined to be the conditional expectation

$$\hat{\mathbf{x}}(t_i^-) = \mathbf{E}\{\mathbf{x}(t_i) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} = \mathbf{\Phi}(t_i, t_{i-1})\,\hat{\mathbf{x}}(t_{i-1}^+) \tag{3.12}$$

Now let $\mathbf{P}(t_i^-)$ be the conditional covariance of $\mathbf{x}(t_i)$ before the measurement $\mathbf{z}(t_i) = z_i$ is included. Then it can be found through the propagation of $\mathbf{P}(t_{i-1}^+)$ as follows

$$\begin{aligned}
\mathbf{P}(t_i^-) &= \mathbf{E}\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)][\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-)]^{\mathrm{T}} \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} \\
&= \mathbf{E}\{[\mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{x}(t_{i-1}) + \mathbf{w}_d(t_{i-1}) - \mathbf{\Phi}(t_i, t_{i-1})\,\hat{\mathbf{x}}(t_{i-1}^+)][\mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{x}(t_{i-1}) \\
&\quad + \mathbf{w}_d(t_{i-1}) - \mathbf{\Phi}(t_i, t_{i-1})\,\hat{\mathbf{x}}(t_{i-1}^+)]^{\mathrm{T}} \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} \\
&= \mathbf{E}\{[\mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{x}(t_{i-1})\,\mathbf{x}^{\mathrm{T}}(t_{i-1})\,\mathbf{\Phi}^{\mathrm{T}}(t_i, t_{i-1}) \\
&\quad - \mathbf{\Phi}(t_i, t_{i-1})\,\hat{\mathbf{x}}(t_{i-1})\,\hat{\mathbf{x}}^{\mathrm{T}}(t_{i-1})\,\mathbf{\Phi}^{\mathrm{T}}(t_i, t_{i-1}) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} \\
&\quad + \mathbf{E}\{\mathbf{w}_d(t_{i-1})\,\mathbf{w}_d^{\mathrm{T}}(t_{i-1}) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} \\
&= \mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{E}\{[\mathbf{x}(t_{i-1})\,\mathbf{x}^{\mathrm{T}}(t_{i-1}) - \hat{\mathbf{x}}(t_{i-1})\,\hat{\mathbf{x}}^{\mathrm{T}}(t_{i-1}) \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\}\,\mathbf{\Phi}^{\mathrm{T}}(t_i, t_{i-1}) \\
&\quad + \mathbf{E}\{\mathbf{w}_d(t_{i-1})\,\mathbf{w}_d^{\mathrm{T}}(t_{i-1})\}
\end{aligned}$$

Finally, recall that the expectation between the state transition matrices is the definition of $\mathbf{P}(t_{i-1}^+)$ from equation (3.11), and the mean square value of $\mathbf{w}_d(t)$ was given in Section 2.7. Thus, the covariance is propagated by

$$\mathbf{P}(t_i^-) = \mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{i-1}^+)\,\mathbf{\Phi}^{\mathrm{T}}(t_i, t_{i-1}) + \int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i, \tau) G(\tau)\,\mathbf{Q}(\tau) G^{\mathrm{T}}(\tau)\,\mathbf{\Phi}^{\mathrm{T}}(t_i, \tau) d\tau \quad (3.13)$$

To help in focusing on the model rather than all the computations, an intuitive understanding of the value $\mathbf{P}(t_i^-)$ is that it is the error committed by using the conditional mean $\hat{\mathbf{x}}(t_i^-)$ as the estimate of the state $\mathbf{x}(t_i)$.

Now the density function of $\mathbf{x}(t_i)$ conditioned on the measurements Z_{i-1} can be written

$$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}(\xi \mid Z_{i-1}) = \frac{1}{[(2\pi)^{n/2}|\mathbf{P}(t_i^-)|^{1/2}]} e^{-\frac{1}{2}[\xi - \hat{\mathbf{x}}(t_i^-)]^{\mathrm{T}}\,\mathbf{P}(t_i^-)^{-1}[\xi - \hat{\mathbf{x}}(t_i^-)]} \quad (3.14)$$

The next step is to incorporate the measurement that becomes available at time t_i to generate the density function of $\mathbf{x}(t_i)$ conditioned on $\mathbf{Z}(t_i)$ written $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}(\xi \mid Z_i)$. First, it must be shown that this function is in fact Gaussian. The conditional density function will first be written using Bayes' rule, and broken down piece by piece from there.

$$\begin{aligned}
f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)} &= \frac{f_{\mathbf{x}(t_i),\mathbf{Z}(t_i)}}{f_{\mathbf{Z}(t_i)}} \\
&= \frac{f_{\mathbf{x}(t_i),\mathbf{z}(t_i),\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i),\mathbf{Z}(t_{i-1})}} \\
&= \frac{f_{\mathbf{z}(t_i)|\mathbf{x}(t_i),\mathbf{Z}(t_{i-1})} f_{\mathbf{x}(t_i),\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})} f_{\mathbf{Z}(t_{i-1})}} \\
&= \frac{f_{\mathbf{z}(t_i)|\mathbf{x}(t_i),\mathbf{Z}(t_{i-1})} f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})} f_{\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})} f_{\mathbf{Z}(t_{i-1})}} \\
&= \frac{f_{\mathbf{z}(t_i)|\mathbf{x}(t_i),\mathbf{Z}(t_{i-1})} f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}}{f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}}
\end{aligned} \quad (3.15)$$

To determine if the initial conditional density function is Gaussian, each of the conditional density functions in the final equality will have to be analyzed.

First, note that the second density function in the numerator has already been determined to be Gaussian, and has been written explicitly in equation (3.14). The first numerator term is simpler than it looks at first glance. Note that it is the density of the measurement function at time t_i given the *known* state $\mathbf{x}(t_i) = \xi$ and all previous measurements. Recall that the measurement at time t_i is given by $\mathbf{z}(t_i) = H(t_i)\,\mathbf{x}(t_i) + \mathbf{v}(t_i)$, where $\mathbf{v}(t_i)$ is independent of $\mathbf{x}(t_i)$ and $\mathbf{Z}(t_i)$ and is assumed to be a zero mean white Gaussian process. Therefore, $\mathbf{z}(t_i)$ is the linear combination of a known matrix $H(t_i)$, a realized (and therefore known) vector $\mathbf{x}(t_i) = \xi$, and the zero mean white Gaussian vector, and is therefore Gaussian. To write out its density function, the mean and covariance will be needed.

$$\begin{aligned}\mathbf{E}\{\mathbf{z}(t_i) \mid \mathbf{x}(t_i) = \xi, \mathbf{Z}(t_{i-1}) = \mathscr{Z}_{i-1}\} &= H(t_i)\,\mathbf{E}\{\mathbf{x}(t_i) \mid \mathbf{x}(t_i) = \xi, \mathbf{Z}_{t_{i-1}} = \mathscr{Z}_{i-1}\} \\ &\quad + \mathbf{E}\{\mathbf{v}(t_i) \mid \mathbf{x}(t_i) = \xi, \mathbf{Z}_{t_{i-1}} = \mathscr{Z}_{i-1}\} \\ &= H(t_i)\xi\end{aligned}$$

$$\begin{aligned}&\mathbf{E}\{[\mathbf{z}(t_i) - H(t_i)\xi][\mathbf{z}(t_i) - H(t_i)\xi]^{\mathrm{T}} \mid \mathbf{x}(t_i) = \xi, \mathbf{Z}_{t_{i-1}} = \mathscr{Z}_{i-1}\} \\ &\qquad = \mathbf{E}\{\mathbf{z}(t_i)\,\mathbf{z}^{\mathrm{T}}(t_i)\} - H(t_i)\xi\xi^{\mathrm{T}}H^{\mathrm{T}}(t_i)\} \\ &\qquad = H(t_i)\xi\xi^{\mathrm{T}}H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i) - H(t_i)\xi\xi^{\mathrm{T}}H^{\mathrm{T}}(t_i) \\ &\qquad = \mathbf{R}(t_i)\end{aligned}$$

where the final equality comes from equation (3.6). Therefore, the conditional density function can be written explicitly as

$$f_{\mathbf{z}(t_i)|\mathbf{x}(t_i),\mathbf{Z}(t_{i-1})}(\zeta_i \mid \xi, \mathscr{Z}_{i-1}) = \frac{1}{(2\pi)^{m/2}|\,\mathbf{R}(t_i)|^{1/2}} e^{-\frac{1}{2}[\zeta_i - H(t_i)\xi]^{\mathrm{T}}\,\mathbf{R}(t_i)^{-1}[\zeta_i - H(t_i)\xi]} \tag{3.16}$$

Finally, the denominator of (3.15) is the density function of the measurement at $\mathbf{z}(t_i)$ conditioned on all the previous measurements $\mathbf{Z}(t_{i-1})$. Intuitively, it is asking "What mea-

surement are we likely to get given the trend from the last "$i-1$" measurements. Again, by writing out $\mathbf{z}(t_i)$ as the linear combination of the vector $\mathbf{x}(t_i)$ and the zero mean white Gaussian process $\mathbf{v}(t_i)$, it can be shown to be Gaussian by analyzing its pieces.

$$\begin{aligned} f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}(\xi \mid \mathscr{Z}_{i-1}) &= f_{\mathbf{x}(t_i),\mathbf{v}(t_i)|\mathbf{Z}(t_{i-1})}(\xi,\nu \mid \mathscr{Z}_{i-1}) \\ &= \frac{f_{\mathbf{x}(t_i),\mathbf{v}(t_i),\mathbf{Z}(t_{i-1})}(\xi,\nu,\mathscr{Z}_{i-1})}{f_{\mathbf{Z}(t_{i-1})}(\mathscr{Z}_{i-1})} \\ &= \frac{f_{\mathbf{x}(t_i),\mathbf{Z}(t_{i-1})}(\xi,\mathscr{Z}_{i-1}) f_{\mathbf{v}(t_i)}(\nu)}{f_{\mathbf{Z}(t_{i-1})}(\mathscr{Z}_{i-1})} \\ &= f_{\mathbf{x}(t_i)|\mathbf{Z}(t_{i-1})}(\xi \mid \mathscr{Z}_{i-1}) f_{\mathbf{v}(t_i)}(\nu) \end{aligned}$$

The first conditional density function has been shown to be Gaussian, and the second function is the white Gaussian process, therefore their product is Gaussian. Again, the mean and covariance are needed to write the explicit conditional density function.

$$\mathbf{E}\{\mathbf{z}(t_i) \mid \mathbf{Z}(t_{i-1}) = \mathscr{Z}_{i-1}\} = H(t_i)\,\mathbf{E}\{\mathbf{x}(t_i) \mid \mathbf{Z}(t_{i-1}) = \mathscr{Z}_{i-1}\} + \mathbf{E}\{\mathbf{v}(t_i)\} = H(t_i)\,\hat{\mathbf{x}}(t_i^-) \tag{3.17}$$

$$\begin{aligned} &\mathbf{E}\{[\mathbf{z}(t_i) - H(t_i)\,\hat{\mathbf{x}}(t_i^-)][\mathbf{z}(t_i) - H(t_i)\,\hat{\mathbf{x}}(t_i^-)]^{\mathrm{T}} \mid \mathbf{Z}(t_{i-1}) = \mathscr{Z}_{i-1}\} \\ &\quad = \mathbf{E}\{\mathbf{z}(t_i)\,\mathbf{z}^{\mathrm{T}} - H(t_i)\,\hat{\mathbf{x}}(t_i^-)\,\hat{\mathbf{x}}^{\mathrm{T}}(t_i^-)H^{\mathrm{T}}(t_i) \mid \mathbf{Z}(t_{i-1}) = \mathscr{Z}_{i-1}\} \\ &\quad = \mathbf{E}\{H(t_i)\,\mathbf{x}(t_i)\,\mathbf{x}^{\mathrm{T}}(t_i)H^{\mathrm{T}}(t_i) + \mathbf{v}(t_i)\,\mathbf{v}^{\mathrm{T}}(t_i) \\ &\qquad - H(t_i)\,\hat{\mathbf{x}}(t_i^-)\,\hat{\mathbf{x}}^{\mathrm{T}}(t_i^-)H^{\mathrm{T}}(t_i) \mid \mathbf{Z}(t_{i-1}) = \mathscr{Z}_{i-1}\} \\ &\quad = H(t_i)\,\mathbf{E}\{\mathbf{x}(t_i)\,\mathbf{x}^{\mathrm{T}}(t_i) \mid \mathbf{Z}(t_{i-1}) = \mathscr{Z}_{i-1}\}H^{\mathrm{T}}(t_i) \\ &\qquad + \mathbf{E}\{\mathbf{v}(t_i)\,\mathbf{v}^{\mathrm{T}}(t_i)\} - H(t_i)\,\hat{\mathbf{x}}(t_i^-)\,\hat{\mathbf{x}}^{\mathrm{T}}(t_i^-)H^{\mathrm{T}}(t_i) \\ &\quad = H(t_i)[\mathbf{P}(t_i^-) + \hat{\mathbf{x}}(t_i^-)\,\hat{\mathbf{x}}^{\mathrm{T}}(t_i^-)]H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i) - H(t_i)\,\hat{\mathbf{x}}(t_i^-)\,\hat{\mathbf{x}}^{\mathrm{T}}(t_i^-)H^{\mathrm{T}}(t_i) \end{aligned}$$

Distributing the matrices $H(t_i)$ and $H^{\mathrm{T}}(t_i)$ and canceling like terms yields

$$\mathbf{E}\{[\mathbf{z}(t_i) - H(t_i)\,\hat{\mathbf{x}}(t_i^-)][\mathbf{z}(t_i) - H(t_i)\,\hat{\mathbf{x}}(t_i^-)]^{\mathrm{T}} \mid \mathbf{Z}(t_{i-1}) = Z_{i-1}\} = H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i) \tag{3.18}$$

Thus, the conditional density function of $\mathbf{z}(t_i)$ conditioned on all previous measurements can be written

$$f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1})}(\zeta_i \mid \mathscr{Z}_{i-1}) = \frac{1}{(2\pi)^{m/2}|H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)|^{1/2}} \times e^{-\frac{1}{2}[\zeta_i - H(t_i)\,\hat{\mathbf{x}}(t_i^-)]^{\mathrm{T}}[H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i)+\mathbf{R}(t_i)]^{-1}[\zeta_i - H(t_i)\,\hat{\mathbf{x}}(t_i^-)]} \tag{3.19}$$

A quick check confirms that $H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)$ is invertible. First, $\mathbf{R}(t_i)$ is symmetric positive definite, and $\mathbf{P}(t_i^-)$ is positive definite, so $H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i)$ is at least positive semidefinite. Thus, the sum is positive definite, and therefore invertible. Now that all conditional density functions have been determined to be Gaussian and have been written explicitly, (3.14) , (3.16), and (3.19) can be substituted into the original equation to obtain the conditional density function

$$f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}(\xi \mid \mathscr{Z}_i) = \frac{|H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)|^{1/2}}{(2\pi)^{n/2}|\,\mathbf{P}(t_i^-)|^{1/2}|\,\mathbf{R}(t_i)|^{1/2}} e^{\{-\frac{1}{2}*\}}$$

$$\begin{aligned}\text{where} \qquad \{*\} = \{&[\zeta_i - H(t_i)\xi]^{\mathrm{T}}\,\mathbf{R}(t_i)^{-1}[\zeta_i - H(t_i)\xi] \\ &+ [\xi - \hat{\mathbf{x}}(t_i^-)]^{\mathrm{T}}\,\mathbf{P}(t_i^-)^{-1}[\xi - \hat{\mathbf{x}}(t_i^-)] \\ &- [\zeta_i - H(t_i)\,\hat{\mathbf{x}}(t_i^-)]^{\mathrm{T}}[H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)]^{-1}[\zeta_i - H(t_i)\,\hat{\mathbf{x}}(t_i^-)]\}\end{aligned} \tag{3.20}$$

A considerable amount of matrix manipulation will be required to work this function into that of a Gaussian function defined explicitly in terms of one mean and one covariance. The first tool required will be the matrix inversion lemma

Theorem 3 (Matrix Inversion Lemma)**.** If the matrices P and R are invertible, and QRS

has the same dimensions as P, then $P + QRS$ is invertible if and only if $R^{-1} + SP^{-1}Q$ is invertible, and $(P + QRS)^{-1} = P^{-1} - P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}SP^{-1}$.

Proof. First, suppose $R^{-1} + SP^{-1}Q$ is singular, then there exists $x \neq 0$ such that

$$\begin{aligned}(R^{-1} + SP^{-1}Q)x &= 0 \\ RR^{-1}x + RSP^{-1}Qx &= 0 \\ RSP^{-1}Qx &= -x\end{aligned}$$

Let $y = P^{-1}Qx$, then $RSy = -x \neq 0$ therefore $y \neq 0$ and

$$\begin{aligned}(P + QRS)y &= Py + QRSy \\ &= Qx + Q(-x) = 0\end{aligned}$$

Thus, $P + QRS$ is also singular. Now, assume $P + QRS$ is singular. Define $\tilde{P} = R^{-1}$, $\tilde{R} = P^{-1}$, $\tilde{Q} = S$, and $\tilde{S} = Q$, then the previous argument can be applied again, and shows that $\tilde{P} + \tilde{Q}\tilde{R}\tilde{S} = R^{-1} + SP^{-1}Q$ is singular.

Now, to show that $(P + QRS)^{-1} = P^{-1} - P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}SP^{-1}$, the product

will be computed.

$$
\begin{aligned}
(P + QRS)(P^{-1} - P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}SP^{-1}) &= \\
&= I - Q(R^{-1} + SP^{-1}Q)^{-1}SP^{-1} + QRSP^{-1} \\
&\qquad - QRSP^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}SP^{-1} \\
&= I - Q[(R^{-1} + SP^{-1}Q)^{-1} - R + RSP^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}]SP^{-1} \\
&= I - QR[R^{-1}(R^{-1} + SP^{-1}Q)^{-1} + SP^{-1}Q(R^{-1} + SP^{-1}Q)^{-1} - I] \\
&= I - QR[(R^{-1} + SP^{-1}Q)(R^{-1} + SP^{-1}Q)^{-1} - I] \\
&= I - QR[I - I] = I
\end{aligned}
$$

Thus, the inverse is confirmed. □

For simplicity in the following computations, all time values and measurement indicators will be left off. This will not cause an issue, as all measurements are occurring at the same instant. They will be substituted back in when presenting their finished forms.

Three equations which make use of the Matrix Inversion Lemma that will be of particular use in the computations are as follows:

$$[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]^{-1} = \mathbf{P} - \mathbf{P}\,H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1}H\,\mathbf{P} \tag{3.21}$$

$$[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]^{-1}H^{\mathrm{T}}\,\mathbf{R}^{-1} = \mathbf{P}\,H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1} \tag{3.22}$$

$$H[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]^{-1}H^{\mathrm{T}} = \mathbf{R} - \mathbf{R}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1}\,\mathbf{R} \tag{3.23}$$

The first goal will be to reduce the exponential term of the distribution function to a single form. First, the three quadratic terms can be expanded and combined into the

following

$$\begin{aligned}\{*\} = \xi^{\mathrm{T}}[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]\xi - 2\xi^{\mathrm{T}}[\mathbf{P}^{-1}\,\hat{\mathbf{x}}^{-} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,\zeta_i] \\ + \zeta_i^{\mathrm{T}}[\mathbf{R}^{-1} - (H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R})^{-1}]\zeta_i + 2\,\hat{\mathbf{x}}^{-\,T} H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1}\zeta_i \\ + \hat{\mathbf{x}}^{-\mathrm{T}}[\mathbf{P}^{-1} - H^{\mathrm{T}}(H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R})^{-1}H]\,\hat{\mathbf{x}}^{-}\end{aligned}$$

A motivation for the following computations is to obtain matrices of the form

$$(\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H)^{-1}$$

This piece will be very important throughout the remainder of the report.

First, consider the third term $\zeta_i^T[\mathbf{R}^{-1} - (H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R})^{-1}]\zeta_i$. Applying the inversion lemma to the matrix yields the following

$$\begin{aligned}\mathbf{R}^{-1} - [H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1} &= \mathbf{R}^{-1} - [\mathbf{R}^{-1} - \mathbf{R}^{-1}\,H(\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H)^{-1}H^{\mathrm{T}}\,\mathbf{R}^{-1}] \\ &= \mathbf{R}^{-1}\,H(\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H)^{-1}H^{\mathrm{T}}\,\mathbf{R}^{-1}\end{aligned}$$

and this term can therefore be written $\zeta_i^{\mathrm{T}}\,\mathbf{R}^{-1}\,H(\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H)^{-1}H^{\mathrm{T}}\,\mathbf{R}^{-1}\,\zeta_i$.

Next, consider the fourth term $2\,\hat{\mathbf{x}}^{-\,T} H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + R]^{-1}\zeta_i$ which is equal to (3.22) if is pre-multiplied by $\mathbf{P}^{-1}$. Therefore, it can be rewritten

$$\begin{aligned}H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1} &= \mathbf{P}^{-1}\,\mathbf{P}\,H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1} \\ &= \mathbf{P}^{-1}[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]^{-1}\,\mathbf{P}^{-1}\end{aligned}$$

to yield $2\,\hat{\mathbf{x}}^{-\,T}\,\mathbf{P}^{-1}[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]^{-1}H^{\mathrm{T}}\,\mathbf{R}^{-1}\,\zeta_i$

Now the fifth term is equal to (3.21) once pre and post multiplied by $\mathbf{P}^{-1}$, and can

therefore be rewritten as follows

$$\begin{aligned}\mathbf{P}^{-1}[\mathbf{P}^{-1} - H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1}H]\,\mathbf{P}^{-1} &= \\ &= \mathbf{P}^{-1}\,\mathbf{P}\,\mathbf{P}^{-1} - \mathbf{P}^{-1}\,\mathbf{P}\,H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1}H\,\mathbf{P}\,\mathbf{P}^{-1} \\ &= \mathbf{P}^{-1}[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]^{-1}\,\mathbf{P}^{-1}\end{aligned}$$

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

Finally, the third, fourth, and fifth terms can be combined from the original sum to the equivalent form

$$\begin{aligned}(*) = \xi^{\mathrm{T}}[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]\xi - 2\xi^{\mathrm{T}}[\mathbf{P}^{-1}\,\hat{\mathbf{x}}^{-} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,\zeta_i] \\ + \ [\zeta_i^{\mathrm{T}}\,\mathbf{R}^{-1}\,H + \hat{\mathbf{x}}^{-\mathrm{T}}\,\mathbf{P}^{-1}][\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]^{-1}[H^{\mathrm{T}}\,\mathbf{R}^{-1}\,\zeta_i + \mathbf{P}^{-1}\,\hat{\mathbf{x}}^{-}]\end{aligned}$$

Now define $\mathbf{A} = [\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]$ and $\mathbf{a} = [\mathbf{P}^{-1}\,\hat{\mathbf{x}}^{-} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,\zeta_i]$, then $(*)$ becomes

$$\begin{aligned}(*) &= \xi^{\mathrm{T}}\,\mathbf{A}\,\xi - 2\xi^{\mathrm{T}}\,\mathbf{a} + \mathbf{a}^{\mathrm{T}}\,\mathbf{A}^{-1}\,\mathbf{a} \\ &= \xi^{\mathrm{T}}\,\mathbf{A}\,\xi - 2\xi^{\mathrm{T}}\,\mathbf{A}\,\mathbf{A}^{-1}\,\mathbf{a} + \mathbf{a}^{\mathrm{T}}\,\mathbf{A}^{-1}\,\mathbf{A}\,\mathbf{A}^{-1}\,\mathbf{a} \\ &= (\xi - \mathbf{A}^{-1}\,\mathbf{a})^{\mathrm{T}}\,\mathbf{A}(\xi - \mathbf{A}^{-1}\,\mathbf{a})\end{aligned}$$

To see the reduction of the determinant fraction to $|\,\mathbf{A}^{-1}\,|$, consider the matrix relation (3.22), and take the determinant of both sides, then apply the properties of the

determinant as follows

$$
\begin{aligned}
[\mathbf{P}^{-1} + H^{\mathrm{T}}\,\mathbf{R}^{-1}\,H]^{-1} H^{\mathrm{T}}\,\mathbf{R}^{-1} &= \mathbf{P}\,H^{\mathrm{T}}[H\,\mathbf{P}\,H^{\mathrm{T}} + \mathbf{R}]^{-1} \\
\frac{|H|}{|\,\mathbf{P}^{\text{-}1} + H^{\mathrm{T}}\,\mathbf{R}^{\text{-}1}\,H|\,|\,\mathbf{R}\,|} &= \frac{|\,\mathbf{P}\,|\,|H|}{|H\,\mathbf{P}\,H^{\mathrm{T}} + R|} \\
\frac{|H\,\mathbf{P}\,H^{\mathrm{T}} + R|}{|\,\mathbf{P}\,|\,|\,\mathbf{R}\,|} &= |\,\mathbf{P}^{\text{-}1} + H^{\mathrm{T}}\,\mathbf{R}^{\text{-}1}\,H| \\
&= \frac{1}{|[\mathbf{P}^{\text{-}1} + H^{\mathrm{T}}\,\mathbf{R}^{\text{-}1}\,H]^{-1}|} \\
&= \frac{1}{|\,\mathbf{A}^{-1}\,|}
\end{aligned}
$$

Thus, because $f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}(\xi \mid Z_i)$ was shown to be Gaussian, it is clear that is has mean $\mathbf{A}^{-1}\,\mathbf{a}$ and covariance $\mathbf{A}^{-1}$ written again below

$$
\begin{aligned}
\hat{\mathbf{x}}(t_i^+) &= \mathbf{E}\{\mathbf{x}(t_i) \mid \mathbf{Z}(t_i) = Z_i\} = \mathbf{A}^{-1}\,\mathbf{a} \\
\mathbf{P}(t_i^+) &= \mathbf{E}\{[\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)][\mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^+)]^{\mathrm{T}} \mid \mathbf{Z}(t_i) = Z_i\} \\
&= [H(t_i)P(t_i^-)H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)]^{-1} = \mathbf{A}^{-1} \\
f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i)}(\xi \mid \mathbf{Z}_i) &= \frac{1}{(2\pi)^{n/2}|\mathbf{P}(t_i^+)|^{1/2}} e^{-\frac{1}{2}[\xi - \hat{\mathbf{x}}(t_i^+)]^{\mathrm{T}}\,\mathbf{P}(t_i^+)^{-1}[\xi - \hat{\mathbf{x}}(t_i^+)]}
\end{aligned}
$$

Lastly, the deterministic control function can be included in the initial system to complete the model. As was mentioned before, because there is no uncertainty added from the control function, the covariance matrix is unaffected. The only change takes place in the mean and through that, the conditional density function by shifting the location of the mean.

$$
\hat{\mathbf{x}}(t_i^+) = \mathbf{A}^{-1}\,\mathbf{a} + \int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i, \tau) B(\tau)\,\mathbf{u}(\tau) d\tau
$$

It should be noted that at every time step, the positive definite property of the matrix $\mathbf{P}(t_i^+)$ is required. An induction proof is provided to guarantee this holds. Recall that $\mathbf{P}(t_0) = \mathbf{P}(t_0^+)$ is assumed symmetric positive definite. Let $\mathbf{P}(t_{i-1}^+)$ be positive definite, and

recall from (3.13) that the covariance is propagated by

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

The first term is positive definite by the properties of the state transition matrix $\mathbf{\Phi}(\cdot, \cdot)$, and the integral is at least positive semidefinite due to $\mathbf{Q}(\tau)$ being positive definite. Therefore $\mathbf{P}(t_i^-)$ is positive definite. Then, $\mathbf{P}(t_i^+) = [H(t_i)\,\mathbf{P}(t_i^-) H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)]^{-1}$, where $\mathbf{R}(t_i)$ was assumed positive definite, and $H(t_i)\,\mathbf{P}(t_i^-) H^{\mathrm{T}}(t_i)$ is at least positive semidefinite, so their sum must be positive definite, and therefore the inverse is as well. Thus, the covariance matrix is positive definite at any time step.

3.2 The Kalman Filter Algorithm

As the previous section has shown, all of the conditional density functions of interest are Gaussian, and fairly simple equations have been presented which detail the propagation of the mean and the covariance. The algorithm itself has conditioned an estimate of the state of the system based on all previous measurements, and thus, satisfied the objective of an optimal estimate according to Bayesian criteria. Below, the *Kalman filter gain* will be defined, and a summary of the Kalman filter algorithm just derived will be summarized.

Prior to the measurement at time t_i, the mean and covariance matrix are given by

$$\hat{\mathbf{x}}(t_i^-) = \mathbf{\Phi}(t_i, t_{i-1})\,\hat{\mathbf{x}}(t_{i-1}^+) + \int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i, \tau) B(\tau)\,\mathbf{u}(\tau) d\tau \tag{3.24}$$

$$P(t_i^-) = \mathbf{\Phi}(t_i, t_{i-1})\,\mathbf{P}(t_{i-1}^+)\,\mathbf{\Phi}(t_i, t_{i-1}) + \int_{t_{i-1}}^{t_i} \mathbf{\Phi}(t_i \tau) G(\tau)\,\mathbf{Q}(\tau) G^{\mathrm{T}}(\tau)\,\mathbf{\Phi}^{\mathrm{T}}(t_i, t_{i-1}) \tag{3.25}$$

Then, when the measurement $\mathbf{z}(t_i)$ becomes available, the estimate is updated through the

Kalman filter gain $\mathbf{K}(t_i)$ (defined below) and is applied in both the mean and covariance.

$$\mathbf{K}(t_i) = \mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i)[H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)]^{-1} \tag{3.26}$$

$$\hat{\mathbf{x}}(t_i^+) = \hat{\mathbf{x}}(t_i^-) + \mathbf{K}(t_i)[\mathbf{z}_i - H(t_i)\,\hat{\mathbf{x}}(t_i^-)] \tag{3.27}$$

$$\mathbf{P}(t_i^+) = \mathbf{P}(t_i^-) - \mathbf{K}(t_i)H(t_i)\,\mathbf{P}(t_i^-) \tag{3.28}$$

with the initial conditions for the recursion

$$\hat{\mathbf{x}}(t_0) = \mathbf{E}\{\mathbf{x}(t_0)\} = \hat{\mathbf{x}}_0 \tag{3.29}$$

$$\mathbf{P}(t_0) = \{[\mathbf{x}(t_0) - \hat{\mathbf{x}}_0][\mathbf{x}(t_0) - \hat{\mathbf{x}}_0]^{\mathrm{T}}\} = \mathbf{P}_0 \tag{3.30}$$

Figure 3.1 depicts the Kalman Filter Algorithm stepping from time $k-1$ to k.

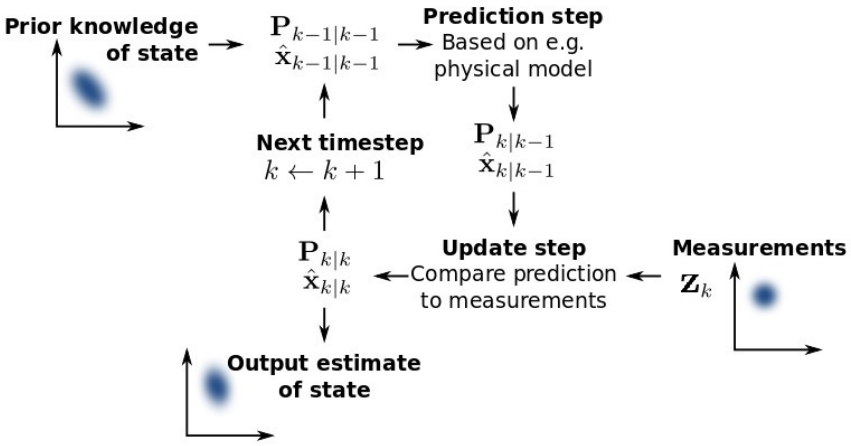

Figure 3.1: The Kalman Filter Algorithm (c) Petteri Aimonen (http://en.wikipedia.org/wiki/File:Basic_concept_of_Kalman_filtering.svg). Used under Creative Commons license CC0 1.0.

With the Kalman filter defined, it is much easier to see how the correction portion works. In the update of the mean, the *residual* (the difference between the guess being passed

through the measurement matrix and the measurement itself), is passed through the optimal weighting matrix $\mathbf{K}(t_i)$ to generate a correction term, then added to the best guess that existed before that. Though it is not as easy to see, given any measurement, the covariance is (generally) reduced by the measurement matrix $H(t_i)$ acting on the covariance matrix $P(t_i^-)$ which determines which portions of the model can be improved (and how), then the Kalman filter weighting these measurements and subtracting from $\mathbf{P}(t_i^-)$ (the error committed by using the conditional mean $\hat{\mathbf{x}}(t_i^-)$ as the estimate of the state $\mathbf{x}(t_i)$). Another thought which can help to understand how the Kalman filter improves the model is found by considering the strength of the diffusion matrix $\mathbf{R}(t_i)$. If $\mathbf{R}(t_i)$ is very small, then $[H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i) + \mathbf{R}(t_i)]^{-1}$ is relatively close to $[H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i)]^{-1}$, and substituting that into $\mathbf{K}(t_i)$ and applying it to the model shows that the $\hat{\mathbf{x}}(t_i)$ is nearly $\mathbf{x}(t_i)$ and the covariance is nearly zero. Conversely, if $\mathbf{R}(t_i)$ is very large, then $[H(t_i)\,\mathbf{P}(t_i^-)H^{\mathrm{T}}(t_i)+\mathbf{R}(t_i)]^{-1}$ is relatively small, and little contribution is made by applying the Kalman filter.

Chapter 4

A Oscillatory Example

To see the how the Kalman Filter Algorithm improves the estimate of a measurement, a simple example was constructed in $\mathbb{R}^n$. F was chosen as a simple matrix to guarantee complex eigenvalues with real parts equal to zero. Thus, the system will have oscillatory behavior. All other values for the system

$$\mathbf{dx}(t) = F(t)\,\mathbf{x}(t)dt + B(t)\,\mathbf{u}(t)dt + G(t)\,\mathbf{d}\boldsymbol{\beta}(t)$$
$$\mathbf{z}(t_i) = H(t_i)\,\mathbf{x}(t_i) + \mathbf{v}(t_i)$$

are defined as follows:

$$F = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad \mathbf{x}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} 0.01 & -0.0025 \\ -0.0025 & 0.02 \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \qquad H = \begin{bmatrix} 1 & 1 \end{bmatrix} \qquad R = 0.1$$

Note that the only measurement available for the system is the sum of the two random variables, corrupted by some white Gaussian noise. The following images show the evolution of the system $\mathbf{x}(t)$ throughout time steps of size $\triangle t = 0.2$. In the first two figures, the exact state of the system is shown as a black circle "o", the estimate prior to the inclusion of the measurement is a red asterisk "*", and the estimate including the measurement through the Kalman Filter is visible as a blue plus sign "+".

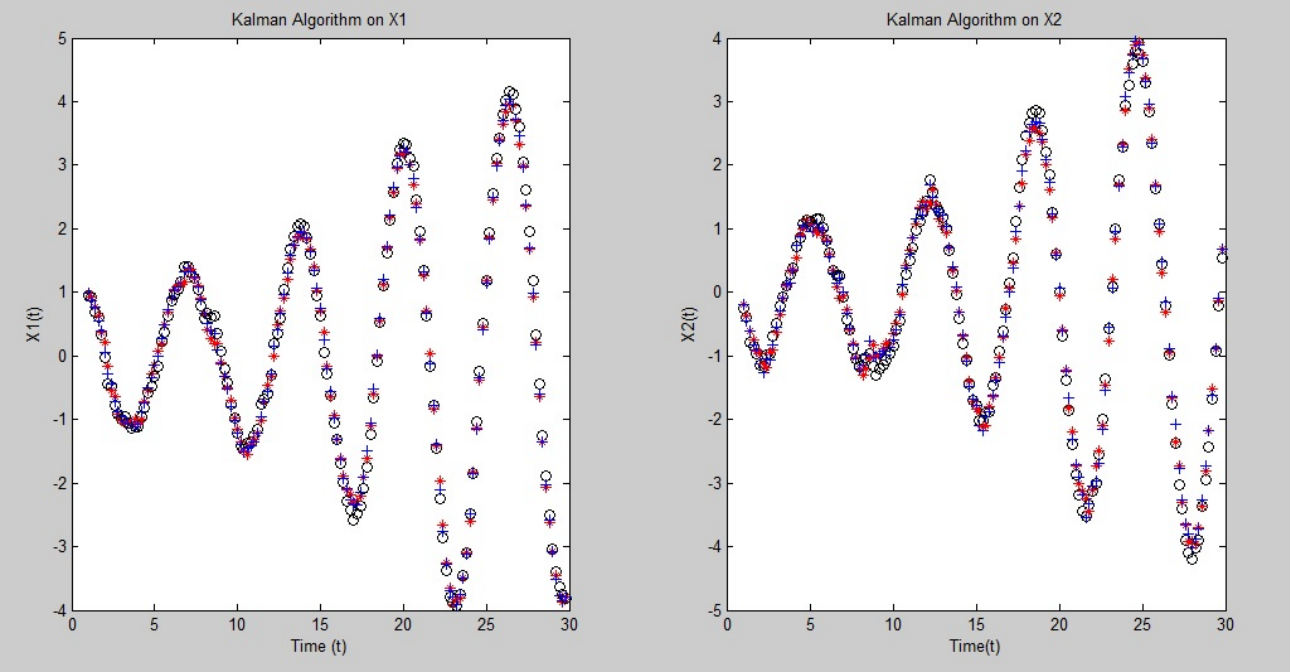

Figure 4.1: Unmeasured Estimate vs. Kalman Filter

In Figure 4.1, it is visible that both the estimate prior to the measurement, and the estimate including the measurement follow the true value of the system fairly well. Figure 4.2 zooms into a sample of the model to show that in general, the updated measurement is significantly closer than the prior estimate. Occasionally, there is an outlier where the prior estimate lands closer to the actual state. However, with the random noise in the system it is bound to happen occasionally.

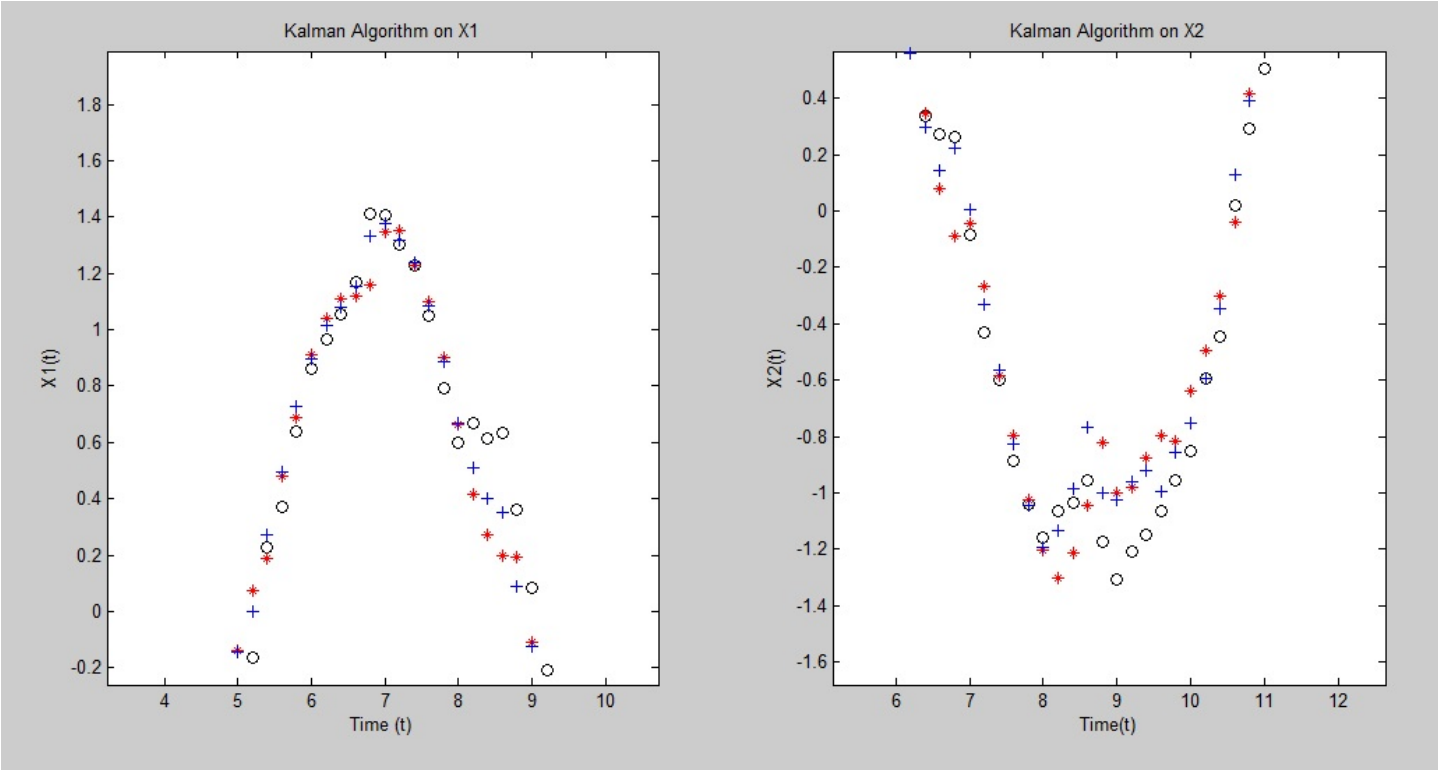

Figure 4.2: Unmeasured Estimate vs. Kalman Filter Zoomed

Finally, an additional program was written that simulates a "noisy orbit" of the system, as well as the Kalman Filter estimate. In Figure 4.3, the green asterick "*" shows the true location of $\mathbf{x}(t)$, while the blue dot "." show the estimated location with absolutely no measurements involved, with the blue curve indicating approximately two standard deviations away (most likely to contain the true location). Finally, the red dot "." indicates the estimated state incorporating all available measurements, with the red curve similarly showing two standard deviations from the updated measurement. As the system evolved, several iterations were selected (indicated in their title), which show the evolution of the estimate.

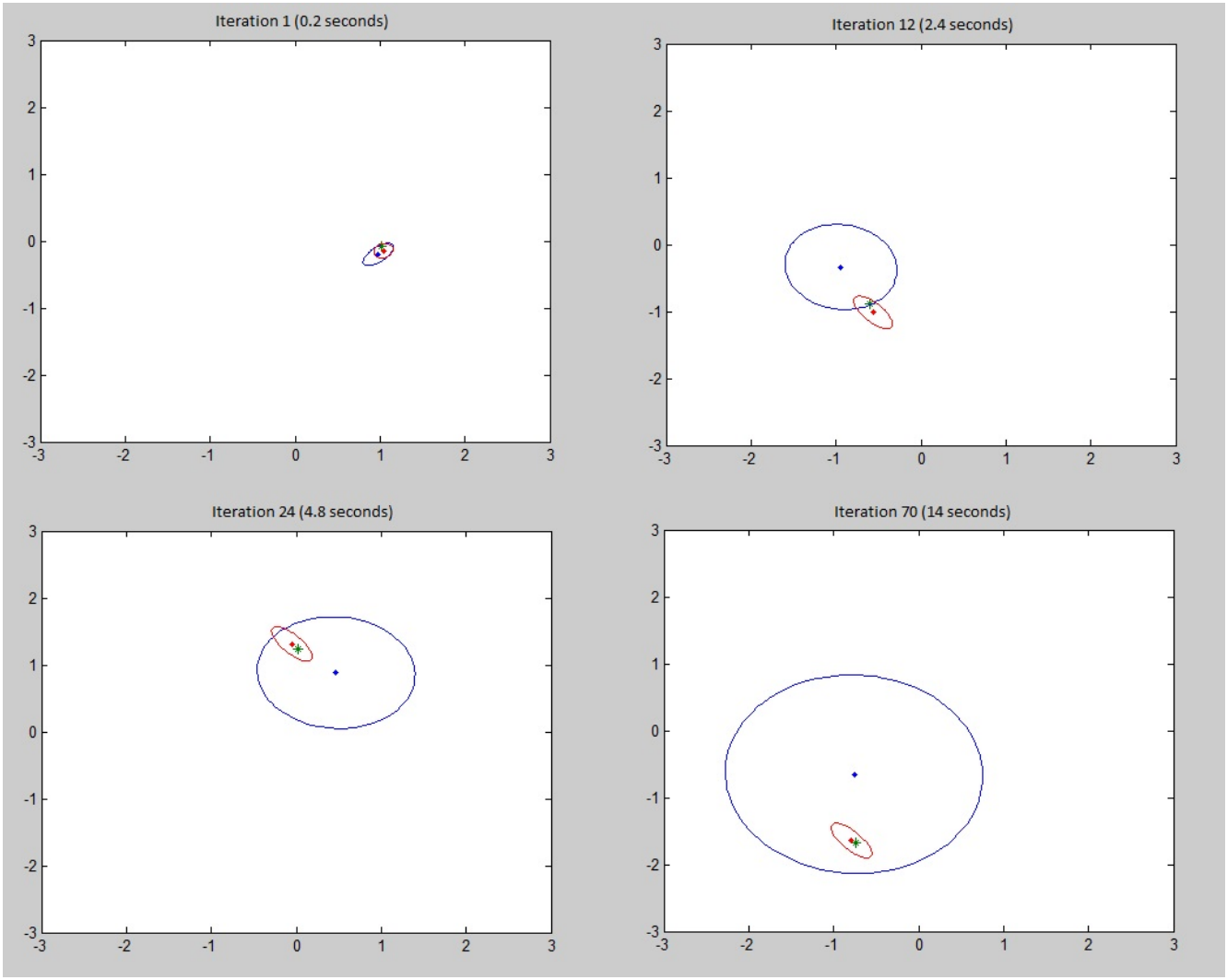

Figure 4.3: Noisy Orbit of the System

As time progresses, the unmeasured estimate quickly loses track of the true location, while the updated measurement hardly varies from it. What's even more interesting, is that the measured estimate is able to do so despite being a fairly poor measurement. In fact, the estimate is accurate throughout extended time periods as the model progresses.

Chapter 5

Conclusion

While deterministic models can provide great insight into real world problems, their lack of an error term limits the realism of the model. This motivated the development and analysis of the stochastic model. First, a background in probability theory was required to define a stochastic process and the measurements available to it. With access to stochastic processes, Gaussian noise could be modeled by ensuring that the process was white (independent) and the joint density functions were normally distributed for any time partition. However, the infinite variance of the white noise prevented the model from being Riemann integrable. To present a solvable differential equation, Brownian processes were defined, and the relation $d\boldsymbol{\beta}(t) = \mathbf{w}(t)dt$ was implemented to define and apply to stochastic integrals. With the existence of a stochastic integral, stochastic differentials could then be defined and substituted into the model to form the stochastic differential equation. With the model defined and a solution available, the question then became "How can a best estimate be made?". Through the matrix inversion lemma and a significant amount of manipulation, it was shown that the conditional density function of the process conditioned on all available measurements was in fact, still Gaussian, and therefore Bayesian reasoning determined that the best estimate for the state of the system at time t_i was the conditional expectation of the state at time t_i. The Kalman Gain matrix $\mathbf{K}(t_i)$ was then defined, and implemented in

the Kalman Filter Algorithm, which makes a best guess before a measurement occurs, then a weighted correction after the measurement is available. Upon inspection of the algorithm and through the oscillatory example, it is visible that any measurement (even a poor one) improves the accuracy of the state equation estimate.

The next step I would like to pursue in stochastic modeling would be observability and controllability of stochastic models. Both are very important topics in deterministic system theory, and have similar extensions in stochastic models. I imagine with the existence of the noise function, both properties will be much more difficult to obtain, and will require interesting properties to do so.

Bibliography

[1] L Perko. *Differential equations and dynamical systems*. Springer-Verlag, New York, 1991.

[2] Geoffrey Grimmett and David Stirzaker. Probability and random processes. 1992.

[3] Peter S Maybeck. *Stochastic models, estimation, and control*, volume 3. Academic press, 1982.

[4] John K. Hunter. Lecture notes on applied mathematics. https://www.math.ucdavis.edu/~hunter/m280_09/ch.pdf, 2009.

Appendix A

The Fundamental Theorem for Linear Systems

Theorem 4. Let A be an $n \times n$ matrix. Then for a given $\mathbf{x}_0 \in \mathbb{R}^n$, the initial value problem

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{F}\,\mathbf{x} \\ \mathbf{x}(0) &= \mathbf{x}_0 \end{aligned} \tag{A.1}$$

has the unique solution given by

$$\mathbf{x}(t) = e^{Ft}\,\mathbf{x}(0).$$

Proof. [1] First, the derivative of $e^{\mathbf{F}\,t}$ with respect to t is shown to be

$$\frac{d}{dt} e^{\mathbf{F}\,t} = \mathbf{F}\,e^{\mathbf{F}\,t}$$

for all $t \in \mathbb{R}^n$. Then

$$\begin{aligned}\frac{d}{dt}[\mathbf{x}(t)] &= \frac{d}{dt}e^{\mathbf{F}\,t} \\ \dot{\mathbf{x}}(t) &= \mathbf{F}[e^{\mathbf{F}\,t}\,\mathbf{x}_0] \\ \dot{\mathbf{x}}(t) &= \mathbf{F}\,\mathbf{x}(t)\end{aligned}$$

To see that this is the only solution, let $\mathbf{x}(t)$ be any solution of the initial value problem, and set

$$\mathbf{y}(t) = e^{-\mathbf{F}\,t}\,\mathbf{x}(t)$$

so that

$$\begin{aligned}\dot{\mathbf{y}}(t) &= e^{-\mathbf{F}\,t}\dot{\mathbf{x}}(t) - \mathbf{F}\,e^{-\mathbf{F}\,t}\,\mathbf{x}(t) \\ &= e^{-\mathbf{F}\,t}\,\mathbf{F}\,\mathbf{x}(t) - \mathbf{F}\,e^{-\mathbf{F}\,t}\,\mathbf{x}(t)\end{aligned}$$

Then because F commutes with $e^{-\mathbf{F}\,t}$, $\dot{y}(t) = 0$. Therefore, $\mathbf{y}(t)$ is constant, and since it satisfies the initial conditions, $\mathbf{y}(t) = \mathbf{x}_0\ \forall\ t$. Thus, any solution of the initial value problem is given by

$$\mathbf{x}(t) = e^{\mathbf{F}\,t}\,\mathbf{y}(t) = e^{\mathbf{F}\,t}\,\mathbf{x}_0$$

□

Appendix B

Central Limit Theorem

Theorem 5. Let $x_1, x_2, \ldots$ be a sequence of independent identically distributed random variables with $\mathbf{E}[x_i] = \mu$ and $\mathrm{Var}(x_i) = \sigma^2 > 0$. Let

$$S_n = \sum_{i=1}^{n} x_i$$

Then

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \to \mathrm{N}(0,1) as n \to \infty$$

Where $\mathrm{N}(0,1)$ is the standard normal distribution.

Proof. [2] Let $Y_i = \frac{X_i - \mu}{\sigma}$, the $\mathbf{E}[Y_i] = 0$ and $\mathbf{E}[Y_i^2] = 1$. Then the characteristic function can be written using Taylor's Theorem

$$\phi_{Y_i} = 1 - \frac{t^2}{2} + o(t^2)$$

Now let

$$\begin{aligned} U_n &= \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \\ &= \sum_{i=1} n \frac{X_i - \mu}{\sqrt{n\sigma^2}} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1} n \left(\frac{X_i - \mu}{\sigma} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \end{aligned}$$

Then the characteristic function is given by

$$\begin{aligned} \phi_{U_n}(t) &= \mathbf{E}\left[e^{itU_n}\right] \\ &= \mathbf{E}\left[e^{\frac{it}{\sqrt{n}}\sum_{i=1}^{n} Y_i}\right] \\ &= \prod_{i=1}^{n} \mathbf{E}\left[e^{\frac{it}{\sqrt{n}}Y_i}\right] \\ &= \prod_{i=1}^{n} \phi_{Y_i}(\frac{t}{\sqrt{n}}) \\ &= \left(\phi_{Y_i}\left(\frac{t}{\sqrt{n}}\right)\right)^n \\ &= \left[1 - \frac{t^2}{2n} + o(t^2)\right]^n \end{aligned}$$

$$\left[1 - \frac{t^2}{2n} + o(t^2)\right]^n \to e^{-\frac{1}{2}t^2} \text{ as } n \to \infty$$

Thus, because $e^{-\frac{1}{2}t^2}$ is the characteristic function for the normal distribution, the sum converges to the normal distribution. □