Parameter Identification and the Design of Experiments for Continuous Non-Linear Dynamical Systems

Adam F. Childers

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

John A. Burns, Chair
Joseph A. Ball
Eugene M. Cliff
Terry L. Herdman

June 15, 2009
Blacksburg, Virginia

Keywords: Parameter Identification, Bounded Error, D-Optimal, Sensitivity Functions

Parameter Identification and the Design of Experiments for Continuous Non-Linear Dynamical Systems

Adam F. Childers

(ABSTRACT)

Mathematical models are useful for simulation, design, analysis, control, and optimization of complex systems. One important step necessary to create an effective model is designing an experiment from which the unknown model parameter can be accurately identified and then verified. The strategy which one approaches this problem is dependent on the amount of data that can be collected and the assumptions made about the behavior of the error in the statistical model. In this presentation we describe how to approach this problem using a combination of statistical and mathematical theory with reliable computation. More specifically, we present a new approach to bounded error parameter validation that approximates the membership set by solving an inverse problem rather than using the standard forward interval analysis methods. For our method we provide theoretical justification, apply this technique to several examples, and describe how it relates to designing experiments. We also address how to define infinite dimensional designs that can be used to create designs of any finite dimension. In general, finding a good design for an experiment requires a careful investigation of all available information and we provide an effective approach to the problem.

This research was supported in part by the Air Force Office of Scientific Research under Grant FA9550-07-1-0273.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis we focus on problems of modeling, parameter identification, and experimental design for problems in science and engineering. Mathematical models are useful for simulation, design, analysis, control, and optimization of complex systems. Creating and effectively using a mathematical model is a multi-step iterative process that consists of understanding the relevant process, developing the mathematical equations that describe this process, using data to estimate parameters in the model and validating the model on real systems.

We focus on systems where the mathematical models are continuous in $t \in \mathbb{R}^+$ and $\theta \in \mathbb{R}^p$ and have the form

$$y = f(t, \theta),$$

where θ is a vector of model parameters and t represents time. The model is designed to describe some measurable aspect of the system in question and changes in the model parameter allow the model to mirror the behavior of the system under a variety of conditions. The actual measurements one takes from a system are data, $d = \{d_j\}_{j=1}^N$, collected at corresponding times $\xi = \{t_j\}_{j=1}^N$. One of our goals is to find a parameter $\hat{\theta}$, so that for each j, $y(t_j, \hat{\theta})$ matches the data d_j "as well as possible". This process is known as parameter identification or parameter estimation. The related process of finding the times $\xi_N = \{t_j\}_{j=1}^N$, for which data collected at those times will lead to a good parameter estimate is a problem in the area of the design of experiments (DOE).

This thesis addresses some fundamental issues involving the modeling process and how these issues impact the design of experiments in the context of parameter identification. In addition, we focus on the DOE problem in the cases that the amount of data is too small for significant statistical analysis or when the sensitivity to the parameters is large. In these cases, poorly designed experiments can lead to at best, accurate parameter estimates that cannot be validated, and at worst, inaccurate parameter estimates that are used to create incorrect models. Experimental design is a crucial component of the modeling process, and reliable models allow researchers to gain insight on how the systems they describe behave as time and inputs to the system change. We approach the problem of identification and validation by a flexible method based on mathematical and statistical theory combined with reliable computation.

Modeling is a multifaceted process so it is important to define exactly what role our method plays in the process. We are working to accurately identify unknown model parameters and validate these estimates. We assume that the model has unknown parameters and these unknown parameters determine how the system evolves. In order to analyze the model, one requires some apriori knowledge about the unknown parameters. This information could come from a previous experiment or from a theoretical framework that provides insight into the dynamics of the system. We will use this information to analyze the statistical and mathematical properties of the model to obtain insight on how to design an experiment to gain the most information about the unknown parameters. The questions we consider are broken down into two parts: i) How many samples are used and ii) When are the best times to take data. The goal is to be able to provide insight into the DOE problem of when and how to collect data in order to both estimate and validate the unknown model parameters. This is especially difficult for small data sets.

The specific goal of this research is to develop methods that can be used to accurately identify and validate unknown model parameters by designing efficient and informative experiments. We are considering the problem of how to design experiments where the controllable variables are the location in time and number of data collection. Ideally we would have a sample at each time contained in the time domain of the experiment, however for almost all real problems this is impossible. Restrictions on the collection of data come in many forms: financial costs, comfort of the test subjects, or limitations from machinery. For example, suppose we were trying to identify parameters in an HIV model and collecting a sample translates into taking blood from a patient. In this example, we have to take into account that it is uncomfortable for the patient to have blood drawn, there is a cost associated with processing the blood into data, and there is a ceiling to the number of times blood can be taken from a patient each day. Therefore, before experimentation begins, it is crucial to know how many samples have to be taken and when to take these sample in order to obtain data sufficient for parameter estimation and validation. We provide answers to these questions by combining the statistical assumptions inherent in the model and the mathematical properties that describe the dynamics of the model.

This thesis is composed of five main parts. Chapter 1 provides an introduction , sets up the notation , and summarizes the basic problem of parameter identification and the DOE. Chapter 2 provides the details of standard parameter estimation methods and discusses standard validation methods as well as new theoretical methods for bounded error parameter identification. Chapter 3 focuses on design of experiments and contains some new theoretical results for designing experiments. Chapter 4 contains a collection of numerical examples to elaborate the theory. Chapter 5 contains a brief conclusion and details the direction our future research is heading.

1.1 Notation

Statistical analysis plays a fundamental role in parameter identification and the DOE. For our work, it is important to be precise about what statistical theory can and cannot imply about a particular DOE. Thus, we review some important mathematical and statistical theories and introduce notation used throughout this thesis. The definitions and theorems may be found in [16].

Definition 1.1.1. *The set, Ω, of all possible outcomes of a particular experiment is called the sample space for the experiment.*

Definition 1.1.2. *A collection of sets $\mathcal{F}$ is called a σ-algebra of a set Ω if*

1. $\Omega \in \mathcal{F}$
2. *If $A \in \mathcal{F}$ then $\Omega/A \in \mathcal{F}$*
3. *If $A_1, A_2, ... \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$*

Definition 1.1.3. *Given a sample space Ω and an associated σ-algebra $\mathcal{F}$, a probability measure is a measure P, with domain $\mathcal{F}$, that satisfies*

1. *$P(A) \geq 0$ for all $A \in \mathcal{F}$*

2. *$P(\Omega) = 1$*

3. *If $A_1, A_2, ... \in \mathcal{F}$ are pairwise disjoint then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$*

Definition 1.1.4. *A probability space is a measure space $(\Omega, \mathcal{F}, P)$, where Ω is sample space, $\mathcal{F}$ is an associated σ-algebra, and P is a probability measure.*

Definition 1.1.5. *Given a measure space $(\Omega, \mathcal{F}, P)$, the function $X(\cdot)$ is $\mathcal{F}$-measurable if for every set A in the range of $X(\cdot)$, the set $\{s : X(s) \in A\}$ is measurable.*

Definition 1.1.6. *Given a probability space $(\Omega, \mathcal{F}, P)$, a real random variable $X(\cdot)$ is a $\mathcal{F}$-measurable function from Ω into $\mathbb{R}^n$.*

Definition 1.1.7. *For a particular outcome of an experiment $s \in \Omega$, $x = X(s)$ is called the realization of the random variable.*

Definition 1.1.8. *The induced probability measure $P_X(A)$ is given by*

$$P_X(X(\cdot)\epsilon A) = P(\{s\epsilon\Omega : X(s)\epsilon A\})$$

The induced probability measure is defined on sets in the range of the random variable, where as the probability measure is defined on sets in the sample space. We see that sometimes they are the same and sometimes they are not. To give a clear illustration of what these definitions mean, we now consider describing the behavior of two systems in the context of the definitions we have stated.

Example 1.1.1. *In this example we will describe flipping a fair coin using Definitions 1.1.1-1.1.8. The experiment is tossing a fair coin and observing the outcome. Either it lands heads up, in which case we observe H or it lands tails up and observe T. The sample space is $\Omega = \{H, T\}$. The associated σ-algebra is $\mathcal{F} = \{\emptyset, H, T, \{H, T\}\}$. The probability measure is defined by $P_{flip}(H) = P_{flip}(T) = .5$. With the sample space, a σ-algebra, and the probability measure we define the probability space $(\Omega, \mathcal{F}, P) = (\{T, H\}, \{\emptyset, H, T, \{H, T\}\}, P_{flip})$. A random variable defined on $(\Omega, \mathcal{F}, P)$ is $X(H) = 1$ and $X(T) = 0$. For this random variable the induced probability measure is defined by $P_X(1) = P_X(0) = .5$.*

Example 1.1.2. *In this example we will describe picking a random number out of the interval $[0, 1]$ using Definitions 1.1.1-1.1.8. The experiment is picking a random number out of the* *interval* $[0,1]$ *and observing the outcome. We make the assumption that each outcome is equally likely. The sample space is* $\Omega = [0,1]$*. The associated* σ*-algebra is* $\mathcal{F} = \beta$*, the Borel set on* $[0,1]$*. The probability measure is the Lebesgue measure* (ℓ)*. With the sample space, a* σ*-algebra, and the probability measure we define the probability space* $(\Omega, \mathcal{F}, P) = ([0,1], \beta, \ell)$*. A random variable defined on* $(\Omega, \mathcal{F}, P)$ *is* $X(s) = s$*. For this random variable the induced probability measure is defined by* $P_X(A) = \ell(A)$*. Notice that in the example the probability measure and the induced probability measure are the same.*

Definition 1.1.9. *The cumulative distribution function (cdf) of a continuous vector valued random variable* $X(\cdot) = [X_1(\cdot), \ldots, X_n(\cdot)]^T$ *denoted by* $F_X(x)$*, is defined by*

$$F_X(x) = P_X(X_1(\cdot) \leq x_1, \ldots, X_n(\cdot) \leq x_n) \text{ for all } x.$$

Definition 1.1.10. *The probability density function (pdf),* $f_X(\cdot)$*, for a vector valued random variable* $X(\cdot) = [X_1(\cdot), \ldots, X_n(\cdot)]^T$ *is the function that satisfies*

$$F_X(x) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_X(t_1, t_2, \ldots, t_n) dt_1 dt_2 \ldots dt_n.$$

The cdf and the pdf describe the behavior and give information about the probability of different realizations of the random variable.

Definition 1.1.11. *The expected value or mean of a real random variable* $g(X(\cdot))$*, denoted by* $Eg(X(\cdot))$ *is*

$$\begin{aligned} Eg(X(\cdot)) &= \int_\Omega g(x) f_X(x) dx \\ &= \int_\Omega g(x) dP_X(x) \end{aligned}$$

provided that the integral exists.

Definition 1.1.12. *The variance of a real scalar random variable* $X(\cdot)$ *is*

$$VarX(\cdot) = E(X(\cdot) - EX(\cdot))^2.$$

The mean can be thought of as the average outcome to the experiment and the variance describes how close to the mean outcomes are likely to be. A small variance would say that the outcome of most experiments will be close to the mean. As the variance increases so does the likelihood that the outcomes will be farther from the mean.

Definition 1.1.13. *The covariance of two random variables* $X(\cdot)$ *and* $Y(\cdot)$ *is*

$$Cov(X(\cdot), Y(\cdot)) = E[(X(\cdot) - EX(\cdot))(Y(\cdot) - EY(\cdot))^T].$$

Definition 1.1.14. *The variance matrix of a vector random variable* $X(\cdot)$ *is*

$$VarX(\cdot) = E[(X(\cdot) - EX(\cdot))(X(\cdot) - EX(\cdot))^T].$$

Like the variance of a scalar random variable, the variance matrix also describes what the outcomes of an experiment are likely to be with respect to the mean. The reason the variance is a matrix and not a vector is that the different components of the vector valued random variable could be related.

Theorem 1.1.1. *The variance matrix is positive semi-definite.*

For a scalar valued random variable, the variance is always non-negative. Theorem 1.1.1 is the equivalent for the vector valued case.

A fundamental assumption in the statistical theory we will uses for parameter identification is that the different components of the statistical model are identically independently distributed (i.i.d.). The next definitions set up what it means for a set of random variables to be i.i.d.

Definition 1.1.15. *If* $X_1(\cdot), X_2(\cdot), \ldots, X_n(\cdot)$ *are real valued random variables with corresponding probability density functions* $f_1(\cdot), f_2(\cdot), \ldots, f_n(\cdot)$ *and* $f(\cdot)$ *is the probability density function for the vector valued random variable* $[X_1(\cdot), X_2(\cdot), \ldots, X_n(\cdot)]^T$*, then the random variables* $X_1(\cdot), X_2(\cdot), \ldots, X_n(\cdot)$ *are independent iff*

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot, \ldots, \cdot f_n(x_n)$$

for all $[x_1, x_2, \ldots, x_n]^T$ *in the range of* $[X_1(\cdot), X_2(\cdot), \ldots, X_n(\cdot)]^T$*.*

Definition 1.1.16. *Given a probability space* $(\Omega, \mathcal{F}, P)$*, two random variables* $X_1(\cdot)$ *and* $X_2(\cdot)$ *are said to be identically distributed on* $(\Omega, \mathcal{F}, P)$ *if*

$$P_{X_1}(X_1(\cdot) \in A) = P_{X_2}(X_2(\cdot) \in A)$$

for all $A \in \mathbb{R}^n$

Then a set random variables are i.i.d. if they are all independent of one another and are all identically distributed. For a quick and intuitive example of i.i.d. random variables we will return to the example of flipping a fair coin. Let the experiment be observing the outcome of flipping two fair coins. Let $X_1(\cdot)$ and $X_2(\cdot)$ both be defined by Example 1.1.1. Consider the vector valued random variable defined by $Y(\cdot) = [X_1(\cdot), X_2(\cdot)]^T$. The sample space is $\{(H, H), (T, T), (H, T), (T, H)\}$. Without going through the proof rigorously we can intuitively say that X_1 and X_2 are independent since the outcome of flipping one coin has no affect on the outcome of flipping the other coin. Additionally they are identically distributed since by definition they both have the same distribution.

Next we will define some common distributions. First, a random variable has a uniform distribution if each outcome is equally likely to occur.

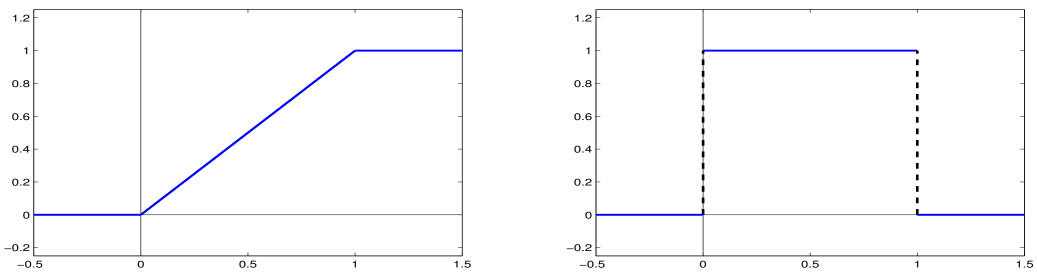
Definition 1.1.17. *A real vector valued random variable $X(\cdot)$ is said to have a uniform distribution on $I \subset \mathbb{R}^n$ if*

$$f_X(x) = \begin{cases} \frac{1}{\ell(I)} \text{ if } x \epsilon I \\ 0 \text{ else} \end{cases}$$

where ℓ is the Lesbesgue measure.

We abbreviate the statement "$X(\cdot)$ has a distribution given by $F_X(x)$ (or $f_X(x)$)" by "$X(\cdot) \sim F_X(x)$". The symbol "$\sim$" is defined to be "is distributed as". Then the statement a random variable $X(\cdot)$ has a uniform distribution on the interval I can be expressed as $X(\cdot) \sim U(I)$ where $U(I)$ describes a distribution defined by Definition 1.1.17. The random variable defined in Example 1.1.2 has a uniform distribution on the interval $[0, 1]$ or $X(\cdot) \sim U([0, 1])$. The cdf and pdf for this random variable can be seen in Figure 1.1.

The normal distribution is one of the most fundamental distributions in probability theory and statistical theory. In the scalar case a random variable with a normal distribution is often said to have a bell curved distribution because the pdf has the shape of a bell. The normal distribution is often used because one can obtain precise analytical results and second because it can, in many cases be, used to approximate how real systems behave. The first time the normal distribution will come up in this thesis is in the statistical model. We make the assumption that error associated with collected data has a normal distribution. Making this assumption requires an investigation of the system in question and is not always, but in many cases appropriate. It will also appear in our discussion of asymptotic theory in the



(a) cdf (b) pdf

Figure 1.1: $X(\cdot) \sim U([0,1])$

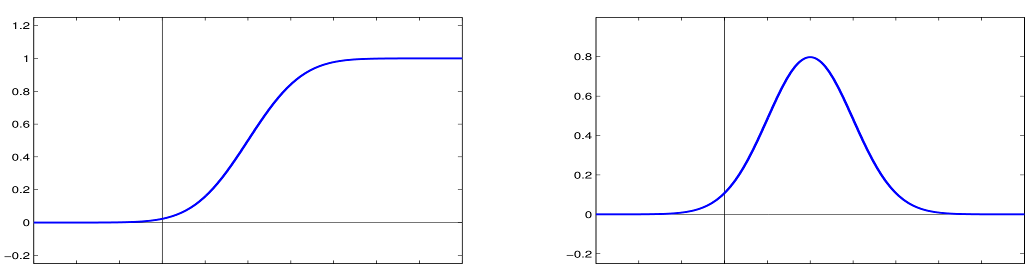
central limit theorem.

Definition 1.1.18. *A real scalar valued random variable $X(\cdot)$ is said to have normal distribution on $\mathbb{R}$ or $X(\cdot) \sim N(\mu, \sigma^2)$ if*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean and σ^2 is the variance.

Figure 1.2 shows plots of the random variable with a normal distribution with mean 1 and variance 0.5.



(a) cdf (b) pdf

Figure 1.2: $X(\cdot) \sim N(1, 0.5)$

Definition 1.1.19. *A real vector valued random variable $X(\cdot)$ is said to have normal distribution on $\mathbb{R}^n$ or $X(\cdot) \sim N(\mu, V)$ if*

$$f_X(x) = \frac{1}{2\pi^{n/2}|V|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T V(x-\mu)}$$

where μ is the mean and V is the variance matrix.

Note : Throughout this presentation, random variables will be capital letters and their realizations will be lower case letters. We will consider three main random variables. The statistical model will be the random variable $D_j(\cdot)$ that has realization d_j. The error that arises from collecting data will be denoted by e_j and e_j is the realization of the random variable $\mathcal{E}_j(\cdot)$. Finally, a parameter estimator is the random variable $\hat{\Theta}(\cdot)$ with realization $\hat{\theta}$. $\hat{\Theta}(\cdot)$ is a random variable since it is a function of data collected from a system and we consider data to be random variables. It is important to remember that a random variable is a **function** defined on a sample space. Where there is no chance for confusion we will sometimes write $D_j(\cdot)$ as simply D_j. There will be two operators, L and E, the expected value operator. We will be working in the space $\mathcal{X} \subset \mathbb{R}^n$ which is the range of the model, $\mathcal{Y} \subset \mathbb{R}^m$ which is the range of the model output, and $\Phi \subset \mathbb{R}^p$ which is the parameter space.

Although we do not present a complete overview of the statistical background, it is sufficient to begin with the outline of our problem.

1.2 The Mathematical Model

Let $f(\cdot,\cdot) : [0,T] \times \Phi \to \mathbb{R}^m$ be a continuously differentiable function in both arguments. We assume that t is contained in the closed interval $[0,T] \subset \mathbb{R}^1$ and the parameter $\theta \in \Phi \subset \mathbb{R}^p$. The general continuous model has the form

$$y = f(t, \theta),$$

for all $t \in [0,T]$ and $\theta \in \Phi \subset \mathbb{R}^p$. If the function $f(\cdot,\cdot)$ comes from an expression where it can be solved for exactly, the model is called exact. Let $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{Y} \subset \mathbb{R}^m$. An exact model has the form

$$x(t,\theta) = h(t,\theta),$$
$$y(t,\theta) = w(x(t,\theta)),$$

where $h : [0, T] \times \Phi \to \mathcal{X}$ is a function and $w : \mathcal{X} \to \mathcal{Y}$ is a function.

In our study, the mathematical model is described by a differential equation. We will also consider models of the form

$$\dot{x}(t, \theta) = g(t, \theta, x(t, \theta)) \qquad x(0, \theta) = x_0, \tag{1.2.1}$$

$$y(t, \theta) = w(x(t, \theta)), \tag{1.2.2}$$

where $g : [0, T] \times \Phi \times \mathcal{X} \to \mathcal{X}$ is a function and $w : \mathcal{X} \to \mathcal{Y}$ is a function. Since there is error associated with collecting data, it is very unlikely that any model will exactly fit experimental data. To compensate for this, we introduce a statistical model that accounts for discrepancies between the mathematical model and experimental data.

1.2.1 The Statistical Model

The statistical model is defined by

$$D_j = y(t_j, \theta_0) + \mathcal{E}_j,$$

where $\mathcal{E}_j$ is a random variable that represents noise in the data collection process and θ_0 is assumed to be "the true" model parameter. Since $\mathcal{E}_j$ is a random variable, D_j is also a random variable. We make the assumptions that for $j = 1, \ldots, N$, the random variables $\mathcal{E}_j$ are i.i.d. with $E[\mathcal{E}_j] = 0$. For the scalar case we assume $var[\mathcal{E}_j] = \sigma^2$ and for the vector case we assume $var[\mathcal{E}_j] = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$. From these assumptions it follows that

$$E[D_j] = y(t_j, \theta_0).$$

As mentioned previously, we are interested in determining where in time to sample in order to gain as much information about the unknown parameter as possible. Because of this, we will be very careful to define sets of sampling times. A set of sampling times will be called a **time sample** or an **experiment** and is defined to be

$$\xi_N = \{t_j\}_{j=1}^N.$$

Notice that an experiment ξ_N is defined by the number of samples N and the time to take the N samples. Given a time sample, we can define a random process as

$$\begin{aligned} D(\xi_N) &= [D_1, \ldots, D_N]^T \\ &= [y(t_1, \theta_0) + \mathcal{E}_1, \ldots, y(t_N, \theta_0) + \mathcal{E}_N]^T. \end{aligned}$$

The assumptions we make about the statistical model translate to the assumption that the mathematical model is "good". In particular, for some $\theta_0 \in \Phi$, $y(t, \theta_0)$ exactly describes the system for all $t \in \xi_N$. For an experiment ξ_N, we consider a set of experimental data to be a realization of the statistical process $D(\xi_N)$ and will be denoted by

$$\begin{aligned} d(\xi_N) &= [d_1, \ldots, d_N]^T \\ &= [y(t_1, \theta_0) + \varepsilon_1, \ldots, y(t_N, \theta_0) + \varepsilon_N]^T, \end{aligned}$$

where ε_j is a realization of the random variable $\mathcal{E}_j$.

Given a mathematical model and a set of experimental data, the next step is to find a parameter that "fits" the mathematical model to the experimental data. This process is known as the parameter identification problem and what we assume about the distribution of the random variable that describes the error, determines how we solve the parameter identification problem. However, before discussing parameter identification, we present a short summery on sensitivity equations that are defined by the model. As we will see later, sensitivities play a role in parameter identification, parameter validation and in the design of experiments.

1.2.2 Sensitivity Equations

Roughly speaking, sensitivities provide information on how changes in the model parameters impact the model output. Accurately solving for the sensitivities in a model is one of the most important parts in both parameter estimation and parameter validation. In the parameter estimation process we use gradient based algorithms that require numerous calculations of the sensitivities and in the parameter validation process the sensitivities are necessary to describe the distribution of the estimated parameter. All of the methodology for designing experiments require accurate sensitivity computations.

Definition 1.2.1. *The sensitivities with respect to the model parameters are defined to be*

$$s_{\theta_1}(t,\theta) \equiv [\frac{\partial y_1(t,\theta)}{\partial \theta_1}\ \frac{\partial y_2(t,\theta)}{\partial \theta_1}\ \dots\ \frac{\partial y_m(t,\theta)}{\partial \theta_1}]^T \tag{1.2.3}$$

$$s_{\theta_2}(t,\theta) \equiv [\frac{\partial y_1(t,\theta)}{\partial \theta_2}\ \frac{\partial y_2(t,\theta)}{\partial \theta_2}\ \dots\ \frac{\partial y_m(t,\theta)}{\partial \theta_2}]^T \tag{1.2.4}$$

$$\vdots \tag{1.2.5}$$

$$s_{\theta_p}(t,\theta) \equiv [\frac{\partial y_1(t,\theta)}{\partial \theta_p}\ \frac{\partial y_2(t,\theta)}{\partial \theta_p}\ \dots\ \frac{\partial y_m(t,\theta)}{\partial \theta_p}]^T \tag{1.2.6}$$

where $\theta = [\theta_1, \theta_2, \dots, \theta_p]^T$.

For exact models, solving for the equations is not a problem, we simply differentiate the mathematical model with respect to the unknown parameter $\theta = [\theta_1, \theta_2, \dots, \theta_p]^T$.

For systems described by exact models, computing the sensitivities is straight forward but much more care has to be taken when solving for the sensitivity of a model described by a differential equation. We shall employ a continuous sensitivity method ([11], [14]). In order to derive the sensitivity equation, one assumes that integrating $\dot{x}(t,\theta)$ with respect to t and then differentiating with resect to θ is the same as differentiating $\dot{x}(t,\theta)$ with respect to θ and then integrating with respect to t. Of course, this assumption places requirements on the function $g(t,\theta)$. Formally differentiating

$$\dot{x}(t,\theta) = g(t,\theta,x(t,\theta)) \tag{1.2.7}$$

with respect to θ yields the state sensitivities

$$\begin{aligned}
\dot{x}_{\theta_1}(t,\theta) &= g_x(t,\theta,x(t,\theta))x_{\theta_1}(t,\theta) + g_{\theta_1}(t,\theta,x(t,\theta)) \\
\dot{x}_{\theta_2}(t,\theta) &= g_x(t,\theta,x(t,\theta))x_{\theta_2}(t,\theta) + g_{\theta_2}(t,\theta,x(t,\theta)) \\
&\vdots \\
\dot{x}_{\theta_p}(t,\theta) &= g_x(t,\theta,x(t,\theta))x_{\theta_p}(t,\theta) + g_{\theta_p}(t,\theta,x(t,\theta)).
\end{aligned}$$

If one similarly differentiates the initial condition $x(0,\theta)$ respect to θ and define

$$x_\theta(0,\theta) = [\frac{\partial x(0,\theta)}{\partial \theta_1}^T, \frac{\partial x(0,\theta)}{\partial \theta_2}^T, \ldots, \frac{\partial x(t,\theta)}{\partial \theta_p}^T]^T,$$

then it follows that the state sensitivities can be computed by solving the system

$$\dot{x}(t,\theta) = g(t,\theta,x(t,\theta)) \quad (1.2.8)$$

$$\dot{x}_{\theta_1}(t,\theta) = g_x(t,\theta,x(t,\theta))x_{\theta_1}(t,\theta) + g_{\theta_1}(t,\theta,x(t,\theta)) \quad (1.2.9)$$

$$\dot{x}_{\theta_2}(t,\theta) = g_x(t,\theta,x(t,\theta))x_{\theta_2}(t,\theta) + g_{\theta_2}(t,\theta,x(t,\theta)) \quad (1.2.10)$$

$$\vdots$$

$$\dot{x}_{\theta_p}(t,\theta) = g_x(t,\theta,x(t,\theta))x_{\theta_p}(t,\theta) + g_{\theta_p}(t,\theta,x(t,\theta)) \quad (1.2.11)$$

$$[x(0,\theta)^T, x_\theta(0,\theta)^T]^T = [x_0^T, x_{\theta,0}^T]^T. \quad (1.2.12)$$

Finally, we obtain the sensitivities of the model by

$$s_{\theta_1}(t,\theta) = y_{\theta_1}(t,\theta) = [w_x(x(t,\theta))]x_{\theta_1}(t,\theta) \quad (1.2.13)$$

$$s_{\theta_2}(t,\theta) = y_{\theta_2}(t,\theta) = [w_x(x(t,\theta))]x_{\theta_2}(t,\theta) \quad (1.2.14)$$

$$\vdots$$

$$s_{\theta_p}(t,\theta) = y_{\theta_p}(t,\theta) = [w_x(x(t,\theta))]x_{\theta_p}(t,\theta). \quad (1.2.15)$$

Solving the sensitivity equations can be a hefty numerical problem when the number of model parameters is large so we introduce a method to solve them in parallel.

1.2.3 The Parallel Method for Computing the State Sensitivities

Solving the sensitivity equations for models that have a large number of parameters is computationally expensive. For example, finding the sensitivities for a model with n states, m outputs and p parameters, requires solving a $n*(p+1)$ dimensional differential equation. The issue is compounded for applications like optimization that require solving the sensitivity equations multiple times. We can address this issue by exploiting the fact that the sensitivity equations are "embarrassingly" parallel. To see how the problem can be solved in parallel we write the j^{th} parallel sensitivity equation as

$$\dot{x}(t,\theta) = g(t,\theta,x(t,\theta)) \tag{1.2.16}$$

$$\dot{x}_{\theta_j}(t,\theta) = g_x(t,\theta,x(t,\theta))x_{\theta_j}(t,\theta) + g_{\theta_j}(t,\theta,x(t,\theta)). \tag{1.2.17}$$

Instead of solving the $n*(p+1)$ dimensional differential equation described in Equations 1.2.8 through 1.2.12 on a single processor, we can the split it into p, $2*n$ dimensional differential equations all to be solved on separate processors. Note that solving Equation 1.2.8 p times is more computationally efficient than solving it one time and passing the information between processors. By not passing the information, we solve more equations, however, each of the j^{th} sensitivity equations are fully independent.

1.2.4 Approximating the Model

If the mathematical model can not be solved exactly, as is often the case when the model is described by a differential equation, the mathematical model and the sensitivity functions must be approximated. We say a partition of the set $A \subset \mathbb{R}^n$ is μ-fine, if for each $a \in A$, there exists a $p \in \rho^\mu$ such that $\|a-p\|_2 < \mu/2$. Suppose that $\rho^\mu = [p_1, \ldots, p_M]$ is a μ-fine partition of $[0,T]$. We will solve the sensitivity equations, Equation (1.2.8) through Equation (1.2.12) on the partition ρ^μ using a variety of differential equation solvers. Further suppose that $[w_1^{\rho^\mu}(p_1,\theta), \ldots, w_M^{\rho^\mu}(p_M,\theta)]$ is the approximation of $[y(p_1,\theta),\ldots,y(p_M,\theta)]$. Define $w^{\rho^\mu} : [0,T], \Phi \to \mathbb{R}^m$ to be the linear interpolation of the points $\{(p_j, w_j^{\rho^\mu}(p_j,\theta))\}$. Now suppose that $[s_1^{\rho^\mu}(p_1,\theta), \ldots, s_M^{\rho^\mu}(p_M,\theta)]$ is the approximation of $[y_\theta(p_1,\theta),\ldots,y_\theta(p_M,\theta)]$ and $s^{\rho^\mu} : [0,T], \Phi \to \mathbb{R}^{m*p}$ to be the linear interpolation of the points $\{(p_j, s_j^{\rho^\mu}(p_j,\theta))\}$.

Now starting with an actual continuous mathematical model we will define the statistical model, derive the sensitivity equations, and approximate the model and sensitivities.

Example 1.2.1. *Consider the mathematical model*

$$\dot{x}(t,[a,b]^T) = ax(t,[a,b]^T) - bx(t,[a,b]^T)^2 \qquad x(0,[a,b]^T) = x_0, \tag{1.2.18}$$

$$y(t,[a,b]^T) = x(t,[a,b]^T), \tag{1.2.19}$$

where $[a,b]^T \in \mathbb{R}_+^2$ *and* $t \in [0,20]$. *Notice that in this example the model parameter is* $\theta = [a,b]^T$, *the parameter space is* $\Phi = \mathbb{R}_+^2$ *and the function* $w(\cdot)$ *is the identity operator.*

The solution to Equations (1.2.18) and (1.2.19) is know as the logistic equation. Figure 1.3 shows a graphical representation of the model.

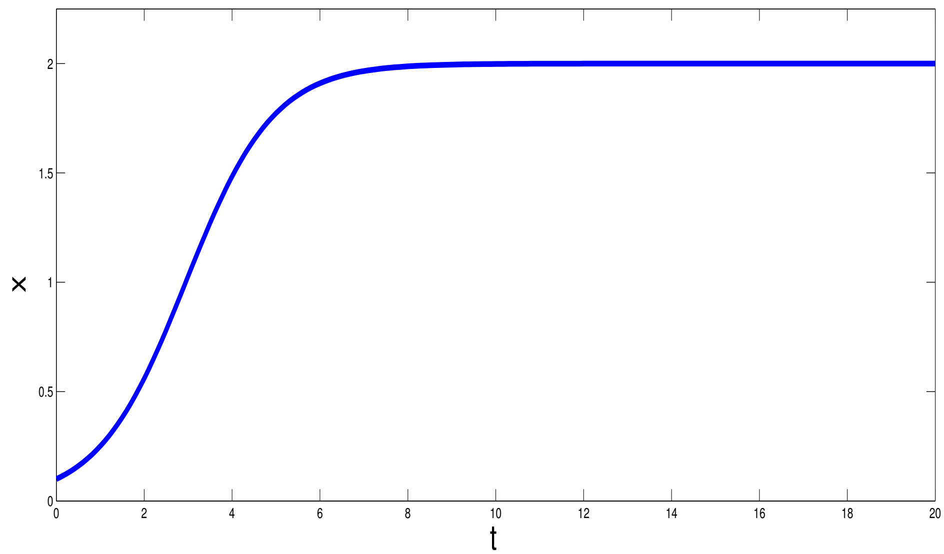


Figure 1.3: The logistic equation with $[a, b]^T = [1, 0.5]^T$ and $x_0 = 0.1$

Suppose that the "true" model parameter is $\theta_0 = [1, 0.5]^T$. *The statistical model is given by*

$$D_j = y(t_j, [1, 0.5]^T) + \mathcal{E}_j. \tag{1.2.20}$$

Let $\xi_3 = \{2, 5, 15\}$ *be an experiment and* $D(\xi_3)$ *be the corresponding random process. Figure 1.4 show a graphical representation of a realization of the random process* $D(\xi_3)$ *in relation to the continuous model* $y(t, [1, 0.5]^T)$.

We next derive the sensitivity equations by differentiating the mathematical model (1.2.18) and (1.2.19) with respect to the parameter $[a, b]^T$. *The sensitivity equations for the logistic equation are given by*

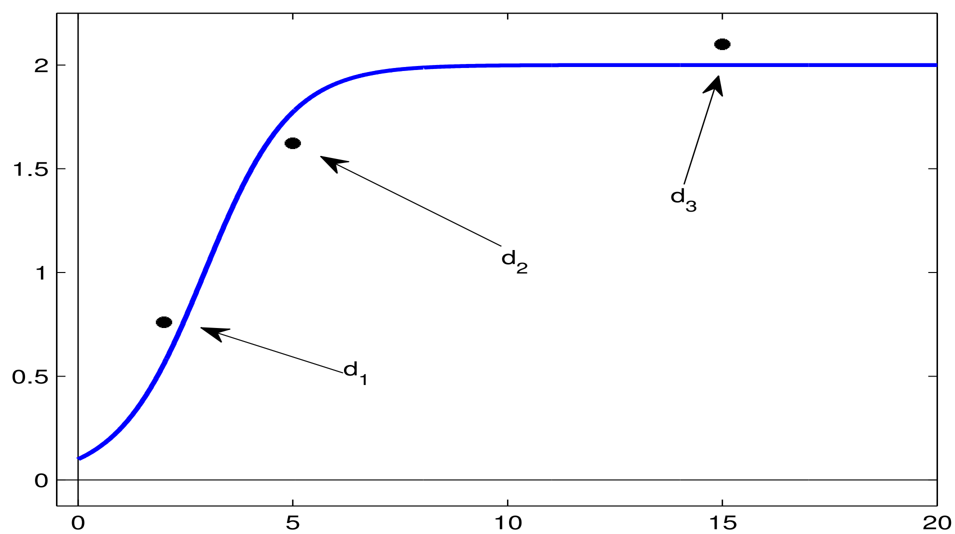


Figure 1.4: A realization of the random process $D(\xi_3)$

$$\dot{x} = ax - bx^2 \tag{1.2.21}$$

$$\dot{x}_a = (ax_a + x) - 2bxx_a \tag{1.2.22}$$

$$\dot{x}_b = ax_b - (x^2 + 2bxx_b) \tag{1.2.23}$$

with initial conditions $[x(0,[a,b]^T), x_a(0,[a,b]^T), x_b(0,[a,b]^T)]^T = [x_0, 0, 0]^T$. *For fixed* $[a, b^T] = [1, 0.5]^T$ *we can approximate the model by solving the sensitivity equations using Euler's explicit method on the uniform partition* $\rho^{0.25} = [0, 0.25, 0.50, \ldots, 19.75, 20]$. *The result is an approximate model* $w^{\rho^{0.25}}(t, [1, 0.5]^T)$ *and the approximate sensitivities* $s^{\rho^{0.25}}(t, [1, 0.5]^T)$. *Figure 1.5 shows the true mathematical model* $y(t, \theta_0)$ *and the approximate model* $w^{\rho^{0.25}}(t, \theta_0)$. *For this example our approximation is not very good but we could easily improve the approximation by decreasing the mesh of the partition or using an appropriate differential equation solver. Regardless, the point of Figure 1.5 is to show that when we are working with models described by differential equitation we often times have to approximate them.*

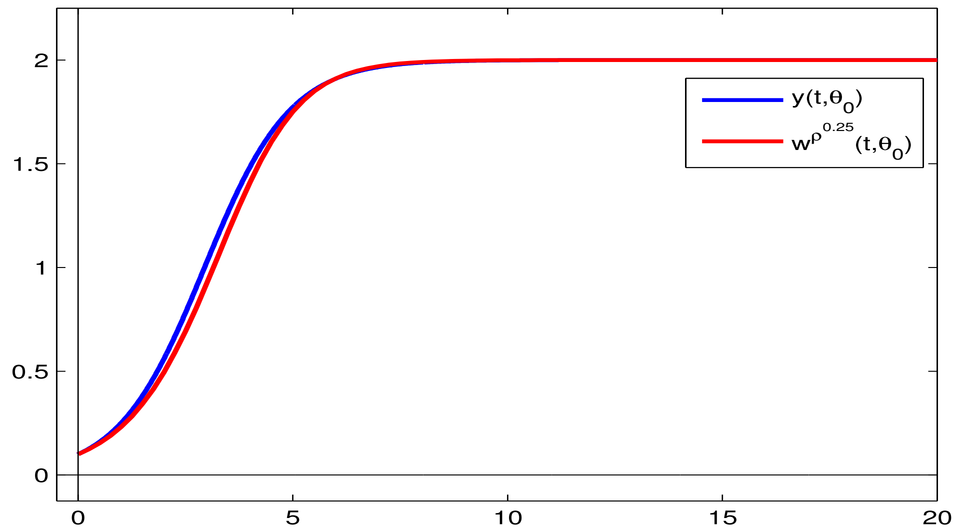
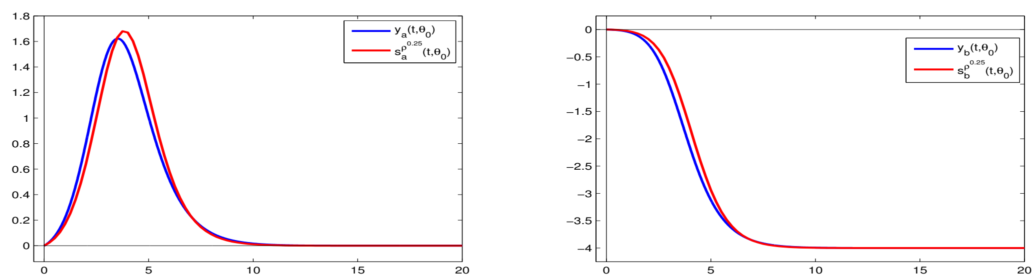


Figure 1.5: The true model $y(t, \theta_0)$ and the approximate model $w^{\rho^1}(t, \theta_0)$



(a) (b)

Figure 1.6: The approximate sensitivities and the true sensitivities

1.3 An Overview of Parameter Estimation and Parameter Validation

The strategy we use to estimate and validate unknown model parameters is based on the assumptions we make about the $\mathcal{E}_j$'s, the random variable that represent error in the statistical model, and the amount of experimental data available. We shall use two parameter estimation techniques: i) least squares parameter estimation and ii) maximum likelihood parameter estimation. In both cases we make standard statistical assumptions about the behavior of the $\mathcal{E}_j$'s, but the actual decision about which technique is superior, is determined by how much is known about their distributions ([8], [16], [39]). If we know approximately the distribution of the $\mathcal{E}_j$'s, then we will use a maximum likelihood estimator in order to use the information we know about the distribution of the $\mathcal{E}_j$'s in the parameter estimate. Likewise, we will use a least squares estimator if it is unreasonable to assume that the distribution of the $\mathcal{E}_j$'s is known, since they do not require this aproiri information [14]. The amount of information known about the distribution of the random variable that represents error is also the key to determining how to validate the parameter estimate.

There is a rich statistical theory for validating maximum likelihood and least squares parameter estimates ([16], [24], [39]). Both techniques use the behavior of the distribution of the parameter estimator as the number of data samples grow large to give insight on the quality of the parameter estimate. While one cannot compute the true distribution of the parameter estimator for a practical problem, one can estimate it using experimental data. By finding an approximation of the distribution of the parameter estimator, one can construct confidence ellipsoids about the parameter estimate. Intuitively, one can conclude that a smaller confidence ellipsoid will correspond to a better estimate. It is import to note here that the parameter estimators are random variables, so experiments that have different time samples, represent different random variables which in turn each have their own distribution. For parameter identification problems where a relatively large amount of data can be collected, these methods of parameter validation are effective. However, for problems where there is a limited amount of data, these methods are either less powerful or not applicable. They are less powerful in the sense that one uses the data to approximate the distribution of the parameter estimator, so less data corresponds to an estimate that might not be able to be sufficiently validated. They are not applicable in the sense that without a minimum amount of data the approximation is essentially meaningless. The approximations rely on the central limit theorem [39] and small data sets might not correspond to a good approximation of the normal distribution.

Another approach is the so called **bounded error method**. Making precise and rigorous statistically conclusive statements about parameter estimates where there is a very small amount of data available is impossible. However, using bounded error parameter validation one can make mathematically precise and rigorous conclusive statements about the parameter estimate ([33], [34]). In any case, if the data set is too small and very noisy then little can be said about the quality of a parameter estimate. On the other hand, if there is a small amount of high quality data then one can begin to describe the quality of the estimate mathematically. To use bounded error parameter validation, the error must be assumed bounded in the statistical model and the bounds are known. This is a strong assumption but it can be realistic for some problems. The method exploits the dynamics of the model to describe the quality of the parameter estimate. Bounded error parameter validation works by finding a membership set that describes the possible parameters that the mathematical model maps to the bounded error data. The advantage of this method is that there are no restrictions about the number of data samples necessary to use the theory, only that the error is bounded. Again, we will intuitively associate a relatively small membership set to a good parameter estimate. This method is inherently more costly than the statistical parameter validation methods, but fills in an important gap that the statistical theory misses.

1.4 An Overview of the Design of Experiments

The next logical step in the parameter identification process is to investigate how to design an experiment that is effective for estimating the unknown model parameter and validating that estimate. We first consider the experimental design problem where time is the only controllable variable. The question we are trying to answer is, where should the experimenter sample in order to get the best parameter estimate? This question can be more precisely defined by asking, if there are M data samples, where should the next N be taken to get a good parameter estimate that can be verified? Note that M can be 0. The question we pose here is very important because two sets of data that each contain N samples but data collected at different times can lead to parameter estimates of very different quality [6]. As was the case in parameter identification, the assumptions one makes about the distribution of the $\mathcal{E}_j$'s determine how to proceed in the design of the experiment.

To design an experiment there has to be some apriori assumptions made, or assumed known, about the parameter being estimated. This information can come from previous experiments or from an investigation of the system being modeled. Regardless of where the information comes from, the design of experiment begins with an initial parameter $\tilde{\theta}$, that is an an estimate of the true parameter θ_0. The experiment is designed based on the estimated parameter $\tilde{\theta}$, so the quality of that estimate strongly influences the quality of the design.

Clearly, designing an experiment for a poor estimate of θ_0 could lead to an experiment that is not meaningful. For this reason, it is crucial to take the quality of the parameter estimate that is used to design the experiment into account when describing an experimental strategy [9]. In some cases a lack of information about the parameter that the experiment is being designed for can be made up by designing an experiment that is more robust ([2], [28]).

The quality of the design of an experiment is measured by the amount of confidence that one has in a parameter estimated from the experiment. There are many ways to measure this quality; for example, the measure of a confidence ellipsoid centered at the parameter estimate or the size of the standard errors for components of the unknown parameter ([3], [6], [19]). This problem can be approached directly by defining ways to measure different optimality conditions or indirectly by studying the behavior of the model. Common approaches for a direct approach are a D-optimal design or a V-optimal design, while a more indirect approach would use traditional sensitivity functions and generalized sensitivity functions. We propose to design experiments that minimize the measure of confidence intervals centered at the parameter estimate, minimizing standard errors of the parameter estimate, and minimizing the measure of membership sets for bounded error problems. The word "optimal" means different things depending on what assumptions one makes about the $\mathcal{E}_j$'s, the amount of data one has, and the goals of the experiment.

We first consider the DOE problem using traditional sensitivity functions and generalized sensitivity functions. We then consider D-optimal designs and V-optimal designs. For all of the problems the quality of the initial parameter estimate must be considered when we design the experiment.

Chapter 2

Parameter Identification

Parameter identification is a two part problem: estimating the unknown parameter and validating that estimate. Parameter estimation deals with fitting the model to experimental data, and parameter validation concerns describing the amount of confidence that the experimenter has in the estimate. These topics go hand in hand, and in this chapter we will detail techniques for estimating unknown parameters and then discuss how to validate different types of estimates. Selecting which parameter estimation technique to use and how to validate the estimate comes down to what is assumed about the random variables that represents error in the statistical model and the amount of experimental data available.

2.1 Parameter Estimation Techniques

Roughly, parameter estimation is finding a parameter estimate that fits the mathematical model to a given set of experimental data. In the following section we will investigate least squares parameter estimation and maximum likelihood parameter estimation. The main difference between the two methods is that maximum likelihood estimation requires that the experimenter knows apriori, the distribution of the random variable that represents error in the statistical model, and least squares parameter estimation does not. We will begin by describing ordinary least squares parameter estimation for the scalar case and then the vector case. Then we will discuss maximum likelihood estimators. MLE's are the most powerful of the estimation techniques, but they also require the strongest assumptions in the statistical model. Determining which estimation technique to use is dependent on what information the experimenter has or assumes that they know about the distribution of the random variables that represent error in the statistical model.

2.1.1 Least Squares

Least squares parameter estimation estimates the unknown parameter by finding the minimizer of the possibly weighted sum of the square of the Euclidean distance from the data to the model. Least squares is very intuitive in terms of how the search for the parameter estimate is found and is applicable to almost all problems because finding the least squares parameter estimate does not require that the distributions of the $\mathcal{E}_j$'s in the statistical model are known.

Ordinary Least Squares

Ordinary least squares (OLS) is a parameter estimation technique that selects the parameter that minimizes the square of the Euclidean distance between the model and the experimental data. We will first consider using OLS to identify the unknown parameters for a scalar model and follow the methodology detailed in ([16],[39]). A nice summary of the theory can be found in [8]. To apply the theory of OLS we must assume that the $\mathcal{E}_j$'s are identically independently distributed with $E[\mathcal{E}_j] = 0$, and that they have constant variance, $var[\mathcal{E}_j] = \sigma_0^2$ for $j = 1, \ldots, N$. However, it is not necessary to know what the distributions of the $\mathcal{E}_j$'s are. Given a realization of the random process $D(\xi_N) = [D_1, ..., D_N]^T$, we want to estimate the parameter θ_0, in the statistical model.

Definition 2.1.1. *For a mathematical model $y(t, \theta)$ and a realization of the random process $D(\xi_N)$, the OLS parameter estimate is defined by*

$$\hat{\theta}_{OLS}(\xi_N) = arg \min_{\theta \in \Phi} \sum_{j=1}^{N} [d_j - y(t_j, \theta)]^2. \tag{2.1.1}$$

It is important to notice that our estimate of the true model parameter is a function of the time sample and we will explicitly write the dependence. In general, we are interested in the behavior of $\hat{\theta}_{OLS}(\xi_N)$ for all of the possible realizations of $D(\xi_N)$. In order to study this behavior we define a random variable called the OLS parameter estimator.

Definition 2.1.2. *For a mathematical model $y(t, \theta)$ and a random process $D(\xi_N)$, the OLS estimator is defined by*

$$\hat{\Theta}_{OLS}(\xi_N) = arg \min_{\theta \in \Phi} \sum_{j=1}^{N} [D_j - y(t_j, \theta)]^2,$$

where Φ is the set of admissible parameters.

Definition 2.1.3. *The variance of the random variable $\hat{\Theta}_{OLS}(\xi_N)$, is defined to be*

$$\sigma_0^2(\xi_N) = \frac{1}{N} E[\sum_{i=1}^{N} [D_j - y(t_j, \theta_0)]^2].$$

We do not know variance $\sigma_0^2(\xi_N)$, since we do not know θ_0, but we can estimate it. By calculating the OLS parameter estimate for $d(\xi_N)$ given by (2.1.1), we can calculate the sample variance.

Definition 2.1.4. *The sample variance of the random variable $\hat{\Theta}_{OLS}(\xi_N)$, is defined by*

$$\hat{\sigma}^2(\xi_N) = \frac{1}{N-p} \sum_{i=1}^{N} [d_j - y(t_j, \hat{\theta}(\xi_N))]^2. \tag{2.1.2}$$

For a vector model, the OLS procedure requires that the $\mathcal{E}_j$'s are identically independently distributed with $E[\mathcal{E}_j] = 0$, and that they have constant variance, $var(\mathcal{E}_j) = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$. Again it is not necessary to know the distribution these random variables that represent error.

Definition 2.1.5. *For a vector mathematical model and a realization of the random process $D(\xi_N)$, the OLS estimate is*

$$\hat{\theta}_{OLS}(\xi_N) = arg \min_{\theta \in \Phi} \sum_{j=1}^{N} [d_j - y(t_j, \theta)]^T [V_0(\xi_N)]^{-1} [d_j - y(t_j, \theta)]. \tag{2.1.3}$$

Then from the definition of a OLS estimate for a vector model, we can define a OLS estimator to explain the behavior of the OLS estimate over all realizations of $D(\xi_N)$.

Definition 2.1.6. *For a vector mathematical model $y(t, \theta)$ and a random process $D(\xi_N)$, the OLS estimator is defined to be*

$$\hat{\Theta}_{OLS}(\xi_N) = arg \min_{\theta \in \Phi} \sum_{j=1}^{N} [D_j - y(t_j, \theta)]^T [V_0(\xi_N)]^{-1} [D_i - y(t_j, \theta)].$$

Definition 2.1.7. *The variance of the random variable $\hat{\Theta}(\xi_N)$ is defined to be*

$$V_0(\xi_N) = diag E\left[\frac{1}{N} \sum_{j=1}^{N} [D_j - y(t_j, \theta_0)][D_i - y(t_j, \theta_0)]^T\right].$$

Given a realization $d(\xi_N)$, of the random process $D(\xi_N)$, we can calculate the OLS estimate of θ_0 given by (2.1.3), and use it to find an unbiased estimate of the of the variance $V_0(\xi_N)$.

Definition 2.1.8. *For a realization of the random process $D(\xi_N)$ and an estimate of the random variable $\hat{\Theta}_{OLS}(\xi_N)$, a unbiased estimate of the sample variance is defined by*

$$\hat{v}(\xi_N) = diag\frac{1}{N-p}\sum_{j=1}^{N}[d_j - y(t_j, \hat{\theta}_{OLS}(\xi_N))][d_i - y(t_j, \hat{\theta}_{OLS}(\xi_N))]^T.$$

Since the calculation of the OLS estimate in (2.1.3) requires the unknown matrix $V_0(\xi_N)$, and the calculation of the estimate $\hat{v}(\xi_N)$ requires the OLS estimate we must solve for them in a coupled system

$$\hat{\theta}_{OLS}(\xi_N) = arg\min_{\theta\in\Phi}\sum_{j=1}^{N}[d_j - y(t_i,\theta)]^T[\hat{v}(\xi_N)]^{-1}[d_j - y(t_j,\theta)]$$
$$\hat{v}(\xi_N) = diag\frac{1}{N-p}\sum_{j=1}^{N}[d_j - y(t_j, \hat{\theta}_{OLS}(\xi_N))][d_j - y(t_j, \hat{\theta}_{OLS}(\xi_N))]^T.$$

The OLS estimate gives us an estimate of the true parameter, but it is only useful if we can statistically validate it. For statistical validation we will use asymptotic theory and this technique is detailed in section (2.2.1).

While we will not discuss the details of the method, we will mention that generalized least squares parameter estimation is another least squares parameter estimation technique. Generalized least squares parameter estimation selects a parameter that minimizes a weighted sum of the square Euclidean distance between the model and the data. The method requires information about the variance of the random variable that represents error. The difference between this method and OLS is that it can take into account differences of the quality or variance of the data at different times. For a detail description of this method see [8], [39].

2.1.2 Maximum Likelihood

For some models, it is reasonable to assume that the distribution of the random variables that represent error are known. In this section we make that assumption and use this extra information to help identify the unknown parameter. This technique is detailed in [16], [39]; we present a summary of the results. A maximum likelihood estimator (MLE), estimates the true parameter θ_0, by maximizing the product of the probability density functions of the random variables that represent error with respect to the parameter estimate. Suppose that the random variable $\mathcal{E}(\xi_N)$, has a distribution that is described by the probability density function $f(\varepsilon)$. Since the possible realizations of $\mathcal{E}(\xi_N)$ are functions of the unknown parameter θ, the probability density functions are more accurately defined by $f(\varepsilon|\theta)$. By the definition of the statistical model

$$\mathcal{E}(\xi_N) = [D_1 - y(t_1, \theta_0), \ldots, D_N - y(t_N, \theta_0)]^T.$$

Definition 2.1.9. *For a realization of the random variable $\mathcal{E}(\xi_N) = [\mathcal{E}_1, \ldots, \mathcal{E}_N]^T$, the likelihood function is defined to be*

$$L(\theta|\varepsilon) = f(\varepsilon|\theta). \tag{2.1.4}$$

It is important to note that the probability density function $f(\varepsilon|\theta)$ is a function of ε with fixed θ, while $L(\theta|\varepsilon)$ is a function of θ with fixed ε. Since we make the assumption that the random variables $\mathcal{E}_j$, are identically independently distributed, we can rewrite the likelihood function as

$$L(\theta|\varepsilon) = \prod_{j=1}^{N} f_j(\varepsilon_j|\theta), \tag{2.1.5}$$

where $f_j(\varepsilon_j|\theta)$ is the probability density function of $\mathcal{E}_j$. Maximum likelihood estimators estimate the unknown parameter θ_0, by finding the θ that maximizes the likelihood function, the product of the probability density functions of the random variables that represent error.

Definition 2.1.10. *For a realization of the random process $\mathcal{E}(\xi_N)$ a maximum likelihood estimate is defined by*

$$\hat{\theta}_{MLE}(\xi_N) = arg \max_{\theta \in \Phi} L(\theta|\varepsilon(\xi_N)).$$

Definition 2.1.11. *For a random process $\mathcal{E}(\xi_N)$ the maximum likelihood estimator is defined by*

$$\hat{\Theta}_{MLE}(\xi_N) = arg \max_{\theta \in \Phi} L(\theta|\mathcal{E}(\xi_N)).$$

Since we assume that the $\mathcal{E}_j$'s are identically independently distributed we can rewrite the maximum likelihood estimate as

$$\hat{\Theta}_{MLE}(\xi_N) = arg \max_{\theta \in \Phi} \prod_{j=1}^{N} f_j(\mathcal{E}(\xi_N)|\theta).$$

Maximizing $L(\theta|\mathcal{E}(\xi_N))$ is equivalent to maximizing $logL(\theta|\mathcal{E}(\xi_N))$ since the natural logarithm is a monotone increasing function. Therefore, we can rewrite the MLE as

$$\hat{\Theta}_{MLE}(\xi_N) = arg \max_{\theta \in \Phi} \sum_{j=1}^{N} log(f_j(\mathcal{E}(\xi_N)|\theta)). \tag{2.1.6}$$

We consider the loglikelihood function to possibly decrease the difficulty of solving for the MLE. These estimators are very flexible and can accommodate many different distributions of the $\mathcal{E}_j$'s. However, optimizing the likelihood function can be quite difficult. For a scalar model, if we further assume that $\mathcal{E}_j \sim N(0, \sigma^2)$ for each $j = 1, \ldots, N$, then optimizing (2.1.7) becomes a problem we are familiar with. This example illustrates one of the benefits of working with the loglikelihood function instead of the likelihood function. In this case the likelihood function is

$$L(\theta|\mathcal{E}(\xi_N)) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{1}{2\sigma}[D_j - y(t_j, \theta)]^2\}. \tag{2.1.7}$$

Maximizing (2.1.7) is equivalent to maximizing the log likelihood function, given by

$$logL(\theta|\mathcal{E}(\xi_N)) = -\frac{N}{2} log(2\pi) - \frac{N}{2} log\sigma^2 - \frac{1}{2\sigma^2} \sum_{j=1}^{N} [D_j - y(t_j, \theta)]^2. \tag{2.1.8}$$

Since the first two terms of (2.1.8) do not depend on θ, (2.1.8) is maximized when

$$\sum_{j=1}^{N} [D_j - y(t_j, \theta)]^2 \tag{2.1.9}$$

is minimized. From (2.1.9), it is clear that the MLE estimator and the OLS estimator are the same. So in the case that the error has a normal distribution, the MLE estimator is equivalently defined as

$$\hat{\Theta}_{MLE}(\xi_N) = arg\min_{\theta\in\Phi}\sum_{j=1}^{N}[D_j - y(t_j,\theta)]^2. \tag{2.1.10}$$

Definition 2.1.12. *The variance of the MLE where the random variable that represents error is assumed to have a normal distribution is given by*

$$\sigma^2(\xi_N) = \frac{1}{N}E[\sum_{j=1}^{N}[D_j - y(t_j,\theta)]^2]. \tag{2.1.11}$$

Now suppose that the statistical model has vector output and $\mathcal{E}_j \sim N_p(0, V_0)$ where

$$V_0 = diag(\sigma_{0,1}^2, \ldots, \sigma_{0,N}^2).$$

The loglikelihood function is defined by

$$\log L(\theta|\mathcal{E}(\xi_N)) = -\frac{n}{2}\sum_{i=1}^{p}\log\sigma_{0,i}^2 - \frac{1}{2}\sum_{i=1}^{p}\frac{1}{\sigma_{0,i}^2}\sum_{j=1}^{N}[D_j^i - y_i(t_j,\theta)]^2 \tag{2.1.12}$$

$$= \frac{N}{2}\sum_{i=1}^{p}\log\sigma_{0,i}^2 - \sum_{j=1}^{N}[D_j - y(t_j,\theta)]^T V^{-1}[D_j - y(t_j,\theta)]. \tag{2.1.13}$$

Then the maximum likelihood estimator for the vector model that has normally distributed error is given by

$$\hat{\Theta}_{MLE}(\xi_N) = arg\min_{\theta\in\Phi}\sum_{j=1}^{N}[D_j - y(t_j,\theta)]^T\hat{V}(\xi_N)^{-1}[D_j - y(t_j,\theta)] \tag{2.1.14}$$

$$\hat{V}(\xi_N) = diagE\left(\frac{1}{N}\sum_{j=1}^{N}[D_j - y(t_j,\hat{\Theta}_{MLE}(\xi_N))][D_j - y(t_j,\hat{\Theta}_{MLE}(\xi_N))]^T\right). \tag{2.1.15}$$

Now in the vector case we so see that the MLE and the OLS estimator are the same if we assume the random variables that represent error have a normal distribution.

2.2 Parameter Validation Techniques

Parameter validation entails describing the amount of confidence that an experimenter has in their parameter estimate. The description of the level of confidence can come in several forms. First, for problems where there is a sufficiently large set of data, we can estimate the distribution of the parameter estimator. From this distribution we can use standard procedures for assessing the quality of the estimate by constructing confidence ellipsoids and examining standard errors. Then we can say that with some statistical certainty that our estimate is close to the true parameter. We will see that for the parameter estimators we are considering, the distribution of $\hat{\Theta}(\xi_N)$ is approximately normal with mean θ_0. Second, we investigate the case when only a small amount of data can be collected and the distribution of the parameter estimator can not be estimated. In this case we will use bounded error parameter validation.

2.2.1 Asymptotic Theory

The asymptotic theory of parameter estimation describes the distribution of a parameter estimator as the number of elements in the time sample approaches infinity. We can use the distribution that the asymptotic theory produces to analyze the quality of a parameter estimator or a parameter estimate. The asymptotic theory can be used to approximate the distribution of a parameter estimator that uses a finite number of samples. The main result of the asymptotic theory [39] uses the central limit theorem to show that the OLS estimator and thus the MLE, where the errors are assumed to have a normal distribution, both have a normal distribution with expected value θ_0, asymptotically. We begin our investigation of models with scalar output with some important definitions.

Definition 2.2.1. *The sensitivity matrix* $\chi(\theta, \xi_N) = \{\chi(\theta, \xi_N)_{jk}\}$ *is defined by*

$$\chi(\theta, \xi_N)_{jk} = \frac{\partial y(t_j, \theta)}{\partial \theta_k},\ j = 1, ..., N,\ k = 1, ..., p.$$

Definition 2.2.2. *The Fisher information matrix is defined to be*

$$F(\theta, \xi_N) = \chi^T(\theta, \xi_N)\chi(\theta, \xi_N).$$

To describe the asymptotic distribution of the parameter estimator there are multiple assumptions that have to be made about the mathematical and statistical model [39].

A1 The $\mathcal{E}_j$ are i.i.d. with mean zero and variance σ_0^2

A2 For each j, $y(t_j, \theta)$ is a continuous function of θ for $\theta \in \Phi$

A3 Φ is a closed, bounded subset of $\mathbb{R}^p$

A4 Let $B_N(\theta, \bar{\theta}) = \sum_{j=1}^{N} y(t_j, \theta) y(t_j, \bar{\theta})$ and $D_N(\theta, \bar{\theta}) = \sum_{j=1}^{N} [y(t_j, \theta) - y(t_j, \bar{\theta})]^2$. (a) $N^{-1} B_N(\theta, \bar{\theta})$ converges uniformly for all θ, $\bar{\theta}$ in Φ to a function $B(\theta, \bar{\theta})$. This implies $N^{-1} D_N(\theta, \bar{\theta})$ converges uniformly for all θ, $\bar{\theta}$ in Φ to a function $D(\theta, \bar{\theta})$. (b) If it is now further assumed that $D(\theta, \theta_0) = 0$ if and only if $\theta = \theta_0$

A5 θ_0 is an interior point of Φ. Let W be an open neighborhood of θ_0 in Φ.

A6 The first and second derivatives, $\partial y(t_j, \theta)/\partial \theta_r$ and $\partial^2 y(t_j, \theta)/\partial \theta_r \partial \theta_s$ $(r, s = 1, ..., p)$, exist and are continuous for all $\theta \in W$.

A7 $1/N \sum_{j=1}^{N} (y_\theta(t_j, \theta))^T (y_\theta(t_j, \theta))$ converges to some matrix $\Omega(\theta, \xi_N)$ uniformly in $\theta \in W$.

A8 $1/N \sum_{j=1}^{N} [\partial^2 y(t_j, \theta)/\partial \theta_r \partial \theta_s]^2$ converges uniformly in θ for $\theta \in W$.

A9 $\Omega = \Omega(\theta_0, \xi_N)$ is nonsingular for $\theta \in W$.

The following result can be found in [39].

Theorem 2.2.1. *Suppose for a given a scalar mathematical model $y(t, \theta)$ and a statistical model $D(\xi_N) = y(t_j, \theta_0) + \mathcal{E}_j$ that A1-A9 are satisfied. Then the OLS parameter estimator has the distribution*

$$\sqrt{N}(\hat{\Theta}_{OLS}(\xi_N) - \theta_0) \sim N_p(0, \sigma_0^2 [lim_{N \to \infty} \frac{1}{N} \chi^T(\theta_0, \xi_N) \chi(\theta_0, \xi_N)]^{-1}) \textit{ asymptotically.} \quad (2.2.1)$$

Before we interpret this result in the context of finding confidence ellipsoids and standard errors for the OLS estimate, we will discuss some background information that will hopefully give insight into the asymptotic theory. We will develop notation that will allow us to consider what an infinite dimensional experiment represents.

Let $([0, T], \beta, m_N)$ be a measure space with $m_N([a, b]) = v_N(b) - v_N(a)$ for $[a, b] \subset [0, T]$ where

- $v_N(0) = 0$

- $v_N(T) = 1$
- v_N non decreasing
- For $t \in [0, T]$, $v_N(t) = c/N$, $c \in \mathbb{N}$ and $c \leq N$.

By definition, $v_N(\cdot)$ is a step function with at most N steps. Let $\{\hat{t}_i\}$ be the points where $v_N(\cdot)$ is discontinuous, that is the points where $v_N(\hat{t}_i) \neq v_N(\hat{t}_i - \varepsilon)$ for sufficiently small $\varepsilon > 0$. Let $\varepsilon_1 > 0$ be such that $\varepsilon_1 < \hat{t}_1$ and $\varepsilon_1 < \hat{t}_{i+1} - \hat{t}_i$ for $i = 1, \ldots, N-1$. Then we can describe an experiment ξ_N, given a measure space $([0, T], \beta, m_N)$ by

$$\xi_N = \{\hat{t}_i \text{ repeated } N * (m_N([\hat{t}_i - \varepsilon_1, \hat{t}_i])) \text{ times } \}. \tag{2.2.2}$$

For a set of measures $\{m_N\}$, defined on $([0, T], \beta)$, that satisfy $m_N \to m$, we can define a new measure space $([0, T], \beta, m)$. These measures are defined to explain what $\lim_{N\to\infty} \xi_N$ means. We can use the measures to rewrite the asymptotic variance. To do this, we first note that

$$\chi^T(\theta, \xi_N)\chi(\theta, \xi_N) = \sum_{j=1}^{N} y_\theta(t_j, \theta)^T y_\theta(t_j, \theta).$$

Suppose that the measures $\{m_N\}$ are defined on the space $([0, T], \beta)$ and satisfy $m_N \to m$. Each measure m_N corresponds to a time sample ξ_N as defined by (2.2.2). As proven by Jennrich in [24], we have that for a model that satisfies $A1 - A9$

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}\chi^T(\theta, \xi_N)\chi(\theta, \xi_N) &= \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} y_\theta(t_j, \theta)^T y_\theta(t_j, \theta) && (2.2.3)\\
&= \lim_{N\to\infty} \int_0^T y_\theta(t, \theta)^T y_\theta(t, \theta) dm_N(t) && (2.2.4)\\
&= \int_0^T y_\theta(t, \theta)^T y_\theta(t, \theta) dm(t). && (2.2.5)\\
& && (2.2.6)
\end{aligned}$$

Applying this idea to equation (2.2.1) we see that asymptotically

$$\sqrt{N}(\hat{\Theta}_{OLS}(\xi_N) - \theta_0) \sim N_p\left(0, \sigma_0^2 \left[\lim_{N\to\infty} \int_0^T y_\theta(t, \theta_0)^T y_\theta(t, \theta_0) dm_N(t)\right]^{-1}\right),$$

and for large N we have

$$\hat{\Theta}_{OLS}(\xi_N) \approx N_p\left(\theta_0, \sigma_0^2 \frac{1}{N}\left[\int_0^T y_\theta(t,\theta_0)^T y_\theta(t,\theta_0) dm_N(t)\right]^{-1}\right).$$

Since we do not know σ_0^2 or θ_0, we can approximate them by $\hat{\sigma}^2(\xi_N)$ and $\hat{\theta}(\xi_N)$ to get

$$\hat{\Theta}_{OLS}(\xi_N) \approx N_p\left(\theta_0, \hat{\sigma}^2(\xi_N) \frac{1}{N}\left[\int_0^T y_\theta(t,\hat{\theta}(\xi_N))^T y_\theta(t,\hat{\theta}(\xi_N)) dm_N(t)\right]^{-1}\right).$$

Thus, given a realization $d(\xi_N)$ of the random process $D(\xi_N)$ that corresponds to the measure m_N, we can calculate a OLS estimate $\hat{\theta}(\xi_N)$, and use (2.2.1) to create a confidence ellipsoid about the estimate. For a given percent likelihood, one way of measuring the quality of the sample times is to find the measure of the confidence ellipsoid centered at the OLS estimate. A smaller measure corresponds to more confidence in the OLS estimate. Another simpler way to describe the quality of the sample times is to find the standard errors of each component of the parameter. The standard errors for the OLS estimate are

$$SE_{\hat{\theta}_{OLS}(\xi_N)_k} = \sqrt{\left(\hat{\sigma}^2(\xi_N) \frac{1}{N}\left[\int_0^T y_\theta(t,\hat{\theta}(\xi_N))^T y_\theta(t,\hat{\theta}(\xi_N)) dm_N(t)\right]^{-1}\right)_{k,k}}$$

As we have stated, we are approximating the distribution of the parameter estimator. Getting better estimates requires a larger amount of data. As a general rule of thumb we will need a minimum of 30 data points to use the asymptotic theory.

There is a similar and equally powerful theory for parameter estimates for a vector model. The theorem for describing the asymptotic distribution of the multivariate OLS parameter estimate requires the mathematical model and the statistical model to meet the following criteria.

B1 The $\mathcal{E}_j$ are iid with mean 0 and positive definite covariance matrix Σ.

B2 For each j, the elements of $y(t_j,\theta)$ are continuous functions of θ for $\theta \in \Phi \subset \mathbb{R}^p$.

B3 Φ is a closed, bounded subset of $\mathbb{R}^p$.

B4 The $\lim N^{-1}\sum_{j=1}^N y(t_j,\theta)y(t_j,\bar{\theta})^T$ exists, and the convergence is uniform for all $\theta, \bar{\theta} \in \Phi$; $\lim N^{-1}\sum_{j=1}^N [y(t_j,\theta) - y(t_j,\theta_0)][y(t_j,\theta) - y(t_j,\theta_0)]^T$ is positive definite for all $\theta \neq \theta_0$ in Φ.

B5 θ_0 is an interior point of Φ. Let W be an open neighborhood of θ_0 in Φ.

B6 The first and second derivatives, $\partial y(t_j,\theta)/\partial\theta_r$ and $\partial^2 y(t_j,\theta)/\partial\theta_r\partial\theta_s$ $(r,s=1,...,p)$, exist and are continuous for all $\theta \in W$.

B7 $1/N\sum_{j=1}^{N}(y_\theta(t_j,\theta))^T(y_\theta(t_j,\theta))$ converges to some matrix $\Omega(\xi_N,\theta)$ uniformly in $\theta\in W$.

B8 The matrix $N^{-1}\sum_{j=1}^{N} g_1(t_j,\theta)g_2(t_j,\bar\theta)$ converges uniformly for all $\theta,\bar\theta\in\Phi$, where $g(t_j,\theta)$ and $g_2(t_j,\theta)$ are each, any one of $y(t_j,\theta)$, $\partial y(t_j,\theta)/\partial\theta_r$, and $\partial^2 y(t_j,\theta)/\partial\theta_r\partial\theta_s$ $(r,s=1,2,\ldots,p)$.

B9 For any positive definite $p\times p$ matrix R, the matrix $M_n(\theta,R)$ with $(r,s)th$ element $N^{-1}\sum_{j=1}^{N}(\partial y(t_j,\theta)/\partial\theta_r)^T R(\partial y(t_j,\theta)/\partial\theta_s)$ has a positive definite limit.

The following result can be found in [39].

Theorem 2.2.2. *Suppose for a given a vector mathematical model $y(t,\theta)$ and a statistical model $D(\xi_N)=y(t_j,\theta_0)+\mathcal{E}_j$, B1-B9 are satisfied, then the OLS parameter estimator has the distribution*

$$\hat\Theta_{OLS}(\xi_N)\sim N_p(\theta_0,\Sigma_0^N),$$

asymptotically, where

$$\Sigma_0^N \approx \left(\sum_{j=1}^{N} S_j^T(\theta,\xi_N)V_0^{-1}S_j(\theta,\xi_N)\right)^{-1}, \tag{2.2.7}$$

and

$$S_j(\theta,\xi_N)=\begin{pmatrix} \frac{\partial y_1(t_j,\theta)}{\partial\theta_1} & \cdots & \frac{\partial y_1(t_j,\theta)}{\partial\theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m(t_j,\theta)}{\partial\theta_1} & \cdots & \frac{\partial y_m(t_j,\theta)}{\partial\theta_p} \end{pmatrix}. \tag{2.2.8}$$

We do not know σ_0^2 and we do not know V_0, but we can use the OLS estimator and estimator of its variance to make the approximation

$$\hat\Sigma^N = \left(\sum_{j=1}^{N} S_j^T(\hat\theta,\xi_N)\hat V(\xi_N)^{-1}S_j(\hat\theta,\xi_N)\right)^{-1} \tag{2.2.9}$$

where $\hat{\Sigma}^N \approx \Sigma_0^N$. From this approximation we have the result that

$$\hat{\Theta}(\xi_N) \approx N_p(\hat{\theta}(\xi_N), \hat{\Sigma}^N). \tag{2.2.10}$$

2.2.2 Validating a Maximum Likelihood Estimate

Asymptotic theory can be used to verify a maximum likelihood estimate if N is sufficiently large and we make the assumption in the scalar case that $\mathcal{E}_j \sim N(0, \sigma_0^2)$, and in the vector case that $\mathcal{E}_j \sim N_m(0, V_0)$. Since the MLE is the same as the OLS estimator under these assumptions, we can use asymptotic theory to investigate the parameter estimate.

2.2.3 Bounded Error

We can use either a least squares estimator or a maximum likelihood estimator to estimate the unknown parameter given a small data set, but we can not use the corresponding statistical theory detailed in Sections 2.2.1 and 2.2.2 to validate the estimate. To validate a parameter estimated from a small data set, we make the stronger assumption that the distributions of the random variables that represent error in the statistical model are bounded. This assumption is the backbone of a parameter validation technique aptly named bounded error parameter validation or bounded error parameter identification. The method defines the membership set to be the set of parameters that map into the given bounded sets in the model's output. Bounded error parameter validation can assign a measure of quality to the estimated parameter when traditional statistical techniques cannot.

The main body of the work that has been done on bounded error parameter identification is for exact models. In 1993, Jaulin and Walter wrote a paper describing a robust bounded error parameter identification technique called set inversion via interval analysis or SILVA [23]. The method maps small sets in the parameter space to the model output using classical interval analysis as described by Moore [31]. There are three possible categorizations for a set: if the set of parameters maps into the predetermined bounded regions in the model output then it is deemed admissible, if the set maps both inside and outside the bounded regions then it is indeterminate, and if it maps outside of the bounded regions then it is not admissible. See Figure 2.1 for a graphical description. The final result of the method is an inner set and outer set that bound the true set of parameters from above and below. The method is reliable but very computationally costly and cannot be applied to approximate models.

More recently several groups have started working on bounded error identification for approximate models. Bounded error parameter estimation has been thoroughly investigated for approximate linear dynamical systems [21], however less work has been done for approximate non-linear dynamical systems. The non-linear continuous time problem was first investigated by T. Raissi, N. Ramdani, and Y. Candau in 2003 [36]. They solved it by using the work of Jualin on non-linear state estimation for ODE [22]. The method is again effective but even more time consuming because accurately estimating the state of the model given an interval in the parameter space is difficult. T. Raissi, N. Ramdani, and Y. Candau further refined the ideas in [37] by bisecting intervals of parameters in the indeterminate set. Thus, by only refining the indeterminate sets the amount of time necessary to improve accuracy is decreased.

The problem of parameter estimation becomes simpler when the system being investigated is cooperative or the model has monotone features. Kieffer and Walter [27] and Johnson and Tucker [25] studied this problem and further decreased the computational time for problems with the required characteristics.

A New Bounded Error Methodology

We are proposing a new method for bounded error parameter validation that uses the implicit function theorem to map sets in the model's output to the boundary of the membership set. The existing methods use state estimation to map sets in the parameter space into sets in the models output in order to find the membership set. The existing methods are in a sense, forward methods of brute force and one of the motivating factors for the method we have introduced is to get away from this style of problem solving. Concentrating our efforts on the boundary of the membership set can decrease computation time and give insight to the characteristics of the membership set by solving an inverse problem. All of the theorems and lemmas in this subsection are original.

All bounded error parameter validation techniques are numerically expensive because the computation cost increases exponentially with the dimension of the unknown parameter. This characteristic makes bounded error parameter validation difficult for problems where the dimension of the unknown parameter is large. To avoid this limitation an investigator may use the technique on smaller dimensional subsets of the unknown parameter.

With a sufficient background of the bounded error parameter validation problem, we introduce our method. Suppose that we assume the random variables that represent error in the statistical model are bounded. This amounts to

$$P(a_j^i \leq \mathcal{E}_j^i \leq b_j^i) = 1$$

for $a_j^i, b_j^i \in \mathbb{R}$, $j = 1, ..., N$, $i = 1, \ldots, m$. With this assumption we can build 100% confidence intervals for the parameter estimate.

Definition 2.2.3. *A scalar bounded error experiment is a set of N sampling times with corresponding intervals defined by*

$$\begin{aligned} E^N &= \{E_j^N\}_{j=1}^N && (2.2.11) \\ &= \{(t_j, [a_j, b_j])\}_{j=1}^N. && (2.2.12) \end{aligned}$$

Definition 2.2.4. *For a scalar bounded error experiment E^N, the membership set is the set of parameters $\mathcal{S}$ that satisfy*

$$\mathcal{S}(E^N, \Phi) = \{\theta \in \Phi | \; y(t_j, \theta) \in [a_j, b_j] \;\; for \;\; j = 1 \ldots N\}.$$

Figure 2.1 provides a graphical representation of how the existing bounded error parameter identification methods identify an admissible set using forward solves. Note the blue set on the left is the membership set and the intervals on the right are the bounded error experiment.

Figure 2.2 provides a graphical representation of how our inverse bounded error parameter identification method works. The figure illustrates how components of a bounded error experiment are mapped back to the parameter space. The blue set on the left is the set of parameters that the model maps into the blue interval on the right at time t_1. The green set on the left is the set of parameters that the model maps into the green interval on the right at time t_2. The black outlined set in the Φ represents the membership set and is the intersection of the blue and green set in the parameter space.

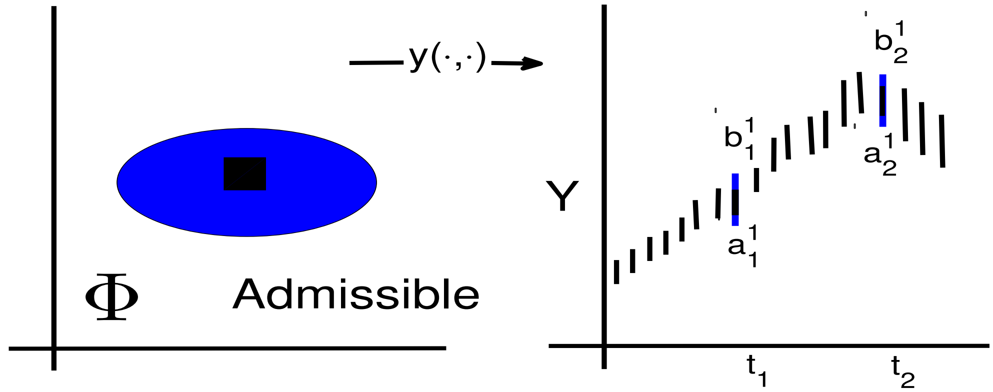


Figure 2.1: Forward method for bounded error parameter estimation.

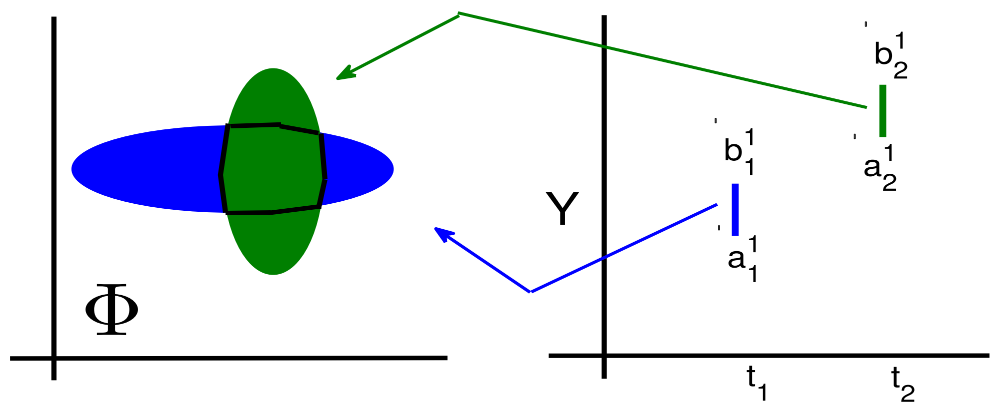


Figure 2.2: Inverse method for bounded error parameter estimation.

Now we will begin the process of describing how to find the membership sets for various models and experiments by stating and proving some theorems for exact models. We use the standard notation $\theta = [\theta_1, \theta_2, \ldots, \theta_p]^T$. Throughout this section we assume that Φ is a **bounded connected set**. Note in all of the theorems and lemmas that follow in this section require that the model is monotone with respect to one component of the unknown parameter at each t_j and all $\theta \in \Phi$. Therefore the theorems can be applied to monotone dynamical systems described in [1] but it is important to note that the systems do not have to be monotone to apply the theorems.

Theorem 2.2.3. *Let $E^1 = (t_1, [a_1, b_1])$ be a bounded error experiment.*

(i) Suppose that $y(t_1, \cdot) : \Phi \to [a_1, b_1]$ is a surjection and without loss of generality $\partial y(t_1, \theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$.

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$ and functions $z_{t_1,a_1} : Q \to \mathbb{R}^1$ and $z_{t_1,b_1} : Q \to \mathbb{R}^1$ such that either

$$\mathcal{S}(E^1, K) = \{[\theta_1, q^T]^T | q \in Q \ \ \textit{and} \ \ \theta_1 \in [z_{t_1,a_1}(q), z_{t_1,b_1}(q)]\}$$

or

$$\mathcal{S}(E^1, K) = \{[\theta_1, q^T]^T | q \in Q \ \ \textit{and} \ \ \theta_1 \in [z_{t_1,b_1}(q), z_{t_1,a_1}(q)]\}$$

where

$$K = \{\theta \in \Phi | [\theta_2, \ldots, \theta_p] \in Q\}.$$

Proof. Since $y(t_1, \cdot) : \Phi \to [a_1, b_1]$ is a surjection and $\partial y(t_j, \theta)/\partial\theta_1 \neq 0$, there exists $\theta \in \Phi$ such that $y(t_1, \theta) = a_1$. Define the function $Z : \mathbb{R}^1 \times \mathbb{R}^{p-1} \times [0, T] \times \mathbb{R}^1 \to \mathbb{R}^1$ by

$$Z(q_1, q_2, t, p) = y(t, [q_1, q_2^T]^T) - p. \tag{2.2.13}$$

Then by the implicit function theorem, there exists open sets $U \subset \mathbb{R}^{1+(p-1)}$ and $W \subset \mathbb{R}^{p-1}$, with $\theta = [\theta_1, \theta_2, \ldots, \theta_p]^T \in U$ and $[\theta_2, \ldots, \theta_p]^T \in W$ such that for each $q \in W$ there exists a unique θ_1 such that

$$[\theta_1, q^T]^T \in U \qquad \text{and} \qquad Z(\theta_1, q, t_1, a_1) = 0.$$

Furthermore, there exists a $\mathcal{C}^1$ function , $z_{t_1,a_1}(q) : W \to \mathbb{R}^1$ such that $z_{t_1,a_1}(q) = \theta_1$ and

$$Z(z_{t_1,a_1}(q), q, t_1, a_1) = 0$$

for all $q \in W$. By the same argument there exists an open set W_1 and $\mathcal{C}^1$ a function, $z_{t_1,b_1}(q) : W_1 \to \mathbb{R}^1$ such that

$$Z(z_{t_1,b_1}(q), q, t_1, b_1) = 0$$

for all $q \in W_1$. Let Q be a compact set contained in W, W_1, and Φ and define a compact set K by

$$K = \{\theta \in \Phi | [\theta_2, \ldots, \theta_p] \in Q\}.$$

Suppose that $\partial y(t_1, \theta)/\partial\theta_1 > 0$ for $\theta \in \Phi$. For each $q \in Q$, if $[\theta_1, q^T]^T \in \mathcal{S}(E_1, K)$ then $\theta_1 \in [z_{t_1,a_1}(q), z_{t_1,b_1}(q)]$ since $y(t_1, \theta)$ increases monotonically with respect to θ_1. Similarly suppose $\partial y(t_1, \theta)/\partial\theta_1 < 0$ for $\theta \in \Phi$. For each $q \in Q$ if $[\theta_1, q^T]^T \in \mathcal{S}(E_1, K)$ then $\theta_1 \in [z_{t_1,b_1}(q), z_{t_1,a_1}(q)]$ since $y(t_1, \theta)$ decreases monotonically with respect to θ_1.

□

Lemma 2.2.1. *Let $E^N = \{(t_j, [a_j, b_j])\}_{j=1}^N$ be a bounded error experiment.*

(i) Suppose that $y(t_j, \cdot) : \Phi \to [a_j, b_j]$ is a surjection and without loss of generality $\partial y(t_j, \theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$ for $j = 1, \ldots, N$.

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$, a compact set $V \subset Q$, and functions $l_{E^N} : V \to \mathbb{R}^1$ and $u_{E^N} : V \to \mathbb{R}^1$ such that

$$\mathcal{S}(E^N, K) = \{[\theta_1, q^T]^T | q \in V \;\; \textit{and} \;\; \theta_1 \in [l_{E^N}(q), u_{E^N}(q)]\} \tag{2.2.14}$$

where

$$K = \{\theta \in \Phi | [\theta_2, \ldots, \theta_p] \in Q\}. \tag{2.2.15}$$

Proof. Let $E_j^N = (t_j, [a_j, b_j])$ and without loss of generality suppose that $\partial y(t_j, \theta)/\partial\theta_1 > 0$ for $\theta \in \Phi$. By Theorem 2.2.3 for each $j = 1, \ldots, N$, there exists a compact set Q_j and functions $z_{t_j,a_j}(q) : Q_j \to \mathbb{R}^1$ and $z_{t_j,b_j}(q) : Q \to \mathbb{R}^1$ such that

$$\mathcal{S}(E_j^N, K_j) = \{[\theta_1, q^T]^T | q \in Q_j \;\; \text{and} \;\; \theta_1 \in [z_{t_j,a_j}(q), z_{t_j,b_j}(q)]\} \tag{2.2.16}$$

where

$$K_j = \{\theta \in \Phi | [\theta_2, \ldots, \theta_p] \in Q_j\}. \tag{2.2.17}$$

Let $Q = \bigcap Q_j$ and define the set

$$V = \{q \in Q| \quad \max_j z_{t_j,a_j}(q) \leq \min_j z_{t_j,b_j}(q)\}.$$

Now define the function $l_{E^N} : V \to \mathbb{R}^1$ by

$$l_{E^N}(q) = \max_j z_{t_j,a_j}(q) \tag{2.2.18}$$

and the function $u_{E^N} : V \to \mathbb{R}^1$ by

$$u_{E^N}(q) = \min_j z_{t_j,b_j}(q). \tag{2.2.19}$$

By definition we have that

$$\begin{aligned} \mathcal{S}(E^N, K) &= \bigcap_j \mathcal{S}(E_j^N, K) &(2.2.20)\\ &= \{[\theta_1, q^T]^T | q \in Q \quad \text{and} \quad \theta_1 \in [l_{E^N}(q), u_{E^N}(q)]\} &(2.2.21)\\ &= \{[\theta_1, q^T]^T | q \in V \quad \text{and} \quad \theta_1 \in [l_{E^N}(q), u_{E^N}(q)]\}. &(2.2.22) \end{aligned}$$

□

Using Theorem 2.2.3 and Lemma 2.2.1 we know there exist functions that describe the boundary of the set $\mathcal{S}(E^N, K)$. It is possible, for select problems, to solve for these boundary functions explicitly, however, this cannot be done for the majority of the problems in which we have interest. While we can not explicitly solve for them exactly, we can explicitly solve for approximations of these functions. The next theorem outlines a procedure for finding approximations of the boundary functions.

Theorem 2.2.4. *Let $E^1 = (t_1, [a_1, b_1])$ be a bounded error experiment.*

(i) Suppose that $y(t_1, \cdot) : \Phi \to [a_1, b_1]$ is a surjection and without loss of generality $\partial y(t_1, \theta)/\partial \theta_1 \neq 0$ for $\theta \in \Phi$. Let Q, $z_{t_1,a_1} : Q \to \mathbb{R}^1$ and $z_{t_1,b_1} : Q \to \mathbb{R}^1$ be defined by Theorem 2.2.3.

Then there exists functions $z_{t_1,a_1}^{\pi^\delta,k} : Q \to \mathbb{R}^1$ and $z_{t_1,b_1}^{\pi^\delta,k} : Q \to \mathbb{R}^1$ such that

$$\lim_{\delta\to 0}\lim_{k\to\infty}\int_Q [z_{t_1,a_1}^{\pi^\delta,k}(q) - z_{t_1,a_1}(q)]^2 dq = 0$$
$$\lim_{\delta\to 0}\lim_{k\to\infty}\int_Q [z_{t_1,b_1}^{\pi^\delta,k}(q) - z_{t_1,b_1}(q)]^2 dq = 0.$$

Further, $z_{t_1,a_1}^{\pi^\delta,k}$ and $z_{t_1,b_1}^{\pi^\delta,k}$ can be explicitly solved for.

Proof. Let π^δ be a δ-fine partition of Q. Consider the function $Z : \mathbb{R}^1\times\mathbb{R}^{p-1}\times[0,T]\times\mathbb{R}^1 \to \mathbb{R}^1$ defined by

$$Z(q_1, q_2, t, a) = y(t, [q_1, q_2^T]^T) - a. \tag{2.2.23}$$

For fixed $q_2 = r_i \in \pi^\delta$, $t = t_1$, and $a = a_1$, let $N^k(r_i, t_1, a_1, r_i^0)$ be the k^{th} iterate of Newton's method on (2.2.23) with initial guess r_i^0. Since $Z(\cdot, r_i, t_1, a_1)$ is continuously differentiable with respect to q_1, for an appropriate initial guess r_i^0, Newton's method will converge uniformly to the solution of $Z(q_1, r_i, t_1, a_1) = 0$.

Let $N^{\pi^\delta,k} = [N^k(r_1, t_1, a_1, r_1^0), N^k(r_2, t_1, a_1, r_2^0), \ldots, N^k(r_M, t_1, a_1, r_M^0)]^T$ and define the function $z_{t_1,a_1}^{\pi^\delta,k} : Q \to \mathbb{R}^1$ to be the linear interpolation of the the points $N^{\pi^\delta,k}$. Then it follows that

$$\lim_{\delta\to 0}\lim_{k\to\infty}\int_Q [z_{t_1,a_1}^{\pi^\delta,k}(q) - z_{t_1,a_1}(q)]^2 dq = 0.$$

Using a similar argument the function $z_{t_1,b_1}^{\pi^\delta,k} : Q \to \mathbb{R}^1$ can be defined and

$$\lim_{\delta\to 0}\lim_{k\to\infty}\int_Q [z_{t_1,b_1}^{\pi^\delta,k}(q) - z_{t_1,b_1}(q)]^2 dq = 0.$$

□

Lemma 2.2.2. *Let $E^N = \{(t_j, [a_j, b_j])\}_{j=1}^N$ be a bounded error experiment.*

(i) Suppose that $y(t_j, \cdot) : \Phi \to [a_j, b_j]$ is a surjection and without loss of generality $\partial y(t_j, \theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$ for $j = 1, \ldots, N$. Let the functions l_{E^N} and u_{E^N} be defined by Lemma (2.2.1).

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$*, a compact set* $V \subset Q$*, and a functions* $l_{E^N}^{\pi^\delta,k} : V \to \mathbb{R}^1$ *and* $u_{E^N}^{\pi^\delta,k} : V \to \mathbb{R}^1$ *such that*

$$\lim_{\delta\to 0}\lim_{k\to\infty}\int_V l_{E^N}^{\pi^\delta,k}(q)dq = \int_V l_{E^N}(q)dq$$
$$\lim_{\delta\to 0}\lim_{k\to\infty}\int_V u_{E^N}^{\pi^\delta,k}(q)dq = \int_V u_{E^N}(q)dq.$$

Furthermore, $l_{E^N}^{\pi^\delta,k}$ *and* $u_{E^N}^{\pi^\delta,k}$ *can be solved for explicitly.*

Proof. Without loss of generality suppose that $\partial y(t_j,\theta)/\partial\theta_1 > 0$ for $\theta \in \Phi$ and $t \in [0,T]$. Let z_{t_j,a_j} and z_{t_j,b_j} be the functions and Q_j be the set defined by Theorem 2.2.3. Suppose that the set V defined by Lemma 2.2.1 is not empty and consider the the function l_{E^N} also defined by Lemma 2.2.1. This function l_{E^N} is composed of the functions z_{t_j,a_j}. Let $Q = \bigcap_{j=1}^N Q_j$ and define the sets

$$A_j = \{q \in Q|\ \ z_{t_j,a_j}(q) = l_{E^N}(q)\}$$

and

$$A_0 = Q\backslash\bigcup_{j=1}^N A_j.$$

If $A_i \bigcap A_j \neq 0$ for $i < j$ then define $B_i = A_i\backslash A_j$ and if $A_i \bigcap A_j = \emptyset$ for $j \neq i$, then let $B_i = A_i$. Also let $B_0 = A_0$. Then $\bigcup_{j=0}^N B_j = Q$, and $B_i \bigcap B_j = \emptyset$ for $i \neq j$. Define the set

$$V = \bigcup_{j=1}^N B_j. \tag{2.2.24}$$

Let $[r_1, r_2, \ldots, r_M] = \pi^\delta$ be a δ-fine partition of Q and let $[r_1^0, r_2^0, \ldots, r_M^0]$ be an associated set of initial guesses for Newton's method. Let the functions $z_{t_j,a_j}^{\pi^\delta,k}$ for $j = 1,\ldots,N$ be defined by Theorem 2.2.4. We define

$$l_{E^N}^{\pi^\delta,k}(q) = \begin{cases} z_{t_1,a_1}^{\pi^\delta,k}(q) & \text{if} \quad q \in B_1 \\ z_{t_2,a_2}^{\pi^\delta,k}(q) & \text{if} \quad q \in B_2 \\ \vdots & \\ z_{t_N,a_N}^{\pi^\delta,k}(q) & \text{if} \quad q \in B_N. \end{cases} \tag{2.2.25}$$

By Theorem 2.2.4 we have that

$$\begin{aligned}\lim_{\delta\to 0}\lim_{k\to\infty}\int_V l_{E^N}^{\pi^\delta,k}(q)dq &= \lim_{\delta\to 0}\lim_{k\to\infty}\sum_{j=0}^{N}\int_{B_j} z_{t_j,a_j}^{\pi^\delta,k}(q)dq\\ &= \sum_{j=0}^{N}\int_{B_j} z_{t_j,a_j}(q)dq\\ &= \int_V l_{E^N}(q)dq.\end{aligned}$$

□

Using the previous theorems and lemmas we can now explicitly construct the membership set for exact scalar models. Now we will consider exact vector models.

Definition 2.2.5. *Suppose that mathematical model has vector output in* $\mathbb{R}^m$*. A vector bounded error experiment is defined to be*

$$\begin{aligned}E^{N,m} &= \{E_{j,i}^{N,m}\}_{i=1\ldots m}^{j=1,\ldots,N} &(2.2.26)\\ &= \{(t_j,[a_j^i,b_j^i])\}_{i=1,\ldots,m}^{j=1,\ldots,N}. &(2.2.27)\end{aligned}$$

Definition 2.2.6. *For a model with vector output and a vector bounded error experiment* $E^{N,m}$*, the membership set is defined to be*

$$\mathcal{S}(E^{N,m},K) = \{\theta\in K|\ y_i(t_j,\theta)\in[a_j^i,b_j^i]\ \ for\ \ i=1\ldots m, j=1\ldots N\}. \tag{2.2.28}$$

Lemma 2.2.3. *Let* $E^{N,m} = \{(t_j,[a_j^i,b_j^i])\}_{i=1,\ldots,m}^{j=1,\ldots,N}$ *be a vector bounded error experiment.*

(i) Suppose that $y_i(t_j,\cdot):\Phi\to[a_j^i,b_j^i]$ *is a surjection and without loss of generality* $\partial y_i(t_j,\theta)/\partial\theta_1\neq 0$ *for* $\theta\in\Phi$ *and* $j=1,\ldots,N$, $i=1,\ldots,m$.

Then there exists a compact set $Q\subset\mathbb{R}^{p-1}$ *, a compact set* $V\subset Q$*, and functions* $l_{E^{N,m}}$:

$V \to \mathbb{R}^1$ and $u_{E^{N,m}} : V \to \mathbb{R}^1$ such that

$$\mathcal{S}(E^{N,m}, K) = \{[\theta_1, q^T]^T | q \in V \ \ and \ \ \theta_1 \in [l_{E^{N,m}}(q), u_{E^{N,m}}(q)]\} \tag{2.2.29}$$

where

$$K = \{\theta \in \Phi | [\theta_2, \ldots, \theta_p] \in Q\}. \tag{2.2.30}$$

Proof. The proof follows from Lemma 2.2.1.

□

Just as in the scalar case, sometimes the functions $l_{E^{N,m}}$ and $u_{E^{N,m}}$ can be solved for explicitly, but most of the time, they can not. However, we can find explicit approximations. The next theorem details the procedure for finding these approximations.

Lemma 2.2.4. *Let $E^{N,m} = \{(t_j, [a^i_j, b^i_j])\}^{j=1,\ldots,N}_{i=1,\ldots,m}$ be a vector bounded error experiment.*

(i) Suppose that $y_i(t_j, \cdot) : \Phi \to [a^i_j, b^i_j]$ is a surjection and without loss of generality $\partial y_i(t_j, \theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$ and $j = 1, \ldots, N$, $i = 1, \ldots, m$.

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$, a compact set $V \subset Q$, and functions $l^{\pi^\delta,k}_{E^{N,m}}$: $Q \to \mathbb{R}^1$ and $u^{\pi^\delta,k}_{E^{N,m}} : Q \to \mathbb{R}^1$ such that

$$\lim_{\delta\to 0}\lim_{k\to\infty}\int_V l^{\pi^\delta,k}_{E^{N,m}}(q)dq = \int_V l_{E^{N,m}}(q)dq$$
$$\lim_{\delta\to 0}\lim_{k\to\infty}\int_V u^{\pi^\delta,k}_{E^{N,m}}(q)dq = \int_V u_{E^{N,m}}(q)dq.$$

Furthermore, can be $l^{\pi^\delta,k}_{E^{N,m}}$ and $u^{\pi^\delta,k}_{E^{N,m}}$ can be solved for explicitly.

Proof. The proof follows from Lemma 2.2.2. □

If the mathematical model can not be solved for exactly, as is often the case when the model is described by a differential equation, the mathematical model must be approximated. Suppose that $\rho^\mu = [p_1, \ldots, p_M]$ is a μ-fine partition of $[0, T]$. Further suppose that $[w^{\rho^\mu}_1(p_1, \theta), \ldots, w^{\rho^\mu}_M(p_M, \theta)]$ is the approximation of $[y(p_1, \theta), \ldots, y(p_M, \theta)]$. Define $w^{\rho^\mu} : [0, T], \Phi \to R^m$ to be the linear interpolation of the points $\{(p_j, w^{\rho^\mu}_j(p_j, \theta))\}$.

In order to prove that the approximate membership set converges to the exact membership set for problems where the model must be approximate we have to make assumptions about the numerical scheme used to approximate the model. We will label the assumption (ii) and refer to it in several of the following theorems,

(ii) Let $\{\rho^\mu\}$ be a set of partition on $[0,T]$ with $\mu \to 0$ and suppose for a given numerical scheme $\lim_{\mu\to 0}\|s(t,\theta) - s^{\rho^\mu}(t,\theta)\|_\infty = 0$ and $\lim_{\mu\to 0}\|y(t,\theta) - w^{\rho^\mu}(t,\theta)\|_\infty = 0$. Further suppose that $w^{\rho^\mu}(t,\theta)$ is continuously differentiable with respect to θ_1.

Theorem 2.2.5. *Let $E^1 = (t_1, [a_1, b_1])$ be a bounded error experiment and suppose that (ii) holds.*

(i) Suppose that for each ρ^μ, $w^{\rho^\mu}(t_1,\cdot) : \Phi \to [a_1, b_1]$ is a surjection and without loss of generality is differentiable with respect to θ_1 and $\partial w^{\rho^\mu}(t_1,\theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$.

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$ and functions $z^{\rho^\mu}_{t_1,a_1} : Q \to \mathbb{R}^1$ and $z^{\rho^\mu}_{t_1,b_1} : Q \to \mathbb{R}^1$ such that

$$\lim_{\mu\to 0}\int_Q z^{\rho^\mu}_{t_1,a_1}(q)dq = \int_Q z_{t_1,a_1}(q)dq$$
$$\lim_{\mu\to 0}\int_Q z^{\rho^\mu}_{t_1,b_1}(q)dq = \int_Q z_{t_1,b_1}(q)dq.$$

Proof. Since $w^{\rho^\mu}(t_1,\cdot) : \Phi \to [a_1,b_1]$ is a surjection and $\partial w^{\rho^\mu}(t_j,\theta)/\partial\theta_1 \neq 0$, there exists $\theta \in \Phi$ such that $w^{\rho^\mu}(t_1,\theta) = a_1$. Define the function $Z^{\rho^\mu} : \mathbb{R}^1 \times \mathbb{R}^{p-1} \times [0,T] \times \mathbb{R}^1 \to \mathbb{R}^1$ by

$$Z^{\rho^\mu}(q_1, q_2, t, p) = w^{\rho^\mu}(t, [q_1, q_2^T]^T) - p. \tag{2.2.31}$$

Then by the implicit function theorem, there exists open sets $U^{\rho^\mu} \subset \mathbb{R}^{1+(p-1)}$ and $W^{\rho^\mu} \subset \mathbb{R}^{p-1}$, with $\theta = [\theta_1, \theta_2, \ldots, \theta_p]^T \in U^{\rho^\mu}$ and $[\theta_2,\ldots,\theta_p]^T \in W^{\rho^\mu}$ such that for each $q \in W$ there exists a unique θ_1 with

$$[\theta_1, q^T]^T \in U^{\rho^\mu} \qquad \text{and} \qquad Z^{\rho^\mu}(\theta_1, q, t_1, a_1) = 0.$$

Furthermore, there exists a $\mathcal{C}^1$ function , $z^{\rho^\mu}_{t_1,a_1}(q) : W \to \mathbb{R}^1$ such that $z^{\rho^\mu}_{t_1,a_1}(q) = \theta_1$ and

$$Z^{\rho^\mu}(z^{\rho^\mu}_{t_1,a_1}(q), q, t_1, a_1) = 0$$

for all $q \in W^{\rho^\mu}$. By the same argument there exists an open set $W_1^{\rho^\mu}$ and $\mathcal{C}^1$ a function, $z_{t_1,b_1}^{\rho^\mu}(q) : W_1 \to \mathbb{R}^1$ such that

$$Z^{\rho^\mu}(z_{t_1,b_1}^{\rho^\mu}(q), q, t_1, b_1) = 0$$

for all $q \in W_1^{\rho^\mu}$. Let Q^{ρ^μ} be a compact set contained in W^{ρ^μ}, $W_1^{\rho^\mu}$, and Φ and define a compact set K^{ρ^μ} by

$$K^{\rho^\mu} = \{\theta \in \Phi | [\theta_2, \ldots, \theta_p] \in Q^{\rho^\mu}\}.$$

We have that $\lim_{\mu\to 0} |Z^{\rho^\mu}(q_1, q_2, t, p) - Z(q_1, q_2, t, p)| = 0$ for all $[q_1, q_2^T]^T \in \Phi$, $t \in [0, T]$ and $p \in \mathbb{R}$. Suppose for $q_2 \in Q \bigcap Q^{\rho^\mu}$, $\lim_{\mu\to 0} Z^{\rho^\mu}(\theta_1^{\rho^\mu}, q_2, t, a) = Z(\theta_1, q_2, t, a) = 0$ and that $\lim_{\mu\to 0} \theta_1^{\rho^\mu} \neq \theta_1$. Then $\lim_{\mu\to 0} \theta_1^{\rho^\mu} = \hat{\theta}_1$ and both $\lim_{\mu\to 0} Z^{\rho^\mu}(\hat{\theta}_1, q_2, t, a) = 0$ and $\lim_{\mu\to 0} Z^{\rho^\mu}(\theta_1^{\rho^\mu}, q_2, t, a) = 0$. This is a contradiction by the Implicit Function Theorem so $\lim_{\mu\to 0} \theta_1^{\rho^\mu} = \theta_1$. We conclude that $\lim_{\mu\to 0} z_{t_1,a_1}^{\rho^\mu}(q) = z_{t_1,a_1}(q)$ for all $q \in Q$. □

Lemma 2.2.5. *Let $E^N = \{(t_j, [a_j, b_j])\}$ be a bounded error experiment and suppose that (ii) holds.*

(i) Suppose that for each ρ^μ and $j = 1, \ldots, N$, $w^{\rho^\mu}(t_j, \cdot) : \Phi \to [a_j, b_j]$ is a surjection and without loss of generality is differentiable with respect to θ_1 and $\partial w^{\rho^\mu}(t_j, \theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$.

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$, a compact set $V^{\rho^\mu} \subset Q$, and functions $l_{E^N}^{\rho^\mu} : V \to \mathbb{R}^1$ and $u_{E^N}^{\rho^\mu} : V \to \mathbb{R}^1$ such that

$$\lim_{\mu\to 0} \int_{V^{\rho^\mu}} l_{E^N}^{\rho^\mu}(q)dq = \int_V l_{E^N}(q)dq$$
$$\lim_{\mu\to 0} \int_{V^{\rho^\mu}} u_{E^N}^{\rho^\mu}(q)dq = \int_V u_{E^N}(q)dq.$$

Proof. Let $\{B_j\}$ be the set of sets defined in the proof of Lemma 2.2.2. Define the function

$$l_{E^N}^{\rho^\mu}(q) = \begin{cases} z_{t_1,a_1}^{\rho^\mu}(q) & \text{if} \quad q \in B_1 \\ z_{t_2,a_2}^{\rho^\mu}(q) & \text{if} \quad q \in B_2 \\ \vdots & \\ z_{t_N,a_N}^{\rho^\mu}(q) & \text{if} \quad q \in B_N. \end{cases} \tag{2.2.32}$$

By Theorem 2.2.5 we have that

$$\begin{aligned}\lim_{\mu\to 0}\int_V l_{E^N}^{\rho^\mu}(q)dq &= \lim_{\mu\to 0}\sum_{j=0}^{N}\int_{B_j} z_{t_j,a_j}^{\rho^\mu}(q)dq\\ &= \sum_{j=0}^{N}\int_{B_j} z_{t_j,a_j}(q)dq\\ &= \int_V l_{E^N}(q)dq.\end{aligned}$$

□

Theorem 2.2.6. *Let $E^1 = (t_1, [a_1, b_1])$ be a bounded error experiment and suppose that (ii) holds.*

(i) Suppose that for each ρ^μ, $w^{\rho^\mu}(t_1, \cdot) : \Phi \to [a_1, b_1]$ is a surjection and without loss of generality is differentiable with respect to θ_1 and $\partial w^{\rho^\mu}(t_1, \theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$. Let Q, $z_{t_1,a_1} : Q \to \mathbb{R}^1$ and $z_{t_1,b_1} : Q \to \mathbb{R}^1$ be defined by Theorem 2.2.3.

Then there exists functions $z_{t_1,a_1}^{\rho^\mu\pi^\delta,k} : Q \to \mathbb{R}^1$ and $z_{t_1,b_1}^{\rho^\mu\pi^\delta,k} : Q \to \mathbb{R}^1$ such that

$$\lim_{\mu\to 0}\lim_{\delta\to 0}\lim_{k\to\infty}\int_Q [z_{t_1,a_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,a_1}(q)]^2 dq = 0$$
$$\lim_{\mu\to 0}\lim_{\delta\to 0}\lim_{k\to\infty}\int_Q [z_{t_1,b_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,b_1}(q)]^2 dq = 0.$$

Further, $z_{t_1,a_1}^{\rho^\mu\pi^\delta,k}$ and $z_{t_1,b_1}^{\rho^\mu\pi^\delta,k}$ can be solved for explicitly.

Proof. For $\varepsilon > 0$, there exist π^δ and k such that

$$\int_Q [z_{t_1,a_1}^{\pi^\delta,k}(q) - z_{t_1,a_1}(q)]^2 dq < \varepsilon, \tag{2.2.33}$$

by Theorem 2.2.4. Theorem 2.2.5 implies that there exists a ρ^μ such that

$$\int_Q [z_{t_1,a_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,a_1}^{\pi^\delta,k}(q)]^2 dq < \varepsilon. \tag{2.2.34}$$

Then we have that

$$\int_Q [z_{t_1,a_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,a_1}(q)]^2 dq = \int_Q [z_{t_1,a_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,a_1}^{\pi^\delta,k}(q) + z_{t_1,a_1}^{\pi^\delta,k}(q) - z_{t_1,a_1}(q)]^2 dq \quad (2.2.35)$$

$$\leq \int_Q [z_{t_1,a_1}^{\rho^\mu\pi^\delta,k}(q) - z_{t_1,a_1}^{\pi^\delta,k}(q)]^2 dq + \int_Q [z_{t_1,a_1}^{\pi^\delta,k}(q) - z_{t_1,a_1}(q)]^2 dq \quad (2.2.36)$$

$$< 2\varepsilon. \quad (2.2.37)$$

The proof that

$$\lim_{\mu\to 0}\lim_{\delta\to 0}\lim_{k\to\infty}\int_Q [z_{t_1,b_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,b_1}(q)]^2 dq = 0, \quad (2.2.38)$$

follows similarly.

□

To measure the distance between sets in $\mathbb{R}^p$ we will use the Hausdorff distance.

Definition 2.2.7. *The Hausdorff distance between two sets $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^m$ is defined to be*

$$d_H(X,Y) = \max\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\} \quad (2.2.39)$$

Lemma 2.2.6. *Let $E^1 = (t_1, [a_1, b_1])$ be a bounded error experiment and suppose that (ii) holds.*

(i) Suppose that for each ρ^μ, $w^{\rho^\mu}(t_1,\cdot) : \Phi \to [a_1, b_1]$ is a surjection and without loss of generality is differentiable with respect to θ_1 and $\partial w^{\rho^\mu}(t_1,\theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$. Let $Q \subset \mathbb{R}^{p-1}$, such that $z_{t_1,a_1}^{\rho^\mu\pi^\delta,k} : Q \to \mathbb{R}^1$ and $z_{t_1,b_1}^{\rho^\mu\pi^\delta,k} : Q \to \mathbb{R}^1$ are defined by Theorem 2.2.6.

Define the set

$$\mathcal{S}^{\rho^\mu,\pi^\delta,k}(E^1,K) = \{[\theta_1, q^T]^T | q \in Q \ \ and \ \ \theta_1 \in [z_{t_1,a_1}^{\rho^\mu,\pi^\delta,k}(q), z_{t_1,b_1}^{\rho^\mu,\pi^\delta,k}(q)]\}$$

or

$$\mathcal{S}^{\rho^\mu,\pi^\delta,k}(E^1,K) = \{[\theta_1, q^T]^T | q \in Q \ \ and \ \ \theta_1 \in [z_{t_1,b_1}^{\rho^\mu,\pi^\delta,k}(q), z_{t_1,a_1}^{\rho^\mu,\pi^\delta,k}(q)]\}$$

where

$$K = \{\theta \in \Phi | [\theta_2, \dots, \theta_p] \in Q\}$$

Then

$$\lim_{\mu \to 0} \lim_{\delta \to 0} \lim_{k \to \infty} d_H(\mathcal{S}^{\rho^\mu, \pi^\delta, k}(E^1, K), \mathcal{S}(E^1, K)) = 0.$$

Proof. Since $z_{t_1,a_1}^{\rho^\mu,\pi^\delta,k}(\cdot)$, $z_{t_1,a_1}(\cdot)$, $z_{t_1,b_1}^{\rho^\mu,\pi^\delta,k}(\cdot)$, and $z_{t_1,b_1}(\cdot)$ are continuous it follows from Theorem 2.2.6 that

$$\lim_{\mu \to 0} \lim_{\delta \to 0} \lim_{k \to \infty} \int_Q [z_{t_1,a_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,a_1}(q)]^2 dq = 0$$
$$\lim_{\mu \to 0} \lim_{\delta \to 0} \lim_{k \to \infty} \int_Q [z_{t_1,b_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,b_1}(q)]^2 dq = 0,$$

and

$$\lim_{\mu \to 0} \lim_{\delta \to 0} \lim_{k \to \infty} sup_{q \in Q} \| z_{t_1,b_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,b_1}(q) \|_2 = 0$$
$$\lim_{\mu \to 0} \lim_{\delta \to 0} \lim_{k \to \infty} sup_{q \in Q} \| z_{t_1,b_1}^{\rho^\mu,\pi^\delta,k}(q) - z_{t_1,b_1}(q) \|_2 = 0.$$

Since the sets $\mathcal{S}^{\rho^\mu,\pi^\delta,k}(E^1, K)$ and $\mathcal{S}(E^1, K)$ are both compact and the boundary of these sets converge, the result follows.

□

Lemma 2.2.7. *Let $E^N = \{(t_j, [a_j, b_j])\}$ be a bounded error experiment and suppose that (ii) holds.*

(i) Suppose that for each ρ^μ and $j = 1, \dots, N$, $w^{\rho^\mu}(t_j, \cdot) : \Phi \to [a_j, b_j]$ is a surjection and without loss of generality is differentiable with respect to θ_1 and $\partial w^{\rho^\mu}(t_j, \theta)/\partial \theta_1 \neq 0$ for $\theta \in \Phi$.

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$, a compact set $V \subset Q$, and a functions $l_{E^N}^{\rho^\mu,\pi^\delta,k} : V \to \mathbb{R}^1$ and $u_{E^N}^{\rho^\mu,\pi^\delta,k} : V \to \mathbb{R}^1$ such that

$$\lim_{\mu\to 0}\lim_{\delta\to 0}\lim_{k\to\infty}\int_V l_{E^N}^{\rho^\mu,\pi^\delta,k}(q)dq = \int_V l_{E^N}(q)dq$$
$$\lim_{\mu\to 0}\lim_{\delta\to 0}\lim_{k\to\infty}\int_V u_{E^N}^{\rho^\mu,\pi^\delta,k}(q)dq = \int_V u_{E^N}(q)dq.$$

Furthermore, $l_{E^N}^{\rho^\mu,\pi^\delta,j}$ *and* $l_{E^N}^{\rho^\mu,\pi^\delta,j}$ *can be solved for explicitly.*

Proof. The result follows from Lemma 2.2.1, Lemma 2.2.2, and Lemma 2.2.5. □

Lemma 2.2.8. *Let* $E^N = (t_j, [a_j, b_j])$ *be a bounded error experiment and suppose that (ii) holds.*

(i) Suppose that for each ρ^μ, $w^{\rho^\mu}(t_j, \cdot) : \Phi \to [a_j, b_j]$ *is a surjection for* $j = 1, \ldots, N$ *and without loss of generality is differentiable with respect to* θ_1 *and* $\partial w^{\rho^\mu}(t_1, \theta)/\partial\theta_1 \neq 0$ *for* $\theta \in \Phi$. *Let* $V \subset Q \subset \mathbb{R}^{p-1}$, *such that* $l_{E^N}^{\rho^\mu\pi^\delta,k} : V \to \mathbb{R}^1$ *and* $u_{E^N}^{\rho^\mu\pi^\delta,k} : V \to \mathbb{R}^1$ *are defined by Theorem 2.2.6.*

Define the set

$$\mathcal{S}^{\rho^\mu,\pi^\delta,k}(E^1, K) = \{[\theta_1, q^T]^T | q \in Q \ \ \text{and} \ \ \theta_1 \in [l_{E^N}^{\rho^\mu,\pi^\delta,k}(q), u_{E^N}^{\rho^\mu,\pi^\delta,k}(q)]\}$$

where

$$K = \{\theta \in \Phi | [\theta_2, \ldots, \theta_p] \in Q\}.$$

Then

$$\lim_{\mu\to 0}\lim_{\delta\to 0}\lim_{k\to\infty} d_H(\mathcal{S}^{\rho^\mu,\pi^\delta,k}(E^N, K), \mathcal{S}(E^N, K)) = 0.$$

Proof. The result follows from Lemma 2.2.7. □

Lemma 2.2.9. *Let* $E^{N,m} = \{(t_j, [a_j^i, b_j^i])\}$ *be a bounded error experiment and suppose that (ii) holds.*

(i) Suppose that for each ρ^μ, $j = 1, \ldots, N$, *and* $i = 1, \ldots, m$, $w_i^{\rho^\mu}(t_j, \cdot) : \Phi \to [a_j^i, b_j^i]$ *is a surjection and without loss of generality is differentiable with respect to* θ_1 *and* $\partial w_i^{\rho^\mu}(t_j, \theta)/\partial\theta_1 \neq$

0 *for* $\theta \in \Phi$.

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$, *a compact set* $V \subset Q$, *and a functions* $l^{\rho^\mu}_{E^{N,m}} : V \to \mathbb{R}^1$ *and* $u^{\rho^\mu}_{E^{N,m}} : V \to \mathbb{R}^1$ *such that*

$$\lim_{\mu \to 0} \int_V l^{\rho^\mu}_{E^{N,m}}(q)dq = \int_V l_{E^{N,m}}(q)dq$$
$$\lim_{\mu \to 0} \int_V u^{\rho^\mu}_{E^{N,m}}(q)dq = \int_V u_{E^{N,m}}(q)dq.$$

Proof. The result follows from Lemma 2.2.3 and Lemma 2.2.5. □

Lemma 2.2.10. *Let* $E^{N,m} = \{(t_j, [a^i_j, b^i_j])\}$ *be a bounded error experiment and suppose that* (ii) *holds.*

(i) Suppose that for each ρ^μ, $j = 1, \ldots, N$, *and* $i = 1, \ldots, m$, $w^{\rho^\mu}_i(t_j, \cdot) : \Phi \to [a^i_j, b^i_j]$ *is a surjection and without loss of generality is differentiable with respect to* θ_1 *and* $\partial w^{\rho^\mu}_i(t_j, \theta)/\partial\theta_1 \neq 0$ *for* $\theta \in \Phi$.

Then there exists a compact set $Q \subset \mathbb{R}^{p-1}$, *a compact set* $V \subset Q$, *and a functions* $l^{\rho^\mu,\pi^\delta,k}_{E^{N,m}} : V \to \mathbb{R}^1$ *and* $u^{\rho^\mu\pi^\delta,k}_{E^{N,m}} : V \to \mathbb{R}^1$ *such that*

$$\lim_{\mu \to 0} \lim_{\delta \to 0} \lim_{k \to \infty} \int_V l^{\rho^\mu\pi^\delta,k}_{E^{N,m}}(q)dq = \int_V l_{E^{N,m}}(q)dq$$
$$\lim_{\mu \to 0} \lim_{\delta \to 0} \lim_{k \to \infty} \int_V u^{\rho^\mu\pi^\delta,k}E^{N,m}(q)dq = \int_V u_{E^{N,m}}(q)dq.$$

Furthermore, can be $l^{\rho^\mu,\pi^\delta,j}_{E^{N,m}}$ *and* $u^{\rho^\mu,\pi^\delta,j}_{E^{N,m}}$ *can be solved for explicitly.*

Proof. The result follows from Lemma 2.2.7 and Lemma 2.2.9 □

Lemma 2.2.11. *Let* $E^{N,m} = (t_j, [a^i_j, b^i_j])$ *be a bounded error experiment and suppose that* (ii) *holds.*

(i) Suppose that for each ρ^μ, $w^{\rho^\mu}_i(t_j, \cdot) : \Phi \to [a^i_j, b^i_j]$ *is a surjection for* $j = 1, \ldots, N$ *and* $i = 1, \ldots, m$ *and without loss of generality is differentiable with respect to* θ_1 *and*

$\partial w_i^{\rho^\mu}(t_1, \theta)/\partial\theta_1 \neq 0$ for $\theta \in \Phi$. Let $V \subset Q \subset \mathbb{R}^{p-1}$, such that $l_{E^{N,m}}\rho^\mu\pi^\delta, k : V \to \mathbb{R}^1$ and $u_{E^{N,m}}^{\rho^\mu\pi^\delta,k} : V \to \mathbb{R}^1$ are defined by Lemma 2.2.10.

Define the set

$$\mathcal{S}^{\rho^\mu,\pi^\delta,k}(E^{N,m}, K) = \{[\theta_1, q^T]^T | q \in Q \text{ and } \theta_1 \in [l_{E^{N,m}}^{\rho^\mu,\pi^\delta,k}(q), u_{E^{N,m}}^{\rho^\mu,\pi^\delta,k}(q)]\}$$

where

$$K = \{\theta \in \Phi | [\theta_2, \ldots, \theta_p] \in Q\}.$$

Then

$$\lim_{\mu\to 0}\lim_{\delta\to 0}\lim_{k\to\infty} d_H(\mathcal{S}^{\rho^\mu,\pi^\delta,k}(E^{N,m}, K), \mathcal{S}(E^{N,m}, K)) = 0.$$

Proof. The result follows from Lemma 2.2.10 □

Attempting to find a globally best method for bounded error parameter validation on approximate models is a fruitless search. Determining which method is best suited for a particular problem requires a careful investigation of the system. For these reasons a direct comparison of the method we have introduced to existing methods is difficult. Therefore, we will attempt to give a clear description of the advantages and disadvantages of the method that we introduced.

The advantages of using our method is that it can be computationally efficient and it can be used to study how the membership set changes as the experiment changes. The computational efficiency stems first from the fact that we are solving differential equations at points in the parameter space rather than on intervals. Solving differential equations on intervals requires care and can lead to overly pessimistic results that can lead to an incorrect estimate of the membership set [22].

The disadvantages of our method come from the assumptions necessary to use the theory. Determining that the sensitivity of the differential equation with respect to θ_1 is non zero on the set Φ can be difficult. Similarly proving that the model is onto the intervals in the experiment can be daunting. However, these requirements are no more difficult to verify than the requirements necessary to use the asymptotic theory (A1-A9, B1-B9) for parameter validation.

Algorithm

1. Partition the set $\{[\theta_2, \ldots, \theta_p] | \theta \in \Phi\}$. This partition is denoted π^δ.

2. Partition the set $[0, T]$. This partition is denoted ρ^μ.

3. Apply Newton's method to the function $F_1^{i,j}(\theta_1) = w_i^{\rho^\mu}(t_j, [\theta_1, r^T]^T) - a_j^i$. Let $N_{1,k}^{i,j}(r)$ denote the result of k^{th} iteration of Newton's method.

4. Apply Newton's method to the function $F_2^{i,j}(\theta_1) = w_i^{\rho^\mu}(t_j, [\theta_1, r^T]^T) - b_j^i$. Let $N_{2,k}^{i,j}(r)$ denote the result of k^{th} iteration of Newton's method.

5. For $r \in \pi^\delta$ check if $max_{i,j}\{N_{x,k}^{i,j}(r)\} < min_{i,j}\{N_{c,k}^{i,j}(r)\}$ where $x = a$ and $c = b$ if $\partial w_i^{\rho^\mu}(t_j, \theta)/\partial\theta_1 > 0$ for $\theta \in \Phi$ and $x = b$ and $c = a$ if $\partial w_i^{\rho^\mu}(t_j, \theta)/\partial\theta_1 < 0$ for $\theta \in \Phi$. Call the set of $r \in \pi^\delta$ the satisfy this requirement V.

6. For each $v \in V$ define

$$N_{x,k}^{i_v,j_v}(v) = max_{i,j}\{N_{x,k}^{i,j}(v)\}$$
$$N_{c,k}^{i_v,j_v}(v) = min_{i,j}\{N_{c,k}^{i,j}(v)\}.$$

   Here $x = a$ and $c = b$ if $\partial w_i^{\rho^\mu}(t_j, \theta)/\partial\theta_1 > 0$ for $\theta \in \Phi$ and $x = b$ and $c = a$ if $\partial w_i^{\rho^\mu}(t_j, \theta)/\partial\theta_1 < 0$ for $\theta \in \Phi$.

7. Apply linear interpolation to the points $(v, N_{x,k}^{i_v,j_v}(v))$ to define the function $l_{E^{N,m}}^{\rho^\mu,\pi^\delta,k}(v)$ and then to the points $(v, N_{c,k}^{i_v,j_v}(v))$ to define the function $u_{E^{N,m}}^{\rho^\mu,\pi^\delta,k}(v)$ for $v \in V$ where $x = a$ and $c = b$ if $\partial w_i^{\rho^\mu}(t_j, \theta)/\partial\theta_1 > 0$ for $\theta \in \Phi$ and $x = b$ and $c = a$ if $\partial w_i^{\rho^\mu}(t_j, \theta)/\partial\theta_1 < 0$ for $\theta \in \Phi$.

8. Define $\mathcal{S}^{\rho^{\mu},\pi^{\delta},k}(E^{N,m},\Phi) = \{[q, v^T] | v \in V \text{ and } q \in [l_{E^{N,m}}^{\rho^{\mu},\pi^{\delta},k}(v), u_{E^{N,m}}^{\rho^{\mu},\pi^{\delta},k}(v)]\}$.

Chapter 3

The Design of Experiments

The quality of the design of an experiment can be the difference between an accurate parameter estimate that can be validated and an uninformative experiment. We define an experiment to a set of points, $\xi = \{t_j\}_{j=1}^{N}$, in the interval $[0, T]$. The points represent times where data will be collected in a scientific experiment. The theory for designing a scientific experiment requires that some information about the behavior of the system is known before experimentation. This information can come from past experiments or insight from studying the system. Since the true model parameter is not known, one can not design an experiment based on that parameter. However, one can design an experiment for an estimate of the true parameter. If the parameter estimate is a good approximation of the true parameter, then the design based on the parameter estimate is often times a good design for the true parameter.

Our goal in this chapter is to identify a set of points in $[0, T]$, from which we can make the most conclusive statement about the quality of a parameter estimate calculated from data collected at those times. The metric we use to measure the quality of the experiment depends on the assumptions we make about the random variable that represents error in the statistical model and the number of data points that can be collected.

3.1 The Design of Experiments Using Sensitivity Analysis

Every aspect of designing experiments uses sensitivity analysis in some way. Throughout this chapter there will be evidence of this and we will begin our discussion of the design of experiments by using traditional sensitivity functions and generalized sensitivity functions. Sensitivity functions measure, as the name states, how sensitive the model outputs to changes to the inputs. Alternatively, generalized sensitivity functions measure how sensitive a parameter estimate is. Intuitively the areas in the time domain where the model or parameter estimate is sensitive, are areas that contain a lot of information, so data collected there has a high information content.

3.1.1 Traditional Sensitivity Functions

Traditional sensitivity functions measure how changes in the parameter affect the model. We will say that the mathematical model is sensitive with respect to θ_1, if small changes in θ_1 translate into large changes in the mathematical model.

Definition 3.1.1. *The traditional sensitivity functions are defined to be*

$$s_{\theta_1}(t,\theta) \equiv [\frac{\partial y_1(t,\theta)}{\partial \theta_1} \; \frac{\partial y_2(t,\theta)}{\partial \theta_1} \; \dots \; \frac{\partial y_m(t,\theta)}{\partial \theta_1}]^T \tag{3.1.1}$$

$$s_{\theta_2}(t,\theta) \equiv [\frac{\partial y_1(t,\theta)}{\partial \theta_2} \; \frac{\partial y_2(t,\theta)}{\partial \theta_2} \; \dots \; \frac{\partial y_m(t,\theta)}{\partial \theta_2}]^T \tag{3.1.2}$$

$$\vdots \tag{3.1.3}$$

$$s_{\theta_p}(t,\theta) \equiv [\frac{\partial y_1(t,\theta)}{\partial \theta_p} \; \frac{\partial y_2(t,\theta)}{\partial \theta_p} \; \dots \; \frac{\partial y_m(t,\theta)}{\partial \theta_p}]^T. \tag{3.1.4}$$

To use traditional sensitivity functions, one would collect data at times when the model is sensitive to the parameter. For example, to gain information on the first component of θ, one would collect data at times when $\|s_{\theta_1}(t,\theta)\|$ is large, where $\|\cdot\|$ is an appropriate norm. Collecting information at these times, intuitively gives more information about θ_1 than at times when $\|s_{\theta_1}(t,\theta)\|$ is small. To estimate the parameter θ, one can select times when the sensitivity is high for each component of the parameter ([6], [7]). This method allows the researcher to concentrate on particular components of the unknown parameter by increasing the number of samples that will give additional information about a certain component.

3.1.2 Generalized Sensitivity Functions

Another way one can identify sampling regions of high information content is to use generalized sensitivity functions (GSF). Thomeseth and Cobelli developed the theory as a parameter identification tool for physiological problems in 1999 [41]. GSF measure how changes in data affect changes in the least squares parameter estimate. This information can be used to design meaningful experiments in a variety of modeling problems. Recently a research group lead by H. T. Banks has investigated GSF from a more mathematical perspective ([6], [7]). They are trying to answer the question, given N samples, where should the next M samples be taken to get the most information about the parameter. To accomplish this, they sample in intervals that the GSF identify as times of high information content. They found that sampling during the intervals of high information content, as opposed to even sampling throughout the entire time interval, results in smaller standard errors. Note, the areas of high information content are not well defined in ([6], [7]) and we will address this later in the chapter.

Thomaseth and Cobelli derived GSF for scalar statistical models as defined by (1.2.1), where the $\mathcal{E}_j$'s are identically independently distributed, have expected value 0, and variance $\sigma^2(t_j)$. These conditions are slightly more relaxed than the conditions necessary for the asymptotic theory for OLS since the variance does not have to be constant as a function of time.

Definition 3.1.2. *Generalized sensitivity functions have the form*

$$s_g(t_k) = \sum_{i=1}^{k} \frac{1}{\sigma^2(t_i)} \left[\sum_{j=1}^{N} \frac{1}{\sigma^2(t_j)} \nabla_\theta y(t_j,\theta) \nabla_\theta y(t_j,\theta)^T \right]^{-1} \nabla_\theta y(t_i,\theta) \bullet \nabla_\theta y(t_i,\theta), \tag{3.1.5}$$

where "$\bullet$" is element by element vector multiplication.

Generalized sensitivity functions approximate how much the least squares estimated parameter changes as the true parameter changes. Intervals when there is large change in GSF correspond to high information content for the estimated parameter. The theory says that increased sampling in these intervals will lead to a more informative sampling strategy than measuring in other regions. They are vector functions that have the same dimension as the unknown parameter. GSF take on the value 0 at $t = 0$ and 1 at $t = T$. Since GSF are always 1 at time T, they display a forced to one characteristic that can lead to a large rate of change around time T, when in reality there is little information there. Examining the GSF with a large T can help address this problem. If the parameters are uncorrelated then GSF increase monotonically; however, if they are correlated then they can be non monotone on the interval $[0, T]$. A measure of how much they increase and decrease can be seen using the function

$$s_{g_{inc}}(t_k) = \frac{1}{\sigma^2(t_k)} \left[\sum_{j=1}^{n} \frac{1}{\sigma^2(t_j)} \nabla_\theta y(t_j,\theta) \nabla_\theta y(t_j,\theta)^T \right]^{-1} \nabla_\theta y(t_k,\theta) \bullet \nabla_\theta y(t_k,\theta). \tag{3.1.6}$$

This function, $s_{g_{inc}}$, measures the change between two consecutive times. Large values of $s_{g_{inc}}$ relate to areas of high information content about the parameter.

For multi-dimensional models GSF have a form similar to the one dimensional case except they take into account changes from all of the components of the vector model output.

Definition 3.1.3. *Generalized sensitivity functions for vector models are defined by*

$$g_s(t_k) = \sum_{i=1}^{k} \sum_{l=1}^{n} \left\{ \left(\left[\sum_{j=1}^{N} \sum_{l=1}^{n} \frac{1}{\sigma_l^2(t_j)} \nabla_\theta y_l(t_j,\theta) \nabla_\theta y_l(t_j,\theta)^T \right]^{-1} \frac{\nabla_\theta y_l(t_i,\theta)}{\sigma_l^2(t_i)} \right) \bullet \nabla_\theta y_l(t_i,\theta) \right\}. \tag{3.1.7}$$

They have associated functions

$$gs_{inc}(t_k) = \sum_{l=1}^{n} \left\{ \left(\left[\sum_{j=1}^{N} \sum_{l=1}^{n} \frac{1}{\sigma_l^2(t_j)} \nabla_\theta y_l(t_j,\theta) \nabla_\theta y_l(t_j,\theta)^T \right]^{-1} \frac{\nabla_\theta y_l(t_k,\theta)}{\sigma_l^2(t_k)} \right) \bullet \nabla_\theta y_l(t_k,\theta) \right\}, \tag{3.1.8}$$

that measure the change in the GSF for vector models.

Continuous Generalized Sensitivity Functions

A clear generalization of the GSF defined by Thomaseth and Cobelli is to define them continuously. The group lead by H. T. Banks has made this generalization rigorous [9]. Consider the mathematical model

$$\dot{x}(t,\theta) = g(x(t,\theta),\theta) \tag{3.1.9}$$

$$y(t,\theta) = \mathcal{L}x(t,\theta) \tag{3.1.10}$$

where $t \in [0,T]$, $\theta \in \Phi$, $g : [0,T] \times \Phi \to \mathbb{R}^n$, and $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator. From the deterministic model the statistical model is given by

$$D(t) = y(t,\theta_0) + \mathcal{E}(t) \tag{3.1.11}$$

where $y(t,\theta_0)$ is the deterministic output of the model and $\mathcal{E}(t)$ is a random variable. For each $t \in [0,T]$, one assumes that $E[\mathcal{E}(t)] = 0$ and variance has $\sigma^2(t)$. Further one assumes that for any $t \neq s$, $\mathcal{E}(t)$ and $\mathcal{E}(s)$ are independent with

$$Cov(\mathcal{E}(s)\mathcal{E}(t)) = \sigma(t)\sigma(s)\delta(t-s) \tag{3.1.12}$$

for all $t, s \in [0,T]$, where $\delta(\cdot)$ is the dirac delta function.

The weighted cost function to estimate θ_0 is given by

$$J(y,\theta) = \int_0^t \frac{(d(t) - y(t,\theta))^2}{\sigma^2(t)} dm(t), \tag{3.1.13}$$

where m is a probability measure defined on $([0,T],\beta)$.

Definition 3.1.4. *A continuous generalized sensitivity function (CGSF) is defined by*

$$s_{cg}(t) = \int_0^t \left[F_G(T)^{-1} \frac{1}{\sigma^2(s)} \nabla_\theta y(s,\theta_0) \right] \bullet \nabla_\theta y(s,\theta_0) dm(s) \tag{3.1.14}$$

where

$$F_G(T) = \int_0^T \frac{1}{\sigma^2(t)} \nabla_\theta y(t,\theta) \nabla_\theta y(t,\theta)^T dm(t) \tag{3.1.15}$$

is the infinite dimensional Fisher information matrix and m *is a probability measure defined on* $([0,T],\beta)$.

Suppose that m is the Lebesgue measure, then the time derivative of the CGSF is given by

$$\frac{\partial s_{cg}(t)}{\partial t} = \left[F_G^{-1}(T)\frac{1}{\sigma^2(t)}\nabla_\theta y(t,\theta_0)\right] \bullet \nabla_\theta y(t,\theta_0). \tag{3.1.16}$$

This function is analogous to gs_{inc} in the discrete version.

Interpreting Generalized Sensitivity Functions

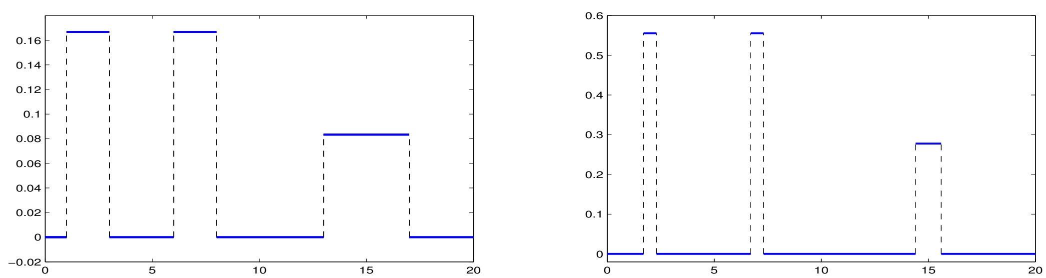
The amount of confidence that the experimenter has in the estimate $\hat{\theta}$ can be used when interpreting the GSF. In the literature, the areas identified to be good places to sample are found by finding regions of sharp increase in the GSF and then sampling uniformly in those regions. We can be more precise with what sharp increase means and sample more aggressively when we are confident with our parameter estimate. Our sampling strategy does not have to be uniform and is most likely not the best way to collect additional data for good parameter estimates.

The goal of this section is to define a sampling strategy using information from the GSF or CGSF. Suppose given an estimate $\hat{\theta}$, that $\hat{t}_j \in [0,T]$ are identified as points where the GSF is sensitive. The next step is to translate this information into a time sample $\xi_N = \{t_j\}_{j=1}^N$ that combines the information about the quality of the estimate $\hat{\theta}$ and the information from the GSF. We will revert to the notation defined in Chapter 2 where we associated time sample with measures. Let us consider measure spaces of the form $([0,T],\beta,m)$. We are trying to answer the question, where should we place N samples? The information from the GSF suggests that sampling close to $\{\hat{t}_j\}$ would be beneficial. We want to be more precise about what we mean by sampling close to $\{\hat{t}_j\}$. To define close will define the parameters $c_{j,1}$ and $c_{j,2}$. The measure we are looking for will have the property that

$$\sum_{j=1}\int_{\hat{t}_j-c_{j,1}}^{\hat{t}_j+c_{j,2}} 1dm(t) \tag{3.1.17}$$

will be close to 1. This says that the majority of the sampling will be in the intervals $[\hat{t}_j - c_{j,1}, \hat{t}_j + c_{j,2}]$. From these infinite-dimensional measures we can define finite dimensional measures, m_N on $([0,T]),\beta)$ that approximate them.

For example, suppose that $\{\hat{t}_j\}_{j=1}^3 = \{2,7,15\}$ represents information gathered from GSF and $([0,T],\beta,m_1)$ and $([0,T],\beta,m_2)$ describe two sampling distributions. Figure 3.1 shows two different ways to use the same information gathered from GSF. The distribution in Figure 3.1 (a) demonstrates less confidence in the parameter estimate than 3.1 (b) because the samples, as whole, are taken close to the points of $\{\hat{t}_j\}_{j=1}^3$.



(a) pdf of sampling distribution given $\hat{\theta}$ with lower confidence in the estimate

(b) pdf of sampling distribution given $\hat{\theta}$ with higher confidence in the estimate

Figure 3.1: Sampling Strategies for GSF

3.2 The Optimal Design of Experiments

We will examine the optimal design of experiments problem for a scalar model. There are several different designs that are each optimal for some aspect of the estimated parameter. The four types of optimal designs that we will look at are A, D, E, and V optimal designs. These design are defined and used in ([33],[40]).

D-optimal designs select sampling points out of a candidate set by finding the points that maximize the determinant of the Fisher information matrix, $\chi^T(\theta,\xi_N)\chi(\theta,\xi_N)$. For linear models finding the D-optimal design is equivalent to minimizing the asymptotic confidence region for the OLS parameter θ. These two procedures are equivalent because $\chi^T(\theta,\xi_N)\chi(\theta,\xi_N)^{-1/2}$ is proportional to the volume of the asymptotic confidence region [39]. For nonlinear models, D-optimal designs minimize the volume of the confidence ellipsoid for a linearization for the nonlinear model [12].

Definition 3.2.1. *An experiment* ξ_N^* *is D-optimal over a candidate set* $\mathcal{T}$ *for* θ *if*

$$\xi_N^* = arg \max_{\{\xi_N \in \mathcal{T}\}} det(\chi^T(\theta, \xi_N)\chi(\theta, \xi_N))$$

where $\chi(\theta, \xi_N)$ *is defined by Definition 2.2.1.*

An A-optimal design minimizes the product of the square of the standard errors for the OLS parameter estimator.

Definition 3.2.2. *An experiment* ξ_N^* *is A-optimal over a candidate set* $\mathcal{T}$ *for* θ *if*

$$\xi_N^* = arg \max_{\{\xi_N \in \mathcal{T}\}} trace([\chi^T(\theta, \xi_N)\chi(\theta, \xi_N)]^{-1})$$

where $\chi(\theta, \xi_N)$ *is defined by Definition 2.2.1.*

An E-optimal design minimizes the length of the largest eigenvalue of the N^{th} dimensional approximation of the asymptotic covariance of the OLS parameter estimator.

Definition 3.2.3. *An experiment* ξ_N^* *is E-optimal over a candidate set* $\mathcal{T}$ *for* θ *if*

$$\xi_N^* = arg \max_{\{\xi_n \in \mathcal{T}\}} \|eig([\chi^T(\theta, \xi_N)\chi(\theta, \xi_N)]^{-1})\|_\infty$$

where $\chi(\theta, \xi_N)$ *is defined by Definition 2.2.1.*

All of these designs, D, A, and E, give insight to the design of experiment problem but it is important to note how the designs are optimal. They are optimal in some sense for an **estimate** of the true parameter. Therefore, their results need to be taken with a grain of salt. As we will see in Section 3.2.1 the optimal designs are not necessarily practical. One way to make the design more robust is to find a design that is optimal for a set of parameters instead of just the estimated parameter. These designs are found by averaging optimal designs for a set of parameters.

Definition 3.2.4. *Given a set* $K \subset \Phi$ *and measure space* (K, β, m)*, a design* $\bar{\xi}^*$ *is D-Optimal with respect to* (K, β, m) *over a candidate set* $\mathcal{T}$ *if*

$$\bar{\xi}_n^* = arg \max_{\{\xi_n \in \mathcal{T}\}} \int_K det(\chi^T(\theta, \xi_N)\chi(\theta, \xi_N))dm$$

where $\chi^T(\theta, \xi_N)$ *is defined by Definition 2.2.1.*

A design of this type will not be optimal for the individual parameter estimate but it will take information about the parameters close to the estimated parameter [33]. King and Wong expanded this idea further by searching for minimax D-optimal designs [28].

Each of these definitions can be extended to vector models, however, the interpretations are not the same [19].

3.2.1 Finding a $k * p$ Point D-Optimal Design

In general finding a D-optimal design when N is large is a very complicated problem to solve. It is easy to see that for moderately large N, solving the optimization problem is extremely costly. The difficulty increases even more when restrictions on the relationships between the samples, such as minimum intervals between samples, are imposed. G.E.P. Box and Lucas [12] were the first to use D-optimal designs for non-linear models and began their work for the case when the number of sample points is the same as the number of parameters. They solved the problem geometrically by investigating the design space. Lucas expanded on this idea [29] for the case $N > p$, the number of sample points N is greater than the dimension of the parameter p, and found that repeated samples at the optimal points for the case $N = p$ were optimal or near optimal for many models. Atkinson and Hunter further expanded these ideas [3] and were able to prove a necessary condition for when a $N = p$ point design is D-optimal and a sufficient condition for the case where repeated samples at the D-optimal points for the case $N = p$ are optimal for $n * p$ points.

Definition 3.2.5. *The attainable region is*

$$R(\theta) = \{\nabla_\theta y(t, \theta) | t \in [0, T]\}. \tag{3.2.1}$$

Let ξ_p^* be a p-point D-optimal design for the model, and define $\Delta_p = |\chi^T(\theta, \xi_p^*)\chi(\theta, \xi_p^*)|$. Now suppose that the u^{th} row of $\chi(\theta, \xi^*)$ is replaced with another row. Denote the determinate of that matrix by $\delta_{p+1,-u}$. We now introduce the necessary condition that for a set of $N = p$ point to be D-optimal found in [3].

Theorem 3.2.1. *A necessary condition for the D-optimality of a p-point design is that it shall consist of points of $R(\theta)$ such that $R(\theta)$ does not lie outside E, the p-dimensional parallelepiped defined by the p pairs of planes*

$$\delta^2_{p+1,-u} = \Delta_p \ u = 1, 2, ..., p.$$

Similarly we introduce the sufficient condition that repeated samples at the D-optimal points for the case $N = p$ are D-optimal for $k * p$ points found in [3].

Theorem 3.2.2. *A sufficient condition for the D-optimal design for $N = np$ experiments, $n = 2, 3, 4, ...,$ to consist solely of replications of p points which are D-optimal for $N = p$ is that $R(\theta)$ be contained in the p-ellipsoid E, the locus of points $(p + 1)$ satisfying the relationship*

$$\sum_{u=1}^{p} \delta^2_{p+1,-u} = \Delta_p.$$

While these theorems give a method for checking models where the exact solution to the differential equation that describes the model is known, they can not be directly applied to problems where we have to approximate the model. However, under certain conditions we can use the results of the theorems on models that we have to approximate. As we did in Chapter 2, suppose that $\rho^\mu = [p_1, \ldots, p_M]$ is a μ-fine partition of $[0, T]$. Further suppose that $[w_1^{\rho^\mu}(p_1, \theta), \ldots, w_M^{\rho^\mu}(p_M, \theta)]$ is the approximation of $[y(p_1, \theta), \ldots, y(p_M, \theta)]$. Define $w^{\rho^\mu} : [0, T], \Phi \to R^m$ to be the linear interpolation of the points $\{(p_j, w_j^{\rho^\mu}(p_j, \theta))\}$. Let $\xi^*_{\rho^\mu}$ be the D-optimal p-point design for the approximate model.

Suppose that $\xi^*_{\rho^\mu,l}$ is a numerical approximation of $\xi^*_{\rho^\mu}$ such that

$$\lim_{l \to \infty} \|\xi^*_{\rho^\mu} - \xi^*_{\rho^\mu,l}\| = 0. \tag{3.2.2}$$

Define $\Delta_p^{\rho^\mu,l} = |\chi^T_{\rho^\mu}(\theta, \xi^*_{\rho^\mu,l})\chi_{\rho^\mu}(\theta, \xi^*_{\rho^\mu,l})|$. where

$$\chi_{\rho^\mu}(\theta, \xi_N)_{jk} = s_k^{\rho^\mu}, \ j = 1, ..., N, \ k = 1, ..., p$$

From this approximation of the Fisher information matrix define the approximate attainable region.

Definition 3.2.6. *The ρ^μ-approximate attainable region is*

$$R^{\rho^\mu}(\theta) = \{s^{\rho^\mu}(t, \theta) | t \in [0, T]\}. \tag{3.2.3}$$

Now suppose that the u^{th} row of $\chi_{\rho^\mu}(\theta, \xi^*_{\rho^\mu,l})$ is replaced with another row. Denote the determinate of that matrix $\delta^{\rho^\mu,l}_{p+1,-u}$. Then for the approximate model we have derived a similar necessary condition for a set of $N = p$ samples to be D-optimal and a sufficient condition for the case when repeated samples at the D-optimal points for the case $N = p$ are D-optimal for $k * p$ points. Before we state the theorems we define two different ways to measure the distance between sets in $\mathbb{R}^p$ and list some of their properties.

Definition 3.2.7. *For a fixed set* A *define the set function*

$$d^A(X) = \sup_{x \in X} \inf_{a \in A} d(x, a) \tag{3.2.4}$$

In our application of the distance functions we will use $d(x, y) = \|x - y\|_p$.

Lemma 3.2.1. *Given compact sets* $A, X, Y \subset \mathbb{R}^n$*, if* $d_H(X, Y) < \varepsilon$ *then* $d^A(X) \leq d^A(Y) + \varepsilon$.

Lemma 3.2.2. *Given compact sets* $A, B, X \subset \mathbb{R}^n$*, if* $d_H(A, B) < \varepsilon$ *then* $d^A(X) \leq d^B(X) + \varepsilon$.

We now introduce a new theorem that describes a necessary condition for a p-point experiment to be D-optimal.

Theorem 3.2.3. *Let* $\varepsilon > 0$*,* $\{\rho^\mu\}$ *be a set of partition with* $\mu \to 0$ *and suppose for a given numerical scheme* $\lim_{\mu\to 0} \|s(t, \theta) - s^{\rho^\mu}(t, \theta)\|_\infty = 0$ *and* $\lim_{\mu\to 0} \|y(t, \theta) - w^{\rho^\mu}(t, \theta)\|_\infty = 0$*. A necessary condition for the optimality of a p-point design of the model* $y(t, \theta)$ *is that there exists a* ρ^μ *and an approximately D-optimal design* $\xi^*_{\rho^\mu,l}$ *for* $w^{\rho^\mu}(t, \theta)$*, consisting of points of* $R^{\rho^\mu}(\theta)$ *such that* $d^{P^{\rho^\mu,l}}(R^{\rho^\mu}(\theta)) < \varepsilon$ *where* $P^{\rho^\mu,l}$ *is the p-dimensional parallelepiped defined by the p pairs of planes*

$$[\delta^{\rho^\mu,l}_{p+1,-u}]^2 = \Delta^{\rho^\mu,l}_p \;\; u = 1, 2, ..., p.$$

Then the $\lim_{\mu\to 0} \lim_{l\to 0} \xi^*_{\rho^\mu,l} = \xi^*$ *is the D-optimal p-point design for* $y(t, \theta)$*.*

Proof. Define a function $G_1 : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ by

$$G_1(w,\theta,\xi) = \begin{vmatrix} w_1 & w_2 & \cdots & w_p \\ s_{\theta_1}(t_2,\theta) & s_{\theta_2}(t_2,\theta) & \cdots & s_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}(t_p,\theta) & s_{\theta_2}(t_p,\theta) & \cdots & s_{\theta_p}(t_p,\theta) \end{vmatrix}^2 \tag{3.2.5}$$

$$- \begin{vmatrix} s_{\theta_1}(t_1,\theta) & s_{\theta_2}(t_1,\theta) & \cdots & s_{\theta_p}(t_1,\theta) \\ s_{\theta_1}(t_2,\theta) & s_{\theta_2}(t_2,\theta) & \cdots & s_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}(t_p,\theta) & s_{\theta_2}(t_p,\theta) & \cdots & s_{\theta_p}(t_p,\theta) \end{vmatrix}^2 \tag{3.2.6}$$

and the functions $G_j : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ for $j = 2, 3, \ldots, p$ by

$$G_j(w,\theta,\xi) = \begin{vmatrix} s_{\theta_1}(t_1,\theta) & s_{\theta_2}(t_1,\theta) & \cdots & s_{\theta_p}(t_1,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}(t_{j-1},\theta) & s_{\theta_2}(t_{j-1},\theta) & \cdots & s_{\theta_p}(t_{j-1},\theta) \\ w_1 & w_2 & \cdots & w_p \\ s_{\theta_1}(t_{j+1},\theta) & s_{\theta_2}(t_{j+1},\theta) & \cdots & s_{\theta_p}(t_{j+1},\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}(t_p,\theta) & s_{\theta_2}(t_p,\theta) & \cdots & s_{\theta_p}(t_p,\theta) \end{vmatrix}^2 \tag{3.2.7}$$

$$- \begin{vmatrix} s_{\theta_1}(t_1,\theta) & s_{\theta_2}(t_1,\theta) & \cdots & s_{\theta_p}(t_1,\theta) \\ s_{\theta_1}(t_2,\theta) & s_{\theta_2}(t_2,\theta) & \cdots & s_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}(t_p,\theta) & s_{\theta_2}(t_p,\theta) & \cdots & s_{\theta_p}(t_p,\theta) \end{vmatrix}^2 . \tag{3.2.8}$$

Then G_j, is quadratic and for fixed $\theta \in \Phi$, $\xi \in D$ the solution to $G_j(w,\theta,\xi) = 0$ is a p-dimensional set of parallel planes [3]. We can rewrite the relation

$$[\delta_{p+1,-u}]^2 = \Delta_p \qquad u = 1, 2, \ldots, p,$$

as $G_j(w,\theta,\xi^*) = 0$ $j = 1, 2, \ldots p$. Define the w that satisfy $G_j(w,\theta,\xi^*) = 0$, $j = 1, 2, \ldots p$ as $\bar{P}$ and define P to be the set of w that satisfy $G_j(w,\theta,\xi^*) \leq 0$ $j = 1, 2, \ldots p$. Then we can similarly define a function $G^{\rho^\mu} : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ by

$$G_1^{\rho^\mu}(w,\theta,\xi) = \begin{vmatrix} w_1 & w_2 & \cdots & w_p \\ s_{\theta_1}^{\rho^\mu}(t_2,\theta) & s_{\theta_2}^{\rho^\mu}(t_2,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}^{\rho^\mu}(t_p,\theta) & s_{\theta_2}^{\rho^\mu}(t_p,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_p,\theta) \end{vmatrix}^2 - \begin{vmatrix} s_{\theta_1}^{\rho^\mu}(t_1,\theta) & s_{\theta_2}^{\rho^\mu}(t_1,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_1,\theta) \\ s_{\theta_1}^{\rho^\mu}(t_2,\theta) & s_{\theta_2}^{\rho^\mu}(t_2,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}^{\rho^\mu}(t_p,\theta) & s_{\theta_2}^{\rho^\mu}(t_p,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_p,\theta) \end{vmatrix}^2$$

and the functions $G_j^{\rho^\mu} : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ for $j = 2, 3, \ldots, p$ by

$$G_j^{\rho^\mu}(w,\theta,\xi) = \begin{vmatrix} s_{\theta_1}^{\rho^\mu}(t_1,\theta) & s_{\theta_2}^{\rho^\mu}(t_1,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_1,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}^{\rho^\mu}(t_{j-1},\theta) & s_{\theta_2}^{\rho^\mu}(t_{j-1},\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_{j-1},\theta) \\ w_1 & w_2 & \cdots & w_p \\ s_{\theta_1}^{\rho^\mu}(t_{j+1},\theta) & s_{\theta_2}^{\rho^\mu}(t_{j+1},\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_{j+1},\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}^{\rho^\mu}(t_p,\theta) & s_{\theta_2}^{\rho^\mu}(t_p,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_p,\theta) \end{vmatrix}^2 - \begin{vmatrix} s_{\theta_1}^{\rho^\mu}(t_1,\theta) & s_{\theta_2}^{\rho^\mu}(t_1,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_1,\theta) \\ s_{\theta_1}^{\rho^\mu}(t_2,\theta) & s_{\theta_2}^{\rho^\mu}(t_2,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}^{\rho^\mu}(t_p,\theta) & s_{\theta_2}^{\rho^\mu}(t_p,\theta) & \cdots & s_{\theta_p}^{\rho^\mu}(t_p,\theta) \end{vmatrix}^2 .$$

Then, $G_j^{\rho^\mu}$, $j = 1, \ldots, p$, is quadratic and for fixed $\theta \in \Phi$, $\xi \in D$ the solution to $G_j^{\rho^\mu}(w,\theta,\xi) = 0$ is a set of p-dimensional parallel planes [3]. We can rewrite the relation

$$[\delta_{p+1,-u}^{\rho^\mu,l}]^2 = \Delta_p^{\rho^\mu,l} \qquad u = 1, 2, \ldots, p,$$

as $G_j^{\rho^\mu}(w,\theta,\xi_{\rho^\mu,l}^*) = 0$, $j = 1, 2, \ldots, p$. Let the set of w that satisfy $G_j^{\rho^\mu}(w,\theta,\xi_{\rho^\mu,l}^*) = 0$, $j = 1, 2, \ldots, p$ be $\bar{P}^{\rho^\mu,l}$ and $P^{\rho^\mu,l}$ be the set that is defined by $G_j^{\rho^\mu}(w,\theta,\xi_{\rho^\mu,l}^*) \leq 0$ $j = 1, 2, \ldots, p$. Similarly let P^{ρ^μ} be the set that satisfies $G_j^{\rho^\mu}(w,\theta,\xi_{\rho^\mu}^*) \leq 0$ $j = 1, 2, \ldots, p$.

Since $\lim_{\mu \to 0} \|s(t,\theta) - s^{\rho^\mu}(t,\theta)\|_\infty = 0$, there exists a partition ρ^μ such that

$$d_H(R^{\rho^\mu}(\theta), R(\theta)) < \varepsilon/3 \tag{3.2.9}$$

and

$$d_H(P, P^{\rho^\mu}) < \varepsilon/3 \tag{3.2.10}$$

Define the function $F^{\rho^\mu}(\theta,\cdot) : [0,T]^p \to \mathbb{R}^1$ by

$$F^{\rho^\mu}(\theta,\xi) = |\chi^T_{\rho^\mu}(\theta,\xi)\chi_{\rho^\mu}(\theta,\xi)| \tag{3.2.11}$$

Since $F^{\rho^\mu}(\theta,\cdot)$ is continuous and the domain is compact it has a maximum on $[0,T]^p$. Let $\{\xi^*_{\rho^\mu,l}\}$ be a sequence that converges uniformly to the point $\xi^*_{\rho^\mu} \subset [0,T]^p$ that maximizes $F^{\rho^\mu}(\theta,\cdot)$.

Then for this ρ^μ there exists an l such that

$$d_H(P^{\rho^\mu}, P^{\rho^\mu,l}) < \varepsilon/3 \tag{3.2.12}$$

Then using Equations 3.2.9, 3.2.10, and 3.2.12 and Lemmas 3.2.1 and 3.2.2 we have

$$\begin{aligned}
d^P(R(\theta)) &\leq d^{P^{\rho^\mu}}(R(\theta)) + d_H(P, P^{\rho^\mu}) && (3.2.13)\\
&\leq d^{P^{\rho^\mu,l}}(R(\theta)) + d_H(P^{\rho^\mu}, P^{\rho^\mu,l}) + \varepsilon/3 && (3.2.14)\\
&\leq d^{P^{\rho^\mu,l}}(R^{\rho^\mu}(\theta)) + d_H(R(\theta), R^{\rho^\mu}(\theta)) + \varepsilon/3 + \varepsilon/3 && (3.2.15)\\
&\leq 2\varepsilon && (3.2.16)
\end{aligned}$$

□

We now state a theorem that describes a sufficient condition for a n*p-point D-optimal experiment to be n repeated samples at the p-point D-optimal design.

Theorem 3.2.4. *Let* $\varepsilon > 0$, $\{\rho^\mu\}$ *be a set of partition with* $\mu \to 0$ *and suppose for a given numerical scheme* $\lim_{\mu\to 0} \|s(t,\theta) - s^{\rho^\mu}(t,\theta)\|_\infty = 0$ *and* $\lim_{\mu\to 0} \|y(t,\theta) - w^{\rho^\mu}(t,\theta)\|_\infty = 0$. *A sufficient condition for the D-optimal design for* $N = np$ *experiments,* $n = 2, 3, 4, \ldots,$ *to* *consist solely to replications of p points which are D-optimal when $N = p$ for the model $y(t,\theta)$, is that there exists an ρ^μ and an approximately D-optimal design $\xi^*_{\rho^\mu,l}$, for $w^{\rho^\mu}(t,\theta)$ such that $d^{E^{\rho^\mu,l}}(R^{\rho^\mu}(\theta)) < \varepsilon$ where $E^{\rho^\mu,l}$ is the p-ellipsoid , the locus of points $(p+1)$ satisfying the relationship*

$$\sum_{u=1}^{p}[\delta^{\rho^\mu,l}_{p+1,-u}]^2 = \Delta^{\rho^\mu,l}_p.$$

Then n replicates of the design $\lim_{\mu\to 0}\lim_{l\to 0}\xi^*_{\rho^\mu,l} = \xi^*$ is the D-optimal np-point design for $y(t,\theta)$.

Proof. Define a function $T : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ by

$$\begin{aligned} T(w,\theta,\xi) = &\begin{vmatrix} w_1 & w_2 & \cdots & w_p \\ s_{\theta_1}(t_2,\theta) & s_{\theta_2}(t_2,\theta) & \cdots & s_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}(t_p,\theta) & s_{\theta_2}(t_p,\theta) & \cdots & s_{\theta_p}(t_p,\theta) \end{vmatrix}^2 \\ &+ \begin{vmatrix} s_{\theta_1}(t_1,\theta) & s_{\theta_2}(t_1,\theta) & \cdots & s_{\theta_p}(t_1,\theta) \\ w_1 & w_2 & \cdots & w_p \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}(t_p,\theta) & s_{\theta_2}(t_p,\theta) & \cdots & s_{\theta_p}(t_p,\theta) \end{vmatrix}^2 \\ &+ \ldots + \begin{vmatrix} s_{\theta_1}(t_1,\theta) & s_{\theta_2}(t_1,\theta) & \cdots & s_{\theta_p}(t_1,\theta) \\ s_{\theta_1}(t_2,\theta) & s_{\theta_2}(t_2,\theta) & \cdots & s_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \cdots & w_p \end{vmatrix}^2 \\ &- \begin{vmatrix} s_{\theta_1}(t_1,\theta) & s_{\theta_2}(t_1,\theta) & \cdots & s_{\theta_p}(t_1,\theta) \\ s_{\theta_1}(t_2,\theta) & s_{\theta_2}(t_2,\theta) & \cdots & s_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s_{\theta_1}(t_p,\theta) & s_{\theta_2}(t_p,\theta) & \cdots & s_{\theta_p}(t_p,\theta) \end{vmatrix}^2. \end{aligned}$$

Then, T is quadratic and for fixed $\theta \in \Phi$, $\xi \in D$ the solution to $T(w,\theta,\xi) = 0$ is a p-dimensional ellipsoid [3]. We can rewrite the relation

$$[\delta_{p+1,-u}]^2 = \Delta_p \qquad u = 1, 2, ..., p,$$

as $T(w,\theta,\xi^*) = 0$. Define the w that satisfy $T(w,\theta,\xi^*) = 0$ as $\bar{E}$ and define E to be the set of w that satisfy $T(w,\theta,\xi^*) \leq 0$. Then we can similarly define a function $T^{\rho^\mu} : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ by

$$
\begin{aligned}
T^{\rho^\mu}(w,\theta,\xi) = &\begin{vmatrix} w_1 & w_2 & \cdots & w_p \\ s^{\rho^\mu}_{\theta_1}(t_2,\theta) & s^{\rho^\mu}_{\theta_2}(t_2,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s^{\rho^\mu}_{\theta_1}(t_p,\theta) & s^{\rho^\mu}_{\theta_2}(t_p,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_p,\theta) \end{vmatrix}^2 \\
&+ \begin{vmatrix} s^{\rho^\mu}_{\theta_1}(t_1,\theta) & s^{\rho^\mu}_{\theta_2}(t_1,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_1,\theta) \\ w_1 & w_2 & \cdots & w_p \\ \vdots & \vdots & \ddots & \vdots \\ s^{\rho^\mu}_{\theta_1}(t_p,\theta) & s^{\rho^\mu}_{\theta_2}(t_p,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_p,\theta) \end{vmatrix}^2 \\
&+ \ldots + \begin{vmatrix} s^{\rho^\mu}_{\theta_1}(t_1,\theta) & s^{\rho^\mu}_{\theta_2}(t_1,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_1,\theta) \\ s^{\rho^\mu}_{\theta_1}(t_2,\theta) & s^{\rho^\mu}_{\theta_2}(t_2,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \cdots & w_p \end{vmatrix}^2 \\
&- \begin{vmatrix} s^{\rho^\mu}_{\theta_1}(t_1,\theta) & s^{\rho^\mu}_{\theta_2}(t_1,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_1,\theta) \\ s^{\rho^\mu}_{\theta_1}(t_2,\theta) & s^{\rho^\mu}_{\theta_2}(t_2,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_2,\theta) \\ \vdots & \vdots & \ddots & \vdots \\ s^{\rho^\mu}_{\theta_1}(t_p,\theta) & s^{\rho^\mu}_{\theta_2}(t_p,\theta) & \cdots & s^{\rho^\mu}_{\theta_p}(t_p,\theta) \end{vmatrix}^2.
\end{aligned}
$$

Then, T^{ρ^μ} is quadratic and for fixed $\theta \in \Phi$, $\xi \in D$ the solution to $T^{\rho^\mu}(w,\theta,\xi) = 0$ is a p-dimensional ellipsoid [3]. We can rewrite the relation

$$\sum_{u=1}^{p} [\delta^{\rho^\mu,l}_{p+1,-u}]^2 = \Delta^{\rho^\mu,l}_p,$$

as $T^{\rho^\mu}(w,\theta,\xi^*_{\rho^\mu,l}) = 0$. Let the set of w that satisfy $T^{\rho^\mu}(w,\theta,\xi^*_{\rho^\mu,l}) = 0$ be $\bar{E}^{\rho^\mu,l}$ and $E^{\rho^\mu,l}$ be the set that is defined by $T(w,\theta,\xi^*_{\rho^\mu,l}) \leq 0$. Similarly let E^{ρ^μ} be the set that is defined by $T(w,\theta,\xi^*_{\rho^\mu}) \leq 0$

Since $\lim_{\mu\to 0} \|s(t,\theta) - s^{\rho^\mu}(t,\theta)\|_\infty = 0$, there exists a partition ρ^μ such that

$$d_H(R^{\rho^\mu}(\theta), R(\theta)) < \varepsilon/3 \tag{3.2.17}$$

and

$$d_H(E, E^{\rho^\mu}) < \varepsilon/3. \tag{3.2.18}$$

Define the function $F^{\rho^\mu}(\theta,\cdot) : [0,T]^p \to \mathbb{R}^1$ by

$$F^{\rho^\mu}(\theta,\xi) = |\chi^T_{\rho^\mu}(\theta,\xi)\chi_{\rho^\mu}(\theta,\xi)|. \tag{3.2.19}$$

Since $F^{\rho^\mu}(\theta,\cdot)$ is continuous and the domain is compact it has a maximum on $[0,T]^p$. Let $\{\xi^*_{\rho^\mu,l}\}$ be a sequence that converges uniformly to the point $\xi^*_{\rho^\mu} \subset [0,T]^p$ that maximizes $F^{\rho^\mu}(\theta,\cdot)$.

Then for this ρ^μ there exists an l such that

$$d_H(E^{\rho^\mu}, E^{\rho^\mu,l}) < \varepsilon/3. \tag{3.2.20}$$

Then using Equations 3.2.17, 3.2.18, and 3.2.20 and Lemmas 3.2.1 and 3.2.2 we have

$$\begin{aligned}
d^E(R(\theta)) &\le d^{E^{\rho^\mu}}(R(\theta)) + d_H(E, E^{\rho^\mu}) && (3.2.21)\\
&\le d^{E^{\rho^\mu,l}}(R(\theta)) + d_H(E^{\rho^\mu}, E^{\rho^\mu,l}) + \varepsilon/3 && (3.2.22)\\
&\le d^{E^{\rho^\mu},l}(R^{\rho^\mu}(\theta)) + d_H(R(\theta), R^{\rho^\mu}(\theta)) + \varepsilon/3 + \varepsilon/3 && (3.2.23)\\
&\le 2\varepsilon. && (3.2.24)
\end{aligned}$$

□

A New Approach to D-Optimal Design

The problem we are trying to solve is, given an initial parameter estimate, design an experiment that gives the most information about the parameter. We want to be able to get a good estimate of the parameter. We measure the quality of the estimate by the size of the confidence ellipsoid and/or the standard errors associated with the estimate. The initial parameter estimate could come from a previous experiment or an educated guess of how the system behaves. When we design an experiment we will factor in how close we believe the true parameter is to the initial parameter estimate.

A clear application of the theory would be to use the OLS parameter estimate to design a p-point D-optimal experiment for the initial parameter estimate. If we can apply Theorem 3.2.2 or Theorem 3.2.4, then the answer would be to sample equally at the optimal points from a p-point design. Unfortunately, using this strategy is neither practical nor optimal in most cases. Additionally the D-optimal experiment for the initial parameter is not the same as the D-optimal experiment for OLS estimate. An inaccurate initial parameter could lead to a design that is far from optimal for the true parameter.

The first method we are proposing is to design an experiment that approximately minimizes the area of the asymptotic confidence ellipsoid while taking into account error in the initial parameter. The idea is to design an experiment using a modified p-point D-Optimal design that can be used for any number of additional samples taken. The designs we are considering are placing sample points in intervals centered at the D-optimal p-point designs. The size of the interval will be based on the perceived accuracy of the initial parameter estimate.

Suppose that $\xi_p^* = \{t_j\}_{j=1}^p$ is a D-optimal p-point design for the initial parameter estimate. Assuming that the $k * p$ point D-optimal design is a p-point design repeated k times, the D-optimal design would have the a sampling distribution described the measure m, defined on $([0,T],\beta)$ with $m([a,b]) = v(b) - v(a)$ where $\xi^* = \{t_j^*\}_{j=1}^p$ and

$$v(t) = \begin{cases} 0 & \text{if } t \in [0, t_1^*), \\ \frac{k}{p} & \text{if } t \in [t_j^*, t_{j+1}^*) \;\; j = 1, \dots, p-1, \\ 1 & \text{if } t \in [t_p^*, T]. \end{cases} \tag{3.2.25}$$

As stated before, sampling according to the distribution defined by (3.2.25) is most likely not optimal for the true parameter and can be unrealistic. Instead of using repeated sampling at the p-point D-optimal design, we will approximate it with a sampling strategy defined by the measure $m_{\vec{\Delta}t}$ defined on $([0,T],\beta)$ with $m_{\vec{\Delta}t}([a,b]) = v_{\vec{\Delta}t}(b) - v_{\vec{\Delta}t}(a)$ where $\xi^* = \{t_j^*\}_{j=1}^p$

and

$$v_{\vec{\Delta}t}(t) = \begin{cases} 0 & \text{if } t \in [0, t_1 + \Delta t_1], \\ (\frac{1}{p\Delta t_j})(t - (t_j - \Delta t_j)) + (j-1)\frac{1}{p} & \text{if } t \in [t_j - \Delta t_j, t_j + \Delta t_j] \;\; j = 1, \ldots, p, \\ \frac{j}{p} & \text{if } t \in [t_j + \Delta t_j, t_{j+1} - \Delta t_{j+1}] \;\; j = 1, \ldots, p-1, \\ 1 & \text{if } t \in [t_p + \Delta t_p, T]. \end{cases} \tag{3.2.26}$$

where $\vec{\Delta}t = [\Delta t_1, \ldots, \Delta t_p]^T$ and $t_1 - \Delta t_1 > 0$, $t_p + \Delta t_p < T$, and $\Delta t_j + \Delta t_{k+1} < t_{k+1} - t_k$ for $k = 1, \ldots, p-1$. This measure describes uniform sampling in each of the intervals $[t_j - \Delta t_j, t_j + \Delta t_j]$ and satisfies

$$\lim_{\|\vec{\Delta}t\|_{\max} \to 0} \|m_{\vec{\Delta}t} - m\| = 0. \tag{3.2.27}$$

The continuous measure can be used to design an experiment with any number of additional points by distributing the points uniformly in the intervals around the design points. The size of the Δt_j's are chosen based on how good the information about the parameter being used to find the D-optimal design is and how sensitive the D-optimal design is to changes in the design points. More aggressive designs can be used when the initial parameter estimate is "good" and more moderate designs can be used when less is known about the quality of the estimate and the design is very sensitive.

3.2.2 Optimal Designs for Bounded Error Problems

When validating a parameter estimate using bounded errors, the idea of a V-Optimal design is very intuitive. Simply put, we want to find the design that minimizes the measure of the membership set that corresponds to the bounded errors. While the concept is easy to grasp, actually finding a design that is V-optimal is a costly numerical task. Each iterations requires an independent calculation of the approximation of the admissible parameters. For small problems this process is reasonable but as the problem grows so does the complexity of finding the V-optimal design. To address the numerical complexity of the problem we will consider approximate V-optimal designs based on techniques from section 2.2.3. Now, instead of searching for the V-optimal design we will search for the optimal design with respect to the the approximations of the membership set.

Designing an optimal experiment for a model where the errors are assumed bounded consists of finding an experiment that minimizes the measure of the membership set. Pronzato and Walter addressed some of the issues involved in designing experiments in [34].

We will be careful to note that the definition of a bounded error experiment contains two parts. Each component of a bounded error experiment contains a time component and an interval. Since we are trying to determine where in time to sample, we will write the bounded error experiment as a function of the time sample. When we design an experiment we have in mind a particular model and a particular parameter. We consider two distinct types of problems, one where the intervals in the bounded error experiment have fixed length and one where the length of the intervals is variable. For clarity we will call the case where the intervals have fixed length an ordinary V-optimal design problem and the case when the intervals vary with time a generalized V-optimal problem.

We begin by expressing an bounded error experiment for an ordinary V-optimal design problem showing the dependance of the experiment $\{t_j\}$. For fixed $\theta \in \Phi$, define

$$E^{N,m}(\{t_j\}_{j=1}^{N}) = \{(t_j, [y_i(t_j,\theta) - \varepsilon_i^1, y_i(t_j,\theta) + \varepsilon_i^2])\}_{j=1,\ldots,N}^{i=1,\ldots,m}$$

where $\varepsilon^1 = [\varepsilon_1^1, \varepsilon_2^1, \ldots, \varepsilon_m^1]^T \subset \mathbb{R}_+^m$ and $\varepsilon^2 = [\varepsilon_1^2, \varepsilon_2^2, \ldots, \varepsilon_m^2]^T \subset \mathbb{R}_+^m$ are constant. Using this notation we can define an ordinary V-optimal design.

Definition 3.2.8. *Given $\varepsilon^1 \subset \mathbb{R}_+^m$, $\varepsilon^2 \subset \mathbb{R}_+^m$ and a set $K \subset \mathbb{R}^p$, an experiment ξ_N^* is an ordinary V-optimal design over a candidate set $\mathcal{T}$ for fixed $\theta \subset \Phi$ if*

$$\xi_N^* = arg \min_{\{\xi_N \in \mathcal{T}\}} \ell(\mathcal{S}(E^{N,m}(\xi_N), K)),$$

where $\updownarrow$ is the Lebesgue measure.

It is also interesting to consider a design problem where ε^1 and ε^2 are time dependant functions instead of constants. We call this a generalized V-optimal design problem. A bounded error experiment for an generalized V-optimal design problem showing the dependance of the experiment $\{t_j\}$ is defined by

$$E^{N,m}(\{t_j\}_{j=1}^N) = \{(t_j, [y_i(t_j,\theta) - \varepsilon_i^1(t_j), y(t_j,\theta) + \varepsilon_i^2(t_j)])\}_{j=1,\ldots,N}^{i=1,\ldots,m}$$

where $\varepsilon^1(\cdot) : [0,T] \to \mathbb{R}_+^m$ and $\varepsilon^2(\cdot) : [0,T] \to \mathbb{R}_+^m$ are functions. Using this notation we can define a generalized V-optimal design.

Definition 3.2.9. *Given functions $\varepsilon^1(\cdot) : [0,T] \to \mathbb{R}_+^m$ and $\varepsilon^2(\cdot) : [0,T] \to \mathbb{R}_+^m$ and a set $K \subset \mathbb{R}^p$, an experiment ξ_N^* is a generalized V-optimal design over a candidate set $\mathcal{T}$ for fixed $\theta \subset \Phi$ if*

$$\xi_N^* = arg \min_{\{\xi_N \in \mathcal{T}\}} \ell(\mathcal{S}(E^{N,m}(\xi_N), K)),$$

where $\updownarrow$ is the Lebesgue measure.

V-optimal designs are naturally related to D-optimal designs for scalar models because both methods try to minimize the measure of a confidence region. For linear scalar models Pronzato and Walter [35] proved that the two methods are equivalent. However, for non-linear models a general comparison does not exist.

Finding a V-optimal design is in general extremely costly because the calculation of the membership set is computationally expensive. The application of the bounded error parameter identification method detailed in Chapter 2 is clearly apparent in this application because of its computational speed.

3.2.3 Approximating V-Optimal Designs

Solving exactly for V-optimal designs essentially impossible for the majority of the problems we are considering. We address this problem by applying the methods detailed in Section 2.2.3 to find approximately V-optimal designs for the model $y(t,\theta)$. To better define the problem we start with a definition,

Definition 3.2.10. *Given $\varepsilon^1 \subset \mathbb{R}_+^m$, $\varepsilon^2 \subset \mathbb{R}_+^m$ and a set $K \subset \mathbb{R}^p$, an experiment $\xi_N^{*,\rho^\mu\pi^\delta,k}$ is an ordinary approximately V-optimal design over a candidate set $\mathcal{T}$ for fixed $\theta \subset \Phi$, k, π^δ, and ρ^μ if*

$$\xi_N^{*,\rho^\mu\pi^\delta,k} = arg \min_{\{\xi_N \in \mathcal{T}\}} \ell(\mathcal{S}^{\rho^\mu\pi^\delta,k}(E^{N,m}(\xi_N), K)),$$

where $\updownarrow$ is the Lebesgue measure.

As we did in the exact problem we also consider a design problem where ε^1 and ε^2 are time dependant functions instead of constants. We call this a generalized approximate V-optimal design problem.

Definition 3.2.11. *Given functions $\varepsilon^1(\cdot) : [0,T] \to \mathbb{R}^m_+$ and $\varepsilon^2(\cdot) : [0,T] \to \mathbb{R}^m_+$ and a set $K \subset \mathbb{R}^p$, an experiment $\xi_N^{*,\rho^\mu\pi^\delta,k}$ is a generalized approximately V-optimal design over a candidate set $\mathcal{T}$ for fixed $\theta \subset \Phi$ k, π^δ, and ρ^μ if*

$$\xi_N^{*,\rho^\mu\pi^\delta,k} = arg \min_{\{\xi_N \in \mathcal{T}\}} \ell(\mathcal{S}^{\rho^\mu\pi^\delta,k}(E^{N,m}(\xi_N), K)),$$

where $\updownarrow$ is the Lebesgue measure.

These approximate designs can capture the majority of the information that a true V-optimal design would give.

As we mentioned for other types of optimal designs, it is important to remember that a V-optimal or approximately V-optimal designs are based off of an estimate of the true parameter. Therefore, when searching for a design to use in a bounded error context, it is important to investigate how the design behaves for parameters near the parameter estimate to ensure that the design is sufficiently robust.

Chapter 4

Numerical Results

4.1 Bounded Error Parameter Validation

In this section we demonstrate the results of Section 2.2.3 applied to a model described by the logistic equation and a SEIR model. We will explain how to use the results the theorem and lemmas through examples based on these models.

4.1.1 The Logistic Model

We chose a model that describes the logistic equation because we can solve for the membership set described in Theorem 2.2.3 exactly and compare it to the approximation defined by Theorem 2.2.7. Additionally, we can generate nice graphical representations of the membership set for two parameter parameter identification problems. We consider a model described by logistic equation,

$$\dot{x}(t,\theta) = cx(t,\theta) - dx(t,\theta)^2, \tag{4.1.1}$$

$$x(0,\theta) = 0.1, \tag{4.1.2}$$

with observation given by

$$y(t,\theta) = x(t,\theta). \tag{4.1.3}$$

Here, $\theta \in \Phi = \{[d,c]^T \in \mathbb{R}^2_+ | c/d > .1\}$, $t \in [0,20]$, and $x : [0,20] \times \Phi \to \mathbb{R}^1_+$.

For all $\theta \in \Phi$ and $t \in [0, 20]$, $\partial y(t, \theta)/\partial d < 0$. Furthermore, the model is onto the interval in the models output given in the experiment

$$E^{1,1} = \{(3, [y(3, [0.5, 1]^T) - .1, y(3, [0.5, 1]^T) + .1])\}.$$

For this experiment we can explicitly solve for the function

$$z_{t_1,a_1}(c) = \frac{ce^{ct_1 - a_1/x_0}}{a_1(-1 + e^{ct_1})}, \tag{4.1.4}$$

defined in Theorem 2.2.3 where $c \in [.85, 1.15]$. Similarly we can explicitly solve for the function

$$z_{t_1,b_1}(c) = \frac{ce^{ct_1 - b_1/x_0}}{b_1(-1 + e^{ct_1})}, \tag{4.1.5}$$

where $c \in Q = [.85, 1.15]$. Figure 4.1 shows z_{t_1,b_1}, the bottom curve, z_{t_1,a_1}, the top curve, and $\mathcal{S}(E^{1,1}, K)$, the area between the curves. This is a very special example and in most practical problems we cannot solve the differential equation exactly so we can not solve for the membership set exactly. We will now proceed by approximating the membership set using only the differential equation and then compare the approximation to the exact result we obtained.

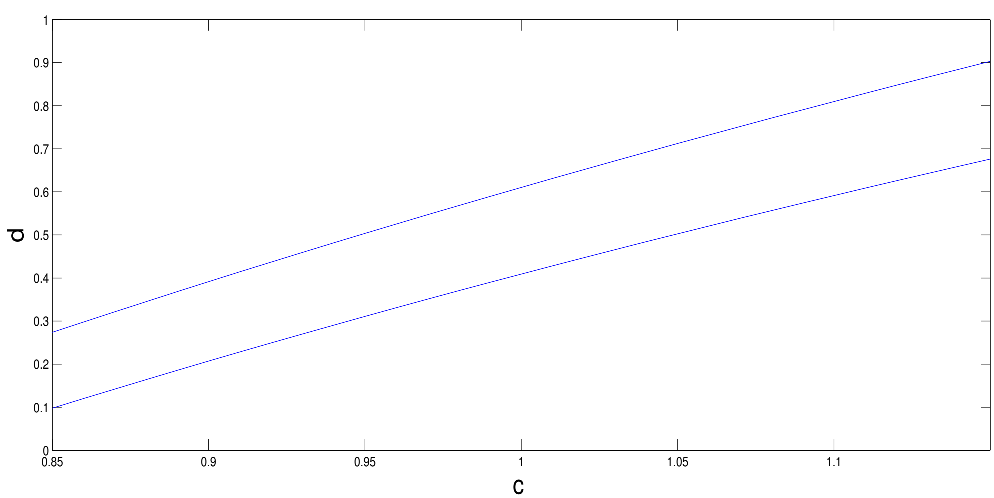


Figure 4.1: $\mathcal{S}(E^{1,1}, K)$

All of the following numerical calculations were preformed on a 2.5Ghz Core 2 Duo CPU using the software package MATLAB 2008a. To approximate the membership set we will first approximate the model described in Equations (4.1.1) and (4.1.2) using the Runge Kutta method on the partition
$\rho^{0.1} = [0, 0.1, 0.2, \ldots, 19.9, 20]$. With this approximation we can find a corresponding approximation for z_{t_1,a_1} and z_{t_1,b_1} at the points of a partition of Q defined by $\pi^{0.1} = [0.85, 0.95, 1.05, 1.15]$ using 4 iterations of Newton's method. Interpolating those points describes the functions $z_{t_1,a_1}^{\rho^\mu \pi^\delta, k}$ and $z_{t_1,b_1}^{\rho^\mu \pi^\delta, k}$ that form the boundary of the membership set $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{1,1}, K)$ seen in Figure 4.2. Finding this set took 0.065885 seconds. Finally, Figure 4.3 shows a graphical comparison of $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{1,1}, K)$, the area between the red lines and $\mathcal{S}(E^{1,1}, K)$, the areas between the blue lines. Visually we can see that $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{1,1}, K)$ is a very good approximation of $\mathcal{S}(E^{1,1}, K)$.

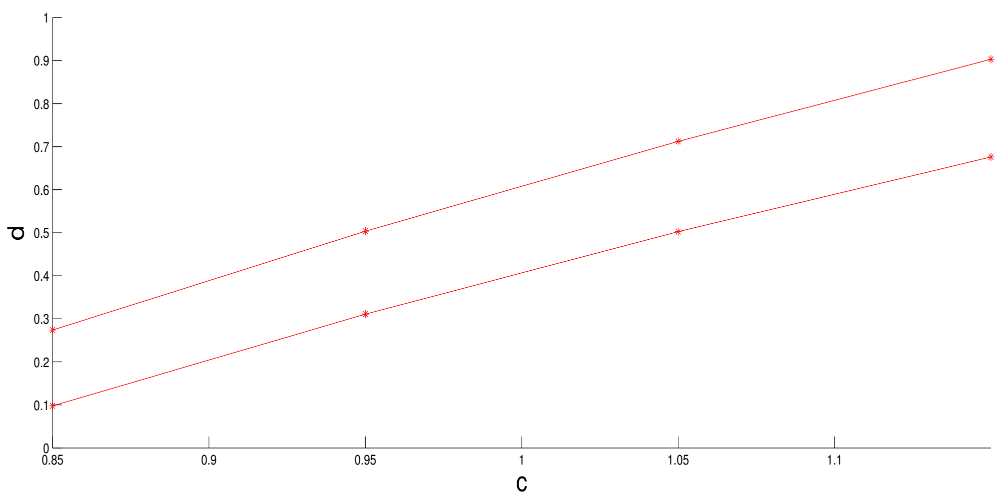


Figure 4.2: $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{1,1}, K)$

With a modest computational cost we are able to find an approximation $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{1,1}, K)$ that satisfies

$$d_H\left(\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{1,1}, K), \mathcal{S}(E^{1,1}, K)\right) < 10^{-3}$$

where $d_H(\cdot,\cdot)$ is the Hausdorff metric. Also, $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{1,1}, K)$ has a relative error of less than 1% from $\mathcal{S}(E^{1,1}, K)$.

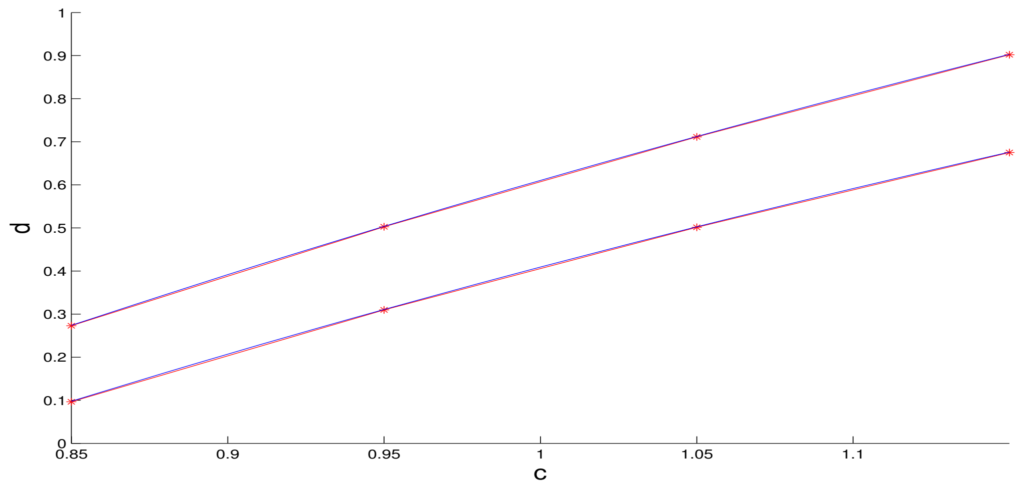


Figure 4.3: $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{1,1}, K)$

Next we consider the experiment

$$\begin{aligned} E^{2,1} &= \{E_1^{2,1}, E_2^{2,1}\} \\ &= \{(3, [y(3, [0.5, 1]^T) - .1, y(3, [0.5, 1]^T) + .1]), \\ &\qquad (15, [y(15, [0.5, 1]^T) - .1, y(15, [0.5, 1]^T) + .1])\}. \end{aligned}$$

To generate approximations of the membership set $\mathcal{S}(E^{2,1}, K)$, we will again solve Equations (4.1.1) and (4.1.2) using the Runge Kutta method on the partition $\rho^{0.1} = [0, 0.1, 0.2, \ldots, 19.9, 20]$ to approximate the model. Additionally we let $Q = [0.85, 1.15]$ and consider the partition $\pi^{0.1} = [0.85, 0.95, 1.05, 1.15]$. Figure 4.4 shows a graphical representation of $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E_1^{2,1}, K)$, the area between the red curves, and $\mathcal{S}(E_2^{2,1}, K)$, the area between the blue curves. The intersection of these two sets represents the set $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{2,1}, K)$.

Figure 4.5 shows the functions $l_{E^{2,1}}^{\rho^{0.1}\pi^{0.1},4}$, the black dashed curve, and $u_{E^{2,1}}^{\rho^{0.1}\pi^{0.1},4}$, the black solid curve, that define the boundary of the membership set.

This example is admittedly simple but this method can implemented on problems where the model has vector output and there are more parameters to be identified. As with any

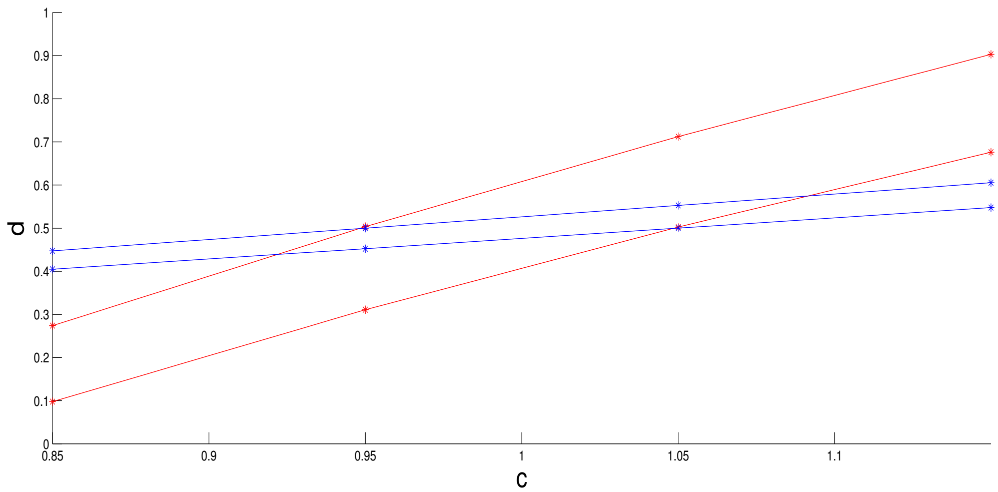


Figure 4.4: $\mathcal{S}^{\rho^{0.1}\pi^{0.1},4}(E^{2,1}, K)$

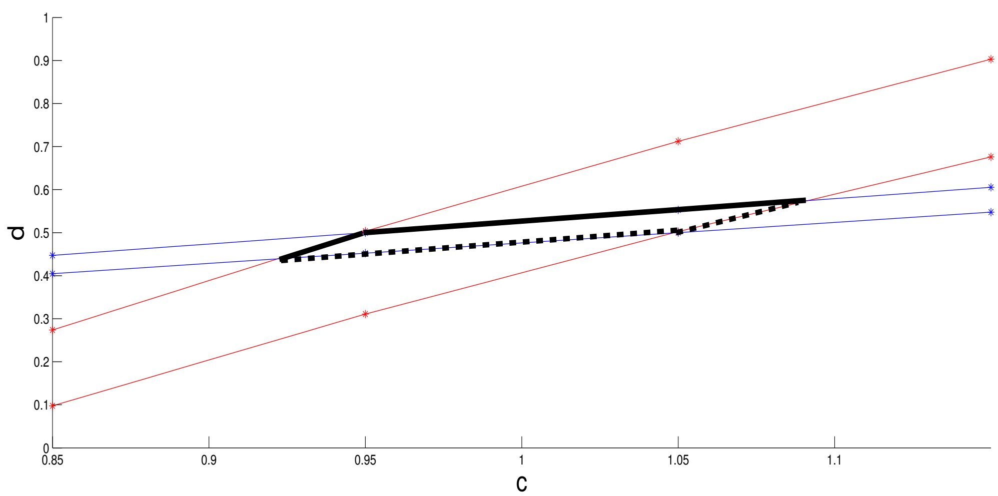


Figure 4.5: $l_{E^{2,1}}^{\rho^{0.1}\pi^{0.1},4}$ (black dashed) and $u_{E^{2,1}}^{\rho^{0.1}\pi^{0.1},4}$ (black solid)

bounded error parameter identification method, the time required to solve for the membership set increases exponentially with the number of unknown parameters.

4.1.2 The SEIR Model

The SEIR model is one way to mathematically explain the spread of an epidemic through a population. The model is 4 dimensional and has components that represent groups in the population that are susceptible, exposed, infected, and recovered. This is a generalization of the SIR model which does not take into account the group that has been exposed to the disease but is not yet showing symptoms. The mathematical model is defined by the relation

$$\begin{aligned}
\dot{S}(t,\theta) &= uN - uS - b\frac{I}{N}S, && (4.1.6)\\
\dot{E}(t,\theta) &= b\frac{I}{N}S - (u+a)E, && (4.1.7)\\
\dot{I}(t,\theta) &= aE - (v+u)I, && (4.1.8)\\
\dot{R}(t,\theta) &= vI - uR, && (4.1.9)
\end{aligned}$$

with initial condition $[S(0,\theta), I(0,\theta), E(0,\theta), R(0,\theta)]^T = [S_0, I_0, E_0, R_0]^T$ where $\theta = [v,a,b,u]^T \in \Phi$ and $t \in [0,T]$. The infection rate is described by b and the recovery rate is described by v. We make the assumption that the population is fixed, that is $N = S(t,\theta) + I(t,\theta) + E(t,\theta) + R(t,\theta)$ for all $t \in [0,T]$. We will be studying this model with output given by

$$y(t,\theta) = \mathcal{L}\begin{bmatrix} S(t,\theta) \\ E(t,\theta) \\ I(t,\theta) \\ R(t,\theta) \end{bmatrix}, \qquad (4.1.10)$$

where $\mathcal{L} \in L(\mathbb{R}^4, \mathbb{R}^m)$ is a linear operator and m is 1, 2, 3, or 4. As we will show, we are interested in the sensitivity of the model with respect to $\theta_1 = v$. Differentiating 4.1.6-4.1.9 with respect to $\theta_1 = v$ we get

$$\begin{aligned}
\dot{S}_v(t,\theta) &= -uS_v - (b\frac{I_v}{N}S + b\frac{I}{N}S_v), && (4.1.11)\\
\dot{E}_v(t,\theta) &= b\frac{I_v}{N}S + b\frac{I}{N}S - (u+a)E_v, && (4.1.12)\\
\dot{I}_v(t,\theta) &= aE_v - ((v+u)I_v + I), && (4.1.13)\\
\dot{R}_v(t,\theta) &= I + vI_v - uR_v, && (4.1.14)
\end{aligned}$$

with initial condition $[S_v(0,\theta), I_v(0,\theta), E_v(0,\theta), R_v(0,\theta)]^T = [0,0,0,0]^T$.

We are considering this problem in a bounded error context. The case we are interested in has the properties that $[S_0, E_0, I_0, R_0]^T = [.9999, 0, .0001, 0]^T$ and $u = 0$. Then the unknown parameter $\theta = [v, a, b, 0]^T$ becomes three dimensional. Additionally we are investigating the model with output given by

$$\begin{bmatrix} y_1(t,\theta) \\ y_2(t,\theta) \\ y_3(t,\theta) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S(t,\theta) \\ E(t,\theta) \\ I(t,\theta) \\ R(t,\theta) \end{bmatrix}. \tag{4.1.15}$$

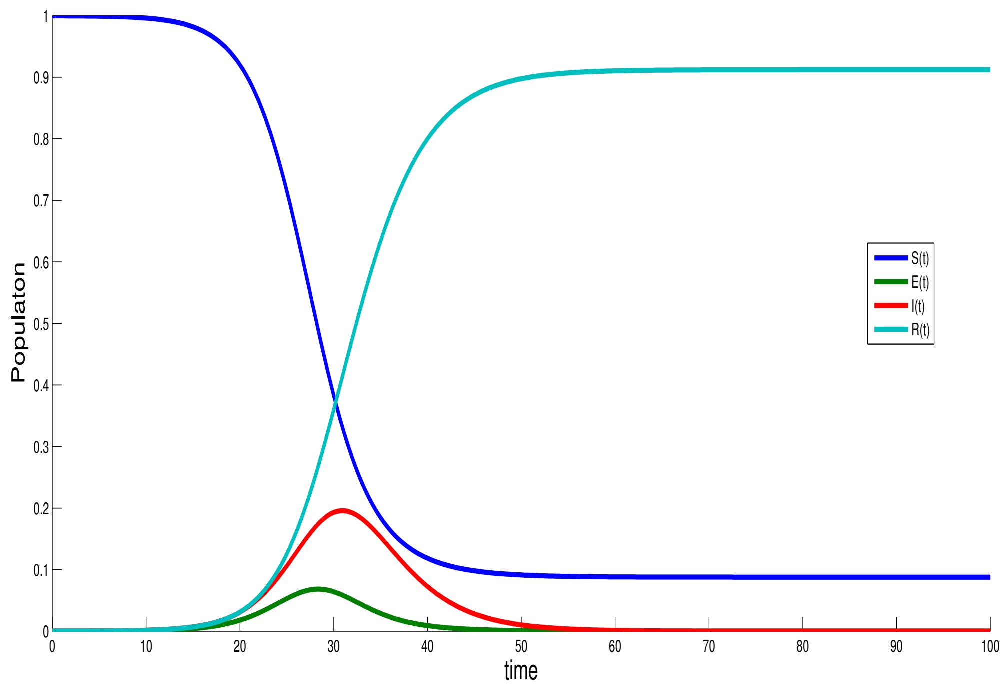


Figure 4.6: The SEIR model: $\theta = [.3, 1, .8, 0]^T$ and $[S_0, E_0, I_0, R_0]^T = [.9999, 0, .0001, 0]^T$.

We consider the experiment

$$E^{1,3} = \{25, [y_i(25, [.3, 1, .8, 0]^T) - .005, y_i(25, [.3, 1, .8, 0]^T) + .005]\}_{i=1}^{3}. \tag{4.1.16}$$

In practice $E^{1,3}$ corresponds to having data for the population exposed, infected, and recovered groups at time $t = 25$ with a maximum error of ± 0.005. All of the following numerical calculations were performed on a 2.5Ghz Core 2 Duo CPU using the software package MATLAB 2008a. For this experiment the hypothesis for Lemma 2.2.11 can be verified. Using the notation from Section 2.2.3 we define $Q = [0.235, 0.365] \times [0.95, 0.105]$. Let $\rho_1^{0.065/25} = [0.235, .235 + 0.065/25, \ldots .365 - 0.065/25, 0.365]$, $\rho_2^{0.05/25} = [0.95, 0.95 + 0.05/25, \ldots 1.05 - 0.05/25, 1.05]$ and $\rho^\mu = \rho_1^{0.065/25} \times \rho_2^{0.05/25}$. To approximate the model we are using a Runga-Kutta method on the uniform partition $\pi^{0.25}$. Then we can approximate the membership set $\mathcal{S}(E^{1,3}, K)$ using Lemma 2.2.11. The area between the surfaces in Figure 4.7 represents $\mathcal{S}^{\rho^\mu,\pi^\delta,10}(E_1^{1,3}, K)$ and took 289.82 seconds to compute. The area between the surfaces in Figure 4.8 represents $\mathcal{S}^{\rho^\mu,\pi^\delta,10}(E_2^{1,3}, K)$ and took 290.12 seconds to compute. The area between the surfaces in Figure 4.9 represents $\mathcal{S}^{\rho^\mu,\pi^\delta,10}(E_3^{1,3}, K)$ and took 290.09 seconds to compute. The intersection of the area between the surfaces in Fig. 4.10 represents $\mathcal{S}^{\rho^\mu,\pi^\delta,10}(E^{1,3}, K)$. The surface in Figure 4.11 represents $l_{a_1,25}^{\rho^\mu,\pi^\delta,10}$. The surface in Figure 4.11 represents $u_{a_1,25}^{\rho^\mu,\pi^\delta,10}$. Finding the approximation of the membership set $\mathcal{S}^{\rho^\mu,\pi^\delta,10}(E^{1,3}, K)$ took 911.71 seconds.

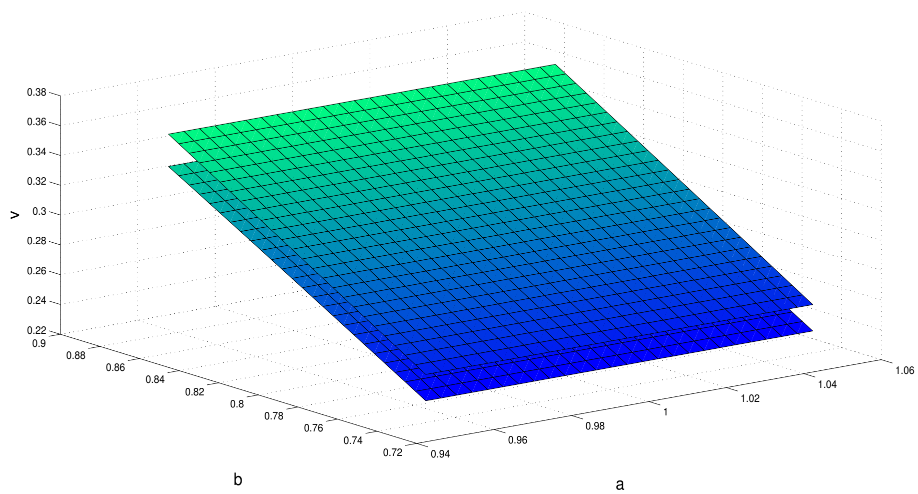


Figure 4.7: $\mathcal{S}^{\rho^\mu,\pi^\delta,10}(E_1^{1,3}, K)$

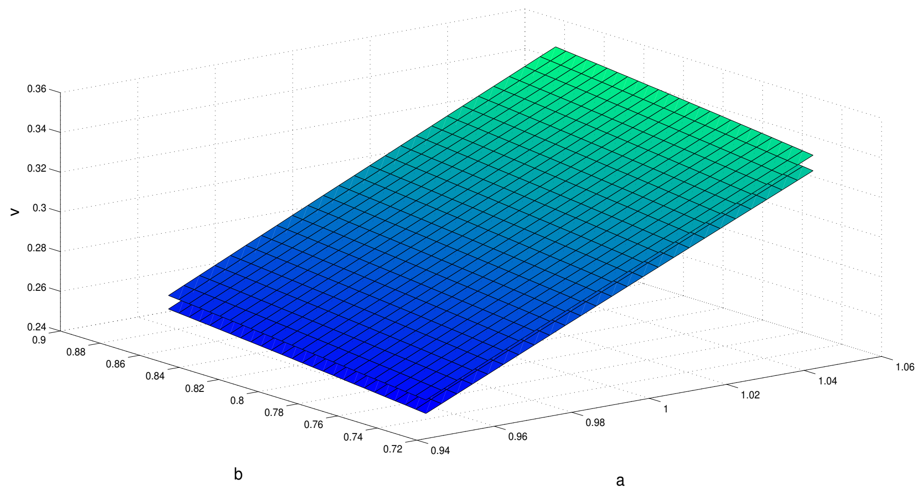


Figure 4.8: $\mathcal{S}^{\rho^{\mu},\pi^{\delta},10}(E_2^{1,3}, K)$

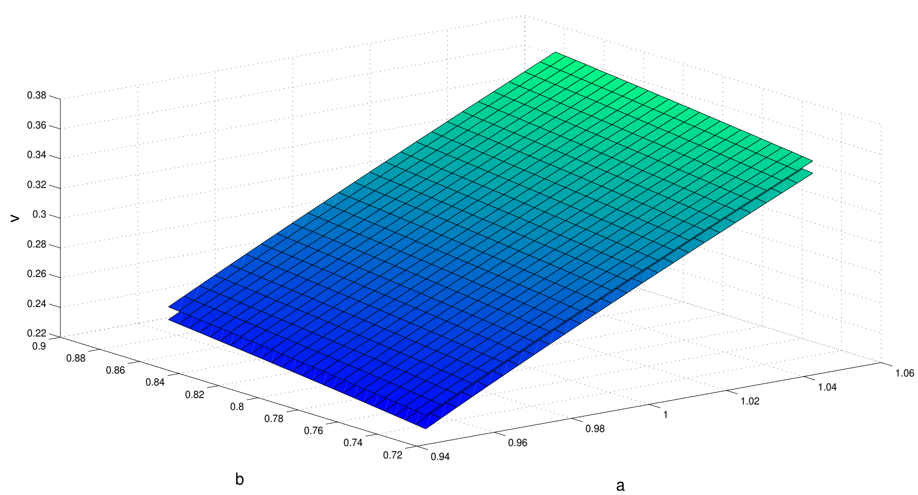


Figure 4.9: $\mathcal{S}^{\rho^{\mu},\pi^{\delta},10}(E_3^{1,3}, K)$

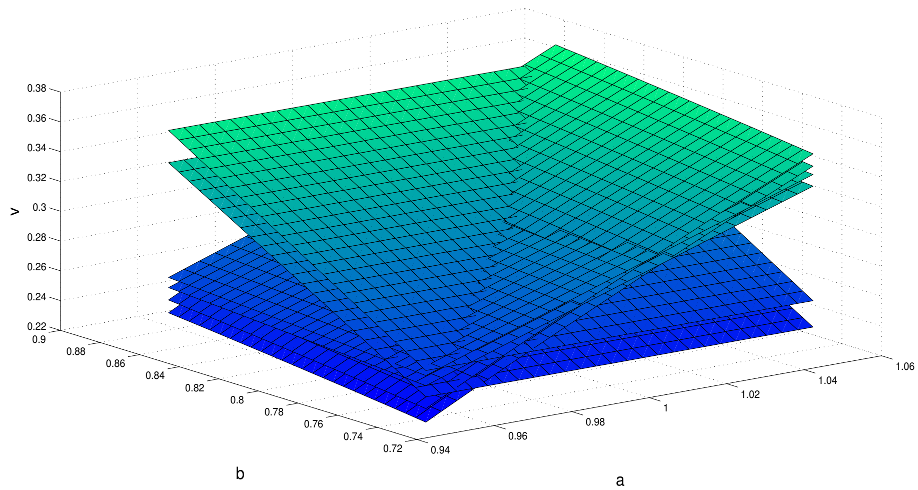


Figure 4.10: $\mathcal{S}^{\rho^{\mu},\pi^{\delta},10}(E^{1,3}, K)$

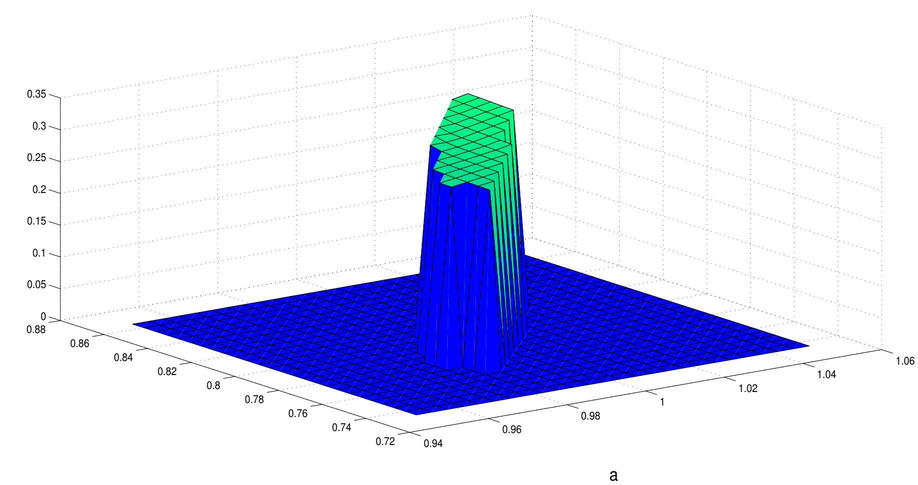


Figure 4.11: $l_{E^{1,3}}^{\rho^{\mu},\pi^{\delta},10}$

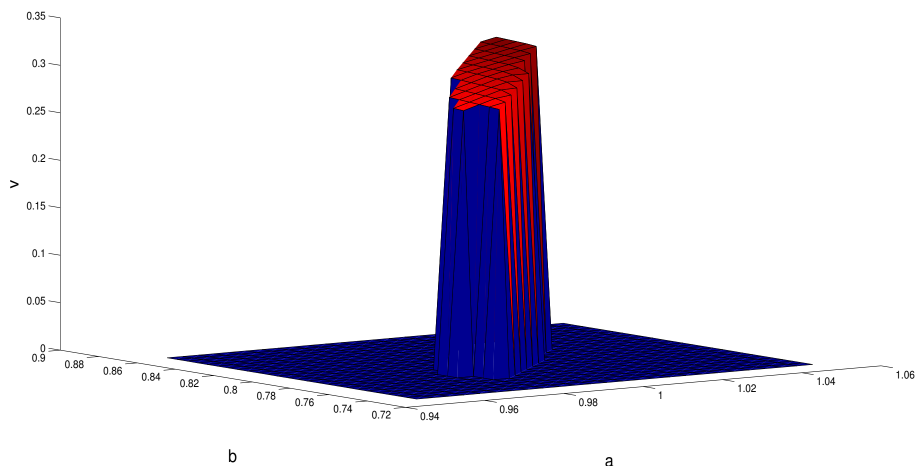


Figure 4.12: $u_{E^{1,3}}^{\rho^{\mu},\pi^{\delta},10}$

4.1.3 A Comparison to Existing Methods

As we mentioned in Section 2.2.3, determining which method of bounded parameter validation to use is heavily dependent on the problem. There are problems where the methods that use interval analysis are superior and as we see in this example, problems where our method is superior. To compare the methods we will revisit the logistic equation described by Equations 4.1.1 and 4.1.2. For the comparison we are going to look at two factors: accuracy and the number of differential equation solves necessary to find the membership set. While the differential equation solves are not all that go into either method, they are what take the majority of the computational time.

Given $K = \mathbb{R}^1_+ \times [0.85, 1.15]$ and the experiment

$$E^{1,1} = \{(3, [y(3, [0.5, 1]^T) - .1, y(3, [0.5, 1]^T) + .1])\},$$

we want to estimate $\mathcal{S}(E^{1,1}, K)$. Using Theorem 2.2.3 and the functions 4.1.4 and 4.1.5 we can solve for the membership set exactly. The area between the blue curves in Figure 4.13 is $\mathcal{S}(E^{1,1}, K)$.

To implement our method defined in Section 2.2.3, we approximate the model using Runge-Kutta on the partition $\rho^{0.1} = [0, 0.1, 0.2, \ldots, 2.9, 3]$ and with the partition of $[.85, 1.15]$ defined by $\pi^{0.1} = [0.85, 0.95, 1.05, 1.15]$ we can approximate $\mathcal{S}(E^{1,1}, K)$ by $\mathcal{S}^{\rho^{0.1},\pi^{0.1},4}(E^{1,1}, K)$. Finding $\mathcal{S}^{\rho^{0.1},\pi^{0.1},4}(E^{1,1}, K)$ requires 64 differential equation solves and has a relative error of less than one percent and takes 0.065885 seconds on a 2.5Ghz Core 2 Duo CPU using the software package MATLAB 2008a. Figure 4.14 shows a graphical comparison of $\mathcal{S}(E^{1,1}, K)$ and $\mathcal{S}^{\rho^{0.1},\pi^{0.1},4}(E^{1,1}, K)$.

Alternatively we could approximate $\mathcal{S}(E^{1,1}, K)$ using a bounded error validation technique based on interval analysis. To complete the first step of this method with a test box with dimension 0.0125 by 0.06 in the cd-plane requires 168 **interval** differential equation solves to produce an approximation that has a relative error of over 37%. This method could improve the accuracy of the solution by bisecting the indeterminate boxes but this will require further interval differential equation solves. By interval differential equation solve, we mean solving the differential equation for $\theta \in A \subset \mathbb{R}^p$. A interval differential equation solve returns intervals in the model's output. Note that an interval differential equation solve

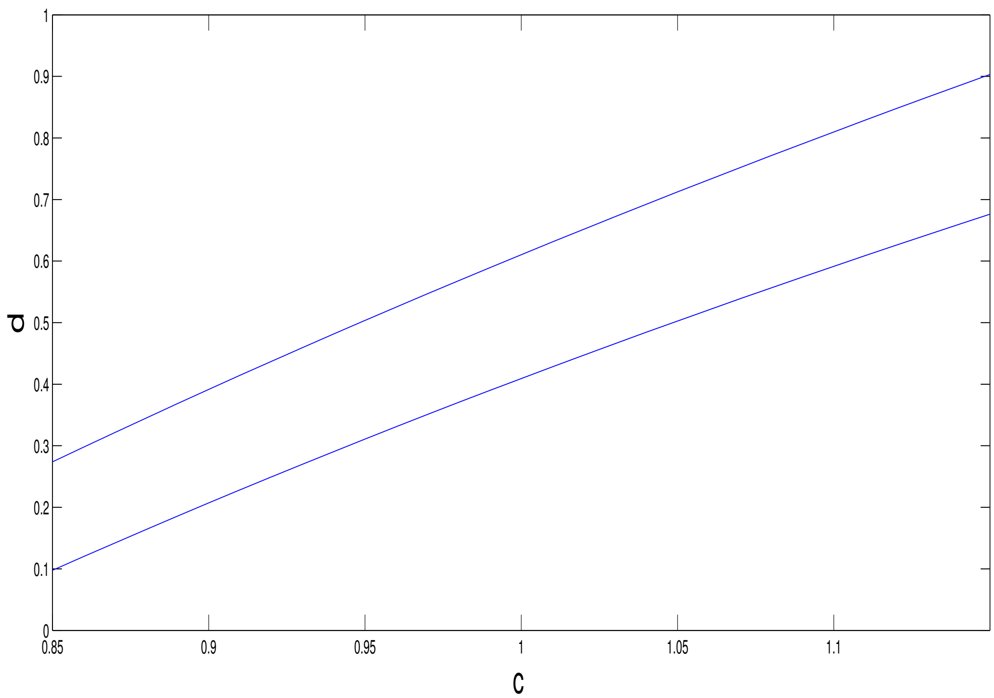


Figure 4.13: $\mathcal{S}(E^{1,1}, K)$

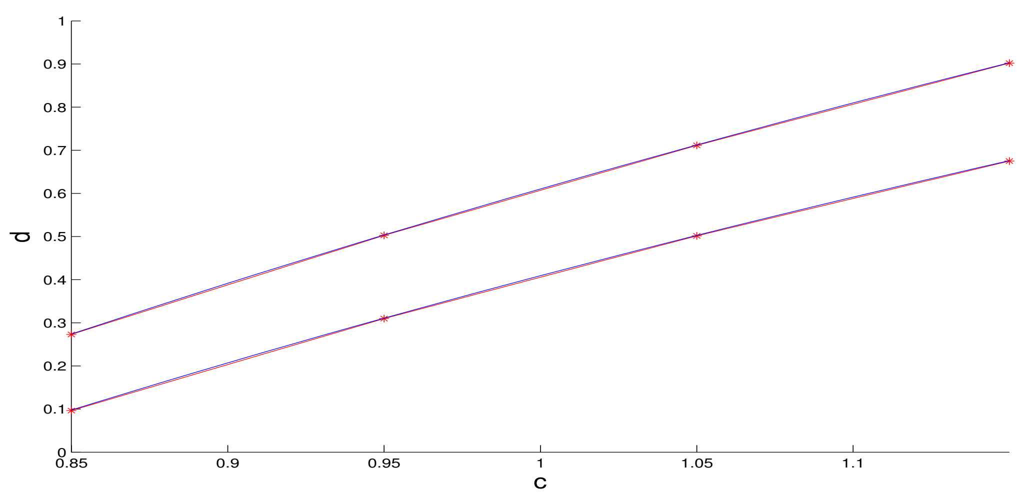


Figure 4.14: $\mathcal{S}(E^{1,1}, K)$ and $\mathcal{S}^{\rho^{0.1},\pi^{0.1},4}(E^{1,1}, K)$

is significantly more costly than solving the differential equation at a point $\theta \in \Phi$. Figure 4.15 show the estimate of $\mathcal{S}(E^{1,1}, K)$ using this method. Figure 4.16 shows the information that the first step of this method provides.

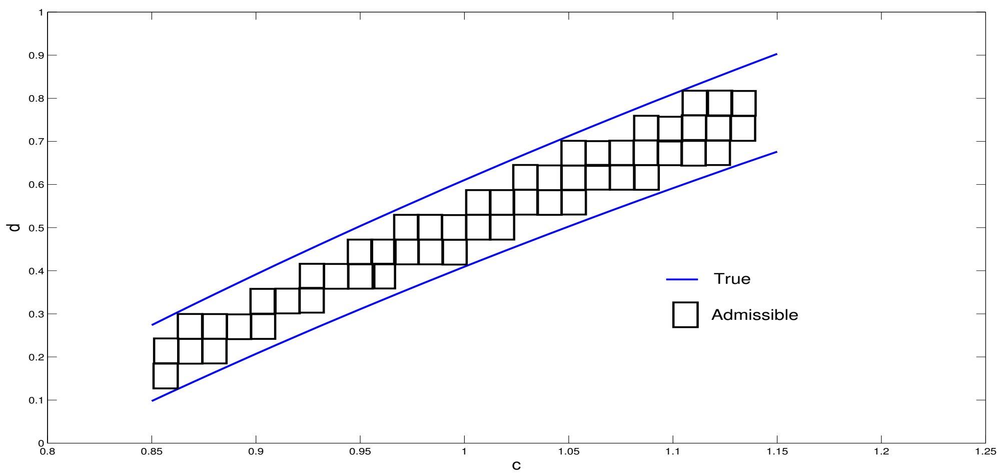


Figure 4.15: An approximation of $\mathcal{S}(E^{1,1}, K)$ using interval analysis.

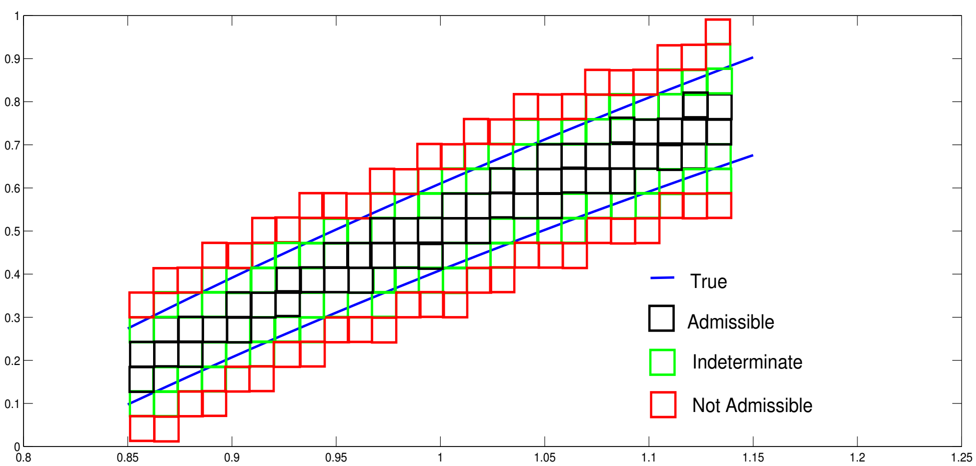


Figure 4.16: An approximation of $\mathcal{S}(E^{1,1}, K)$ using interval analysis.

4.1.4 V-optimal Designs

For a model described by the logistic equation (Equations 4.1.1 and 4.1.1) we consider designing an approximately ordinary V-optimal experiment defined by Definition 3.2.10. To begin the problem we suppose that $\theta = [0.5, 1]^T$ is an estimate of the true parameter and the values $\varepsilon_1 = \varepsilon_2 = 0.1$ explain the bounded regions. To implement our method defined in Section 2.2.3, we first approximate the model using Runge-Kutta on the partition $\rho^{0.1} = [0, 0.1, 0.2, \ldots, 2.9, 3]$. Letting $K = \mathbb{R}^1_+ \times [0.85, B_{max}]$ we define the partition $\pi^{0.1} = [0.85, 0.95, \ldots, B_{max} - 0.1, B_{max}]$ and can then approximate $\mathcal{S}(E^{2,1}, K)$ by $\mathcal{S}^{\rho^{0.1}, \pi^{0.1}, 8}(E^{2,1}, K)$ for a bounded error experiment $E^{2,1}$. Note $Q = [0.85, B_{max}]$ can vary with B_{max} and some experiments will produce membership sets that are not compact for any B_{max}. Given $\mathcal{T} = [0, 20]^2$, our goal will be to find a 2 dimensional experiment that satisfies

$$\xi_2^{*, \rho^{0.1} \pi^{0.1}, 8} = min_{\{\xi_2 \in \mathcal{T}\}} \ell(\mathcal{S}^{\rho^{0.1} \pi^{0.1}, k}(E^{2,1}(\xi_2), K)).$$

With the internal Matlab function 'fminsearch' we found that approximately ordinary V-optimal experiment is $\xi_2^{*, \rho^{0.1} \pi^{0.1}, 8} = \{3.5889, 20\}$. For different experiments the measure of the membership set can change dramatically. Table 4.1 shows the measure of the membership set with respect to several different experiments.

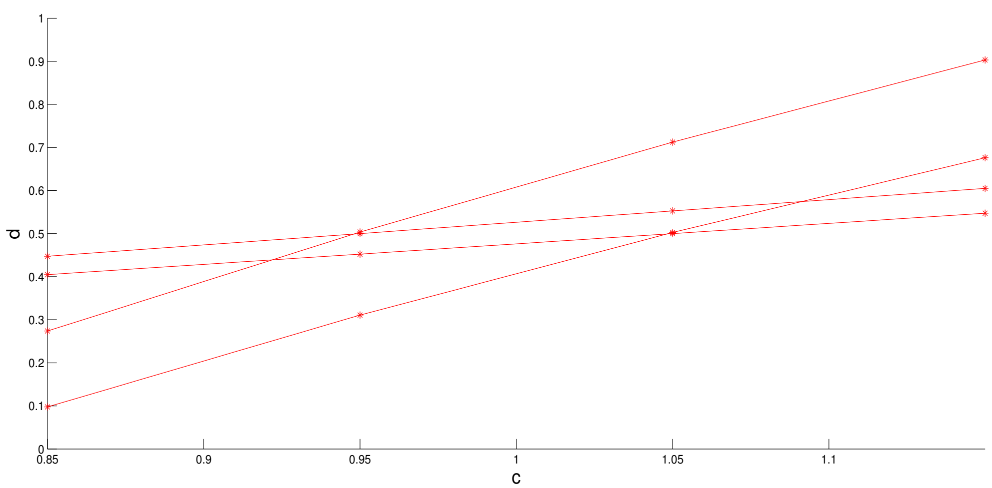


Figure 4.17: The membership set for the V-optimal design.

When we say that the measure of the membership set is ∞, we mean that $\lim_{B_{max} \to \infty} \mathcal{S}^{\rho^{0.1} \pi^{0.1}, k}(E^{2,1}(\xi_2), K) = \infty$.

ξ_2	$\ell(\mathcal{S}^{\rho^{0.1}\pi^{0.1},8}(E^{2,1}(\xi_2),K))$
$\{3.5589, 20\}$	0.006302
$\{3, 15\}$	0.006822
$\{2, 6\}$	0.014464
$\{5, 9\}$	0.013482
$\{8, 15\}$	∞
$\{6, 7\}$	∞

Table 4.1: The measure of the membership set for various experiments

As we see in Table 4.1 the measure of the membership set can change dramatically as the sample times change. For experiments where only a small amount of data can be collected, V-optimal designs give an intuitive and straightforward method for collecting data.

4.2 D-Optimal Designs

Verifying Theorem 3.4

Proving that the hypothesis hold for Theorem 3.2.4 is a difficult if not impossible task. We are going to be investigating a carefully constructed model for which the hypothesis of Theorem 3.2.4 can be verified. Consider the model described by the differential equation

$$\dot{x}(t,\theta) = a\frac{\sqrt{2}}{2}\cos(\frac{5\pi}{4}+t) - b\frac{\sqrt{2}}{2}\sin(\frac{5\pi}{4}+t) \tag{4.2.1}$$

where $\theta = [a,b]^T$ $t \in [0, 2\pi]$, and $x(0,\theta) = 0$. The model has observation described by

$$y(t,\theta) = x(t,\theta). \tag{4.2.2}$$

From the differential equation we can derive the sensitivity equations

$$\dot{s}_a(t,\theta) = \tfrac{\sqrt{2}}{2}\cos(\tfrac{5\pi}{4}+t) \tag{4.2.3}$$

$$\dot{s}_b(t,\theta) = -\tfrac{\sqrt{2}}{2}\sin(\tfrac{5\pi}{4}+t) \tag{4.2.4}$$

with initial condition $[s_a(0,\theta), s_b(0,\theta)]^T = [0,0]^T$.

For this example we can solve for the analytic solution to the differential equation, however, we will also considering this example from an approximate model perspective. Since we will be working with the exact model and an approximate model we will be very careful to distinguish between the two. The exact model is given by

$$y(t,\theta) = a\frac{\sqrt{2}}{2}\sin(\frac{5\pi}{4}+t) + b\frac{\sqrt{2}}{2}\cos(\frac{5\pi}{4}+t) + 1, \tag{4.2.5}$$

and the sensitives of the model with respect to the parameters are

$$s_a(t,\theta) = \frac{\sqrt{2}}{2}\sin(\frac{5\pi}{4}+t) + \frac{1}{2}, \tag{4.2.6}$$

$$s_b(t,\theta) = \frac{\sqrt{2}}{2}\cos(\frac{5\pi}{4}+t) + \frac{1}{2}. \tag{4.2.7}$$

The approximate model will be given by a linear interpolation of the points produced by solving the sensitivity equations using the Runge Kutta method on ρ^μ, a uniform partition of $[0, 2\pi]$. The approximate model and approximate sensitivities will be express as $w^{\rho^\mu}(t,\theta)$, and $s_a^{\rho^\mu}(t,\theta)$, $s_b^{\rho^\mu}(t,\theta)$.

The attainable region for the exact model is

$$R(\theta) = \{[s_a(t,\theta), s_b(t,\theta)]^T | t \in [0, 2\pi]\} \tag{4.2.8}$$

and for the approximate model is

$$R^{\rho^\mu}(\theta) = \{[s_a^{\rho^\mu}(t,\theta), s_b^{\rho^\mu}(t,\theta)]^T | t \in [0, 2\pi]\}. \tag{4.2.9}$$

Figure 4.18 shows the attainable sets for the exact and approximate model.

Since the dimension of the unknown parameter is 2, we will begin investigating the 2-point D-optimal designs. In [12], it is shown that the experiment that maximizes the determinate of the the Fisher information matrix correspond to the area of the simplex formed by the point $(0,0)$ and the points in the attainable set $[r_{1,1}, r_{1,2}]$ and $[r_{2,1}, r_{2,2}]$. For the exact model $R(\theta)$ is a parameterized circle so we can find the 2-point D-optimal design using geometry.

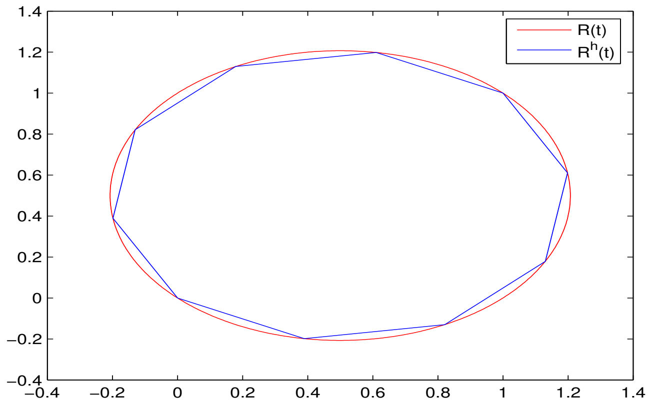
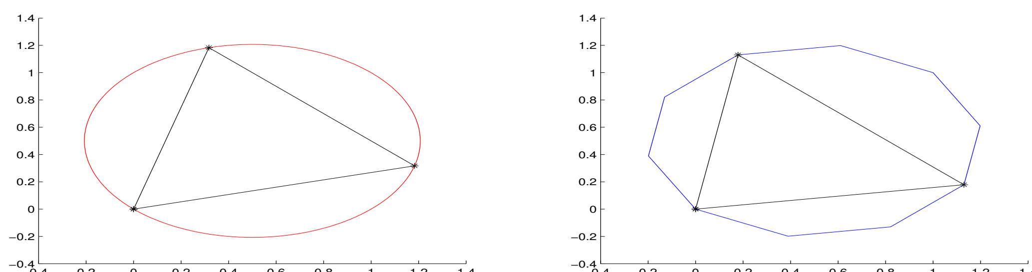


Figure 4.18: The attainable sets for the analytic model and the approximate model

The points $(0,0)$, $[s_a(2\pi/3,\theta), s_b(2\pi/3,\theta)]^T$, $[s_a(4\pi/3,\theta), s_b(4\pi/3,\theta)]^T$ form the simplex of the largest area so $\xi^* = \{2\pi/3, 4\pi/3\}$ is 2-point D-optimal design for the exact model. Finding the 2-point D-optimal design for the approximate model has to be done numerically. We use the internal Matlab function "fminsearch" to find the approximate minimum. For approximation when $\rho^\mu = [0, \frac{1}{9} * 2\pi, \frac{2}{9} * 2\pi, \ldots, \frac{8}{9} * 2\pi, 2\pi]$ a 2-point approximately D-optimal design is $\xi^*_{\rho^{\frac{2\pi}{9}}} \approx [3\pi/5, 7\pi/5]^T$.



(a) Analytic Model (b) Approximate Model

Figure 4.19: 2-point D-optimal designs

The locus of points $[w_1, w_2]^T$ satisfying the relationship

$$\sum_{u=1}^{p} \delta^2_{p+1,-u} = \Delta_p \tag{4.2.10}$$

or equivalently

$$\left|\begin{array}{cc} s_a(2\pi/3,\theta) & s_b(2\pi/3,\theta) \\ w_1 & w_2 \end{array}\right|^2 + \left|\begin{array}{cc} w_1 & w_2 \\ s_a(4\pi/3,\theta) & s_b(4\pi/3,\theta) \end{array}\right|^2 = \left|\begin{array}{cc} s_a(2\pi/3,\theta) & s_b(2\pi/3,\theta) \\ s_a(4\pi/3,\theta) & s_b(4\pi/3,\theta) \end{array}\right|^2 \tag{4.2.11}$$

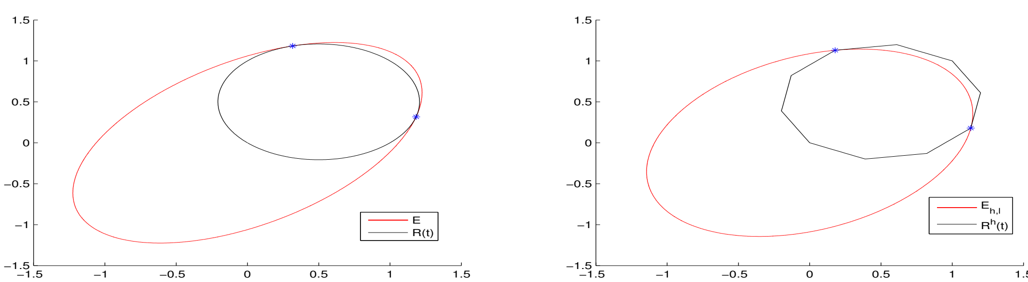
are E, the p-ellipsoid described in Theorem 3.2.2. Similarly the locus of points $[w_1, w_2]^T$ satisfying the relationship

$$\sum_{u=1}^{p} [\delta^{\rho^\mu,l}_{p+1,-u}]^2 = \Delta^{\rho^\mu,l}_p. \tag{4.2.12}$$

or equivalently

$$\left|\begin{array}{cc} s_a^{\rho^\mu}(3\pi/5,\theta) & s_b^{\rho^\mu}(3\pi/5,\theta) \\ w_1 & w_2 \end{array}\right|^2 + \left|\begin{array}{cc} w_1 & w_2 \\ s_a^{\rho^\mu}(7\pi/5,\theta) & s_b^{\rho^\mu}(7\pi/5,\theta) \end{array}\right|^2 = \left|\begin{array}{cc} s_a^{\rho^\mu}(3\pi/5,\theta) & s_b^{\rho^\mu}(3\pi/5,\theta) \\ s_a^{\rho^\mu}(7\pi/5,\theta) & s_b^{\rho^\mu}(7\pi/5,\theta) \end{array}\right|^2 \tag{4.2.13}$$

are $E^{\rho^{\frac{2\pi}{9}},10}$, the p-ellipsoid described in Theorem 3.2.4.



(a) Analytic Model (b) Approximate Model

Figure 4.20: Graphical check of the hypothesis of Theorem 3.2.2.

Notice in Figure 4.20 the exact model appears to satisfy the hypothesis in Theorem 3.2.2 but the approximate model does not. We can prove that the analytic model does indeed satisfy the hypothesis in Theorem 3.2.2 by confirming that there are only two points of intersection between $R(\theta)$ and E. It turns out that regardless of the choice of ρ^μ, the corresponding approximation does not satisfy the hypothesis required to apply Theorem

3.2.2. However, the approximation does satisfy Theorem 3.2.4 since

$$\lim_{\mu\to 0}\lim_{l\to\infty} d_H(E^{\rho^{\mu},l}, E) = 0 \quad (4.2.14)$$
$$\lim_{\mu\to 0} d_H(R^{\rho^{\mu}}(\theta), R(\theta)) = 0 \quad (4.2.15)$$

4.2.1 Application to Theorem 3.2.4

Showing that the hypothesis of Theorem 3.2.4 hold for a particular model is very cumbersome. We were able to prove that they do for the example in the previous section but we used the exact solution in the proof. For many problems we can only solve the sensitivity equations numerically. For these problems we cannot prove that the hypothesis of Theorem 3.2.4 holds but we can use the idea to design an experiment. In this subsection we will check to see if the hypothesis for the theorem appear to hold graphically, and if they do, we will use the techniques describe in Section 3.2.1 to design an experiment.

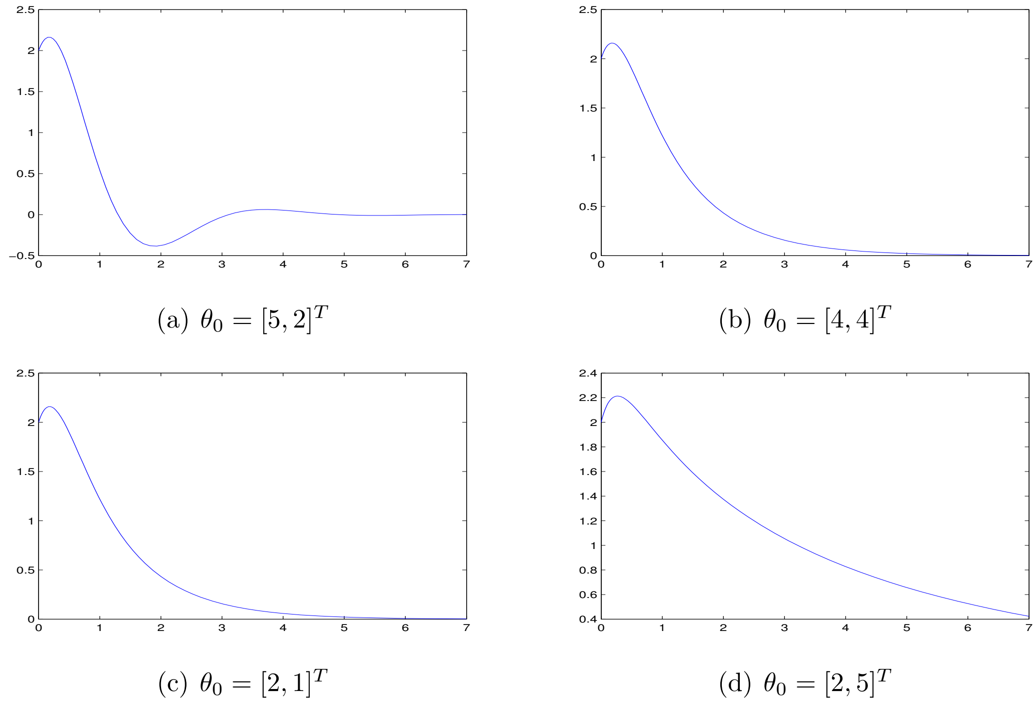
Consider the differential equation described by

$$\dot{\begin{bmatrix} x_1(t,\theta) \\ x_2(t,\theta) \end{bmatrix}} = \begin{bmatrix} 0 & 1 \\ a & b \end{bmatrix}\begin{bmatrix} x_1(t,\theta) \\ x_2(t,\theta) \end{bmatrix} + \begin{bmatrix} 0 \\ sin(x_1(t,\theta)) \end{bmatrix} \quad (4.2.16)$$

with initial condition $[x_1(0,\theta), x_2(0,\theta)]^T = [x_{1,0}x_{2,0}]^T$ with $t \in [0,7]$ and observation

$$y(t,\theta) = [1 \;\; 0]\begin{bmatrix} x_1(t,\theta) \\ x_2(t,\theta) \end{bmatrix} \quad (4.2.17)$$
$$= x_1(t,\theta). \quad (4.2.18)$$

To design an experiment we will need the sensitives of the model with respect to the parameters so we derive the sensitivity equations



(a) $\theta_0 = [5,2]^T$ (b) $\theta_0 = [4,4]^T$

(c) $\theta_0 = [2,1]^T$ (d) $\theta_0 = [2,5]^T$

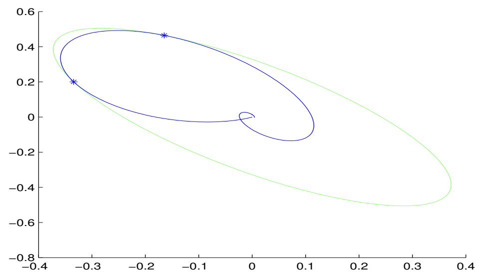
Figure 4.21: Plots of model (4.2.18) when $[x_{1,0} x_{2,0}]^T = [2,2]^T$

$$\begin{bmatrix} \dot{s}_{a1}(t,\theta) \\ s_{a2}(t,\theta) \end{bmatrix} = \begin{bmatrix} \frac{\partial x_2(t,\theta)}{\partial a} \\ x_1(t,\theta) + a\frac{\partial x_1(t,\theta)}{\partial a} + b\frac{\partial x_2(t,\theta)}{\partial a} + \cos(x_2(t,\theta))\frac{\partial x_1(t,\theta)}{\partial a} \end{bmatrix}$$

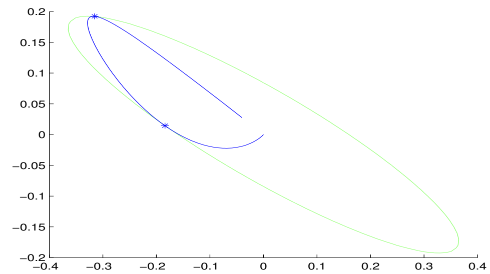
$$\begin{bmatrix} \dot{s}_{b1}(t,\theta) \\ s_{b2}(t,\theta) \end{bmatrix} = \begin{bmatrix} \frac{\partial x_2(t,\theta)}{\partial b} \\ a\frac{\partial x_1(t,\theta)}{\partial b} + x_2(t,\theta) + b\frac{\partial x_2(t,\theta)}{\partial b} + \cos(x_2(t,\theta))\frac{\partial x_1(t,\theta)}{\partial b} \end{bmatrix}.$$

with initial conditions $[s_{a1}(t,\theta), s_{a2}(t,\theta), s_{b1}(t,\theta), s_{b2}(t,\theta)]^T = [\frac{\partial}{\partial a}x_{1,0}, \frac{\partial}{\partial a}x_{2,0}, \frac{\partial}{\partial b}x_{1,0}, \frac{\partial}{\partial b}x_{2,0}]^T$.

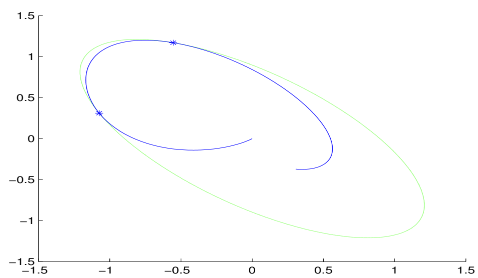
We want to investigate designing an experiment to identify the unknown parameter, θ_0, for the model given by (4.2.18). Using true parameters $[5,2]^T$, $[4,4]^T$, $[2,1]^T$, and $[2,5]^T$ we use "good" initial estimates that are within 5% of the true parameter and "bad" initial estimates that are within 30% of the true parameter to design experiments. The "good" intimal estimates are $[4.75, 2.1]^T$, $[3.6, 4.4]^T$, $[1.9, 1.05]^T$, and $[1.9, 5.25]^T$ and the "bad" initial estimates are $[3.5, 2.6]^T$, $[2.8, 5.2]^T$, $[1.4, 1.3]^T$, and $[1.4, 6.5]^T$. We will not attempt to prove that the hypothesis for Theorem 3.2.4 hold but we will check to see if they appear to hold graphically in Figure 4.22. We see that the model does appear to satisfy the hypothesis of Theorem 3.2.4 for the set of "good" initial estimates. While we do not show them, the graphs of the "bad" in initial estimates also appear to satisfy Theorem 3.2.4.



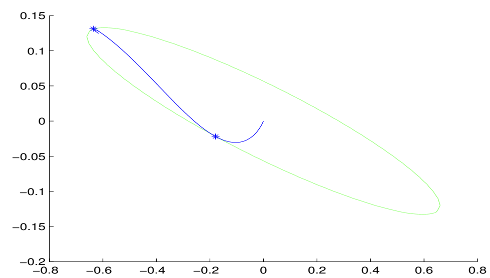
(a) $\theta_0 = [5, 2]^T$



(b) $\theta_0 = [4, 4]^T$



(c) $\theta_0 = [2, 1]^T$



(d) $\theta_0 = [2, 5]^T$

Figure 4.22: Graphical check of they hypothesis of Theorem 3.2.4.

Since we now have reason to believe that $N = k * 2$-point D-optimal design is repeated samples at the 2-point D-optimal design, we will base our design on the theory detailed in Section 3.2.1. We want to design an experiment that is close to the true optimal design but accounts for error in the parameter estimate. To design the experiment, we first find the 2-point D-optimal design for each initial estimate. Then for a given Δt, defined in Section 3.2.1, we design experiments that sample 25 points evenly in intervals defined by Δt around the 2-point D-optimal design. To test the designs we generate numerical data for a given experiment $\xi = \{t_j\}_{j=1}^{N}$ and parameter θ_0 of the form

$$d_j = w^{\rho^{\mu}}(t_j, \theta_0) + e_j \tag{4.2.19}$$

where e_j is realization of a normally distributed random variable with mean 0 and variance

σ^2 generated by the MATLAB function "random". We solve the sensitivity equations using the Runge-Kutta method on appropriately sized partitions. For each parameter and design, we run such an experiment 25 times. From the numerical data, we estimate the unknown parameter and calculate the standard errors for that estimate. We then average the standard errors from the 25 experiments and compare the results.

Table 4.2 shows the mean of the standard errors of 25 numerical experiments for designs based off of the 2-point D-optimal design with several $\vec{\Delta}t$'s for the "good" estimates of the initial parameter. The final column in Table 4.2 shows the mean of the standard errors of 25 numerical experiments with even sampling in the interval. In almost every case the best results came using the 2-point D-optimal design with $\vec{\Delta}t = [.20, .20]^T$. The only time this was not true was when $\theta = [1.9, 5.25]^T$.

Table 4.3 show the mean of the standard errors of 25 numerical experiments for designs based off of the 2-point D-optimal design for the given parameter and three different $\vec{\Delta}t$'s for "bad" estimates of the initial parameter. The final column in Table 4.3 show the mean of the standard errors of 25 numerical experiments with even sampling in the interval. Results for these tests varied widely in comparison of the good initial estimates. The best standard errors came from the smallest $\vec{\Delta}t$, the largest $\vec{\Delta}t$, and even sampling for the different parameters.

The numerical tests show that choosing a proper $\vec{\Delta}t$ is very important to designing a meaningful experiment. The quality of the initial estimate for the parameter and the sensitivity of the D-optimal design of that parameter are key factors for choosing $\vec{\Delta}t$. Numerical experiments can help determine how aggressively you can select $\vec{\Delta}t$. The more confidence you have in the initial parameter estimate, the smaller you can select $\vec{\Delta}t$.

	$\vec{\Delta}t = [.05, .05]^T$	$\vec{\Delta}t = [.20, .20]^T$	$\vec{\Delta}t = [.35, .35]^T$	even sampling
$\theta = [4.75, 2.1]^T$	$\begin{bmatrix} 0.0006439 \\ 0.0002966 \end{bmatrix}$	$\begin{bmatrix} 0.0006300 \\ 0.0002859 \end{bmatrix}$	$\begin{bmatrix} 0.00070628 \\ 0.0003143 \end{bmatrix}$	$\begin{bmatrix} .0021263 \\ 0.0009684 \end{bmatrix}$
$\theta = [3.6, 4.4]^T$	$\begin{bmatrix} 0.0006708 \\ 0.0003836 \end{bmatrix}$	$\begin{bmatrix} 0.0005751 \\ 0.0003180 \end{bmatrix}$	$\begin{bmatrix} 0.0006432 \\ 0.0003390 \end{bmatrix}$	$\begin{bmatrix} 0.0019580 \\ 0.0008978 \end{bmatrix}$
$\theta = [1.9, 1.05]^T$	$\begin{bmatrix} .15600 \\ 0.030550 \end{bmatrix}$	$\begin{bmatrix} 0.013339 \\ 0.0028978 \end{bmatrix}$	$\begin{bmatrix} 0.0045779 \\ 0.0011334 \end{bmatrix}$	$\begin{bmatrix} 0.0019931 \\ 0.0009071 \end{bmatrix}$
$\theta = [1.9, 5.25]^T$	$\begin{bmatrix} 0.0099486 \\ 0.090419 \end{bmatrix}$	$\begin{bmatrix} 0.0019024 \\ 0.011605 \end{bmatrix}$	$\begin{bmatrix} 0.002127 \\ 0.0068301 \end{bmatrix}$	$\begin{bmatrix} 0.0022963 \\ 0.001048 \end{bmatrix}$

Table 4.2: Standard Errors for Estimated Parameters.

	$\vec{\Delta}t = [.05, .05]^T$	$\vec{\Delta}t = [.20, .20]^T$	$\vec{\Delta}t = [.35, .35]^T$	even sampling
$\theta = [3.5, 2.1]^T$	$\begin{bmatrix} 0.0004904 \\ 0.0003200 \end{bmatrix}$	$\begin{bmatrix} 0.0005583 \\ 0.0003413 \end{bmatrix}$	$\begin{bmatrix} 0.0006156 \\ 0.0006156 \end{bmatrix}$	$\begin{bmatrix} .00201934 \\ 0.0009198 \end{bmatrix}$
$\theta = [2.8, 5.2]^T$	$\begin{bmatrix} 0.0014851 \\ 0.0083517 \end{bmatrix}$	$\begin{bmatrix} 0.0013199 \\ 0.0048472 \end{bmatrix}$	$\begin{bmatrix} 0.0013990 \\ 0.0034700 \end{bmatrix}$	$\begin{bmatrix} 0.0021421 \\ 0.0009730 \end{bmatrix}$
$\theta = [1.4, 1.3]^T$	$\begin{bmatrix} 0.0101341 \\ 0.0036910 \end{bmatrix}$	$\begin{bmatrix} 0.0063203 \\ 0.0023167 \end{bmatrix}$	$\begin{bmatrix} 0.0030270 \\ 0.0011508 \end{bmatrix}$	$\begin{bmatrix} 0.0018921 \\ 0.0008582 \end{bmatrix}$
$\theta = [1.4, 6.5]^T$	$\begin{bmatrix} 0.0097946 \\ 0.3613931 \end{bmatrix}$	$\begin{bmatrix} 0.0021961 \\ 0.0261397 \end{bmatrix}$	$\begin{bmatrix} 0.0023849 \\ 0.0113027 \end{bmatrix}$	$\begin{bmatrix} 0.0022814 \\ 0.0010416 \end{bmatrix}$

Table 4.3: Standard Errors for Estimated Parameters.

Chapter 5

Conclusion and Future Work

In this thesis we considered the problem of estimating parameters in models governed by dynamical systems when the number of data samples is limited. We developed a new method for bounded error estimation that accounted for the errors that are due to the numerical approximation of differential equations and the the parameter space grids. We combined these results with the theory of design of experiments to produce an algorithm for identifying and validating model parameters with limited data.

We also investigated these questions by using a continuous version of the Fisher Information Matrix to help guide the selection of sample times. In both cases we combined statistical assumptions inherent in the model and the mathematical properties of the dynamical system to develop new algorithms.

The basic idea guiding the effort is to use all available information to solve the problem. We observed that it is crucial to extrude information from the model by studying its dynamics before deciding on how to solve the problem. Our results on bounded error parameter validation shows that by placing some additional requirements on the hypothesis of standard methods one can significantly reduce the computational time necessary to identify the membership set. Similarly, in the design of experiment problem we found that analyzing the parameter estimate can lead to a more informative design. Using the quality of a parameter estimate instead of just the parameter estimate can go a long way in designing an experiment.

One of our goals in the future is to expand the ideas we have presented in this presentation to problems with more parameters and decrease the restrictions necessary to use them. For the bounded error parameter validation problem we plan to focus on improving the theoretical results to allow for more general application of our inverse method. Also, the standard (forward) bounded error method requires huge numbers of forward simulations which limits its application. We plan to revisit this approach by considering smart sampling algorithms such as Latin Hypercube and Smolyak sampling schemes to speedup computations. We would like to use the method to tackle problems with higher dimensional parameters and relax the requirements necessary to use the theory. Accomplishing these goals would allow us to use the power of these techniques on a broader class of problems and expand the results in this thesis to the design of experiments with large number of parameters and small data sets.

We have already expanded the ideas we present on bounded error parameter identification to delay differential equation [15]. We have successfully run numerical experiments using our method and the next step is to expand the theoretical results to handle delay differential equations. Also, we plan to look at similar problems where the basic models are described by partial differential equations. The bounded error approach to PDE problems will involve limited data in both time and space and hence will require a more complex mathematical framework. For example, the design of experiments for PDE problems must address the issue of selecting "best" time and spatial positions to ensure efficient experiment for parameter estimation and verification.

Bibliography

[1] D. Angeli and E. Sontag, Monotone Control Systems. *IEEE Trans. Automat. Control*, 48:1684-1698, 2003.

[2] S. P. Asprey and S. Macchietto, Designing robust optimal dynamic experiments. *Journal of Process Control*, 12:545-556, 2002.

[3] A. C. Atkinson and W. G. Hunter, The Design of experiments for parameter estimation. *Technometrics*, 10:271-289, 1968.

[4] P. Bai, H. T. Banks, A. Dediu, A. Y. Govan, M. Last, A. L. Loyd, H. K. Nguyen, M. S. Olufsen, G. Rempala, and B. D. Slenning, Stocastic and deterministic models for agricultral production networks. *Math. Biosci. and Engr,* 4:373-402, 2007.

[5] E. Bai and Y. Huang, Convergence of optimal sequential ourter bounding sets in bounded error parameter estimation. *Mathematics and Computers in Simultion*, 49:307-317, 1999.

[6] H. T. Banks, S. L. Ernstberger, and S. L. Grove. Standard Errors and Confidence Intervals in Inverse Problems: Sensitivity and Associated Pitfalls. *Journal of Inverse and Ill-Posed Problems*, 15, 2007

[7] Banks, H. T., Dediu, S., and S. L. Ernstberger. Sensitivity functions and their uses in inverse problems. *Journal of Inverse and Ill-Posed Problems*, 15:683708, 2007.

[8] H. T. Banks, M. Davidian, J. R. Samuels, and K. L. Sutton. An inverse problem statistical methodology summary. *CRSC Technical Reports*, TR08-01, 2008.

[9] H. T. Banks, S. Dediu, S. L. Ernstberger, and F. Kappel A new approach to optimal design. *CRSC Technical Reports*, TR08-012, 2008.

[10] G. Belforte, B. Bona, and V. Cerone. Parameter estimation algorithms for a set-membership description of uncertainty. *Automatica*, 26:887898, 1990.

[11] J. T. Borggaard and J. A. Burns. PDE sensitivity equation method for optimal aerodynamic design. *Journal of Computational Physiscs*, 136:366-384, 1997.

[12] G. E. P. Box and H. Lucas. Design of experiments in non-linear situations. *Biometrika*, 46:77-90, 1959.

[13] M. J. Box The occurrence of replications in optimal designs of experiments to estimate parameters in non-linear models. *Journal of the Royal Statistical Society*, 20:290-302, 1968.

[14] J. A. Burns, S. Doughty, and E. M. Cliff. Sensitivity analysis and parameter estimation for a model of chlamydia trachomatis infection. *Journal of illposed problems*, 15:19-32, 2007.

[15] J. A. Burns, A. F. Childers. Bounded error parameter identification for models described by ordinary and delay differential equations. *Preceeding to the 17th Mediterranean Conference on Control and Automation*, 2009.

[16] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, California, 2002.

[17] A. C. Davison, *Statistical Models*. Cambridge University Press, 2003.

[18] R. D. Driver. *Ordinary and Delay Differential Equations*, Springer-Verlag, New York 1977.

[19] N. R. Draper and W. G. Hunter. Design of experiments for parameter estimation in multiresponse situations. *Biometrika* 53:525-533, 1966.

[20] D. Faller, U. Klingmuller, and J. Timmer, Simulation methods for optimal experimental design in systems biology. *SIMULATION*, 79:717:725, 2003.

[21] E. Fogel and Y. Huang. On the value of information in system identification-bounded noise case. *Automatica*, 2:229-238, 1982.

[22] L. Jaulin. Nonlinear bounded-error state estimation of continuous-time systems. *Automatica*, 38:1979-1982, 2002.

[23] L. Jaulin and E. Walter. Set inversion via interval analysis for non-linear bounded error estimation *Automatica*, 29:1053-1064, 1993.

[24] R. I. Jennrich. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40:633-643, 1969.

[25] T. Johnson and W. Tucker. Rigorous parameter reconstruction for differential equations with noisy data. *Automatica*, 44:24222426, 2008.

[26] N. Kieffer and E. Walter. Garunteed parameter estimation for cooperative models. *Positive Systems*, Benvenuti L., Da Santis, A., Farina, L. (eds), Lecture Notes on Control and Inforamtion Sciences, Vol. 294. Springer, Berlin, 2004.

[27] N. Kieffer and E. Walter. Garunteed parameter estimation for knowledge based models. *Journtal of Computational and Applied Mathematics*, 199:277-285, 2007.

[28] J. King and W. Wong. Minimax D-optimal designs for the logistic equation. *International Biometric Society. Bol.*, 56:1263-1267, 2000.

[29] H. Lucas. The Occurance of Replication in Optimal Designs of Experiments to Estimate Parameters in Non-linear Models. *Journal of the Royal Statistical Society*, 30:190-202, (1968).

[30] M. Milanese, J. Norton, H. Piet-Lahanier and E. Walter, E. (Eds), *Bounding Approaches to System Identification*. Plenum Press, New York, NY, 1996.

[31] R. E. Moore. *Methods and Applications of Interval Analysis*, SIAM, Pennsylvania, 1979.

[32] M. Z. Nashed, Differentiability and related properties of nonlinear operators: some aspects of the role of differentials in nonlinear functional analysis. *Nonlinear Functional Analysis and Application*, Academic Press, 1971.

[33] L. Pronzato. Otimal experimental design and some related control problems. *Automatica* 44:303-325, 2008.

[34] L. Pronzato and E. Walter. Experiment design for bounded-error models. *Mathematics in Computers in Simulation*, 32:571-584, 1990.

[35] L. Pronzato and E. Walter. Experiment design in a bounded error context: comparison with D-optimality. *Automatica*, 25:383-391, 1989.

[36] T. Raissi, N. Ramdani, and Y. Candau. Parameter estimation for non-linear continuous-time systems in a bounded error context. *Proceedings of the 42nd IEEE Conference on Decision and Control*, 2003.

[37] T. Raissi, N. Ramdani, Y. Candau. Set membership state and parameter estimation for systems described by nonlinear differential equations, *Automatica* 40:17711777, 2004.

[38] W. Ruden. *Principles of Mathematical Analysis* McGraw-Hill, Inc., New York, 1976.

[39] G. A. F. Serber and C. J. Wild. *Nonlinear Regression.* John Wiley and Sons, Inc., New York, 1989.

[40] R. C. St. John and N. R. Draper. D-optimality for regression designs: A review. *Technometrics*, 17:15-23, 1975.

[41] K. Thomaseth and C. Cobelli. Generalized sensitivity functions in physiological system identification. *Ann. Biomed. Eng*, 27:607-616, 1999.