ABSTRACT

MATTHEWS, MEGAN LEIGH. Region of Attraction Estimation of Biological Continuous Boolean Models. (Under the direction of Cranos M. Williams.)

Quantitative analysis of biological systems has become an increasingly important research field as scientists look to solve current day health and environmental problems. The development of modeling and model analysis approaches that are specifically geared toward biological processes is a rapidly growing research area. Continuous approximations of Boolean models, for example, have been identified as a viable method for modeling such systems. This is because they are capable of generating dynamic models of biochemical pathways using inferred dependency relationships between components. The resulting nonlinear equations and therefore nonlinear dynamics, however, can present a challenge for most system analysis approaches such as region of attraction (ROA) estimation. Knowledge of the ROA for metabolic steady states in plants can give valuable insight into the robustness and sustainability of favorable phenotypes. This could lead to more efficient production of biofuels from plant biomass, sustained resistance of plants to climate change, and growth of agriculture on marginal lands. Continued progress in the area of biosystems modeling will require that computational techniques used to analyze simple nonlinear systems can still be applied to nonlinear equations typically used to model the dynamics associated with biological processes. In this thesis, we assess the applicability of a state of the art ROA estimation technique based on interval arithmetic to a subnetwork of the Rb-E2F signaling pathway modeled using continuous Boolean functions. We show that this method can successfully be used to provide an estimate of the ROA for dynamic models described using Hillcube continuous Boolean approximations. Additionally, we explore a methodology for selecting, from the many possibilities,

Lyapunov functions that result in improved estimates of the ROA.

© Copyright 2012 by Megan Leigh Matthews

All Rights Reserved

Region of Attraction Estimation of Biological Continuous Boolean Models

by
Megan Leigh Matthews

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Electrical Engineering

Raleigh, North Carolina

2012

APPROVED BY:

Brian L. Hughes

Edward Grant

Cranos M. Williams
Chair of Advisory Committee

DEDICATION

To my parents.

BIOGRAPHY

Megan Leigh Matthews was born in Greensboro, North Carolina on July 14, 1989. While growing up she lived in Asheboro, NC, Rowlands Gill, UK, and finally Avon Lake, Ohio where she spent most of her childhood. In 2007, Megan graduated from Avon Lake High School in 2007, and then immediately fled the cold to attend North Carolina State University. After many great experiences including a 5 week study abroad trip in Egypt and a summer internship in Stuttgart, Germany, she graduated Summa Cum Laude with Bachelors of Science in Electrical Engineering and a Minor in Middle East Studies in May 2011. She began research in the field of Systems Biology that summer joining the EnBiSys research group in the Department of Electrical and Computer Engineering.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Williams for his excellent guidance and patience in helping me develop my research. I thank my family and friends for their constant support, encouragement and interest in my research.

TABLE OF CONTENTS

List of Tables	vii
List of Figures	viii
Chapter 1 Introduction	1
1.1 Overview	1
1.2 Motivation	4
1.3 Contributions of this Work	4
1.3.1 Intellectual Merit	4
1.3.2 Broader Impacts	4
1.4 Organization	6
Chapter 2 Background	8
2.1 Modeling Biological Pathways	8
2.1.1 Mass Action Kinetics	8
2.1.2 Michaelis-Menten Kinetics	9
2.1.3 Boolean Modeling	10
2.2 Interval Arithmetic	15
2.3 Estimating Regions of Attraction	16
2.3.1 Lyapunov Functions	16
2.3.2 ROA Estimation for Nonlinear Polynomial Functions	20
2.3.3 ROA Estimation for Nonlinear Non-Polynomial Functions	22
Chapter 3 Methodology	23
3.1 Interval ROA Estimation Method	24
3.2 Two State Model	26
3.3 Three State Model	28
3.4 Expanding the Calculated Estimate	32
Chapter 4 Results	36
4.1 Two State Model	36
4.2 Three State Model	37
4.3 Expanding Calculated Estimate: Preliminary Results	41
4.3.1 Rotate V	41
4.3.2 Alter Shape of V	43
Chapter 5 Conclusion	47
5.1 Conclusion	47
5.2 Future Work	48

5.2.1	Develop Algorithm to Find Best Quadratic Lyapunov Function . .	48
5.2.2	Increase ROA Estimate	48
References		50

LIST OF TABLES

Table 2.1	Boolean State Table	12
Table 4.1	The area of the ROA estimate when the given P matrix is rotated about the equilibrium point	43
Table 4.2	The ROA Estimate with the largest area from Table 4.1 ($\theta = 15^\circ$) and two additional estimates that result from changing the eigenvalues of that Lyapunov function.	46

LIST OF FIGURES

Figure 2.1	Dependence of reaction rate v , $\frac{d[P]}{dt}$, on substrate concentration S in Michaelis-Menten kinetics. $V_{\max}$ denotes the maximal reaction rate that can be reached for a large substrate concentration. K_m is the substrate concentration that results in a half maximal reaction rate. For low substrate concentrations, v increases almost linearly with S , while for high substrate concentrations v is almost independent of S [21].	11
Figure 2.2	Sample Boolean Network: A activates C and B inhibits C, $C=A\&!B$	11
Figure 2.3	Interval extensions $[f]([\mathbf{x}])$ and $[f]^*([\mathbf{x}])$ (minimal) [19]	16
Figure 2.4	Lyapunov ROA estimate example [26]	18
Figure 2.5	This figure illustrates how initial conditions in the $\dot{V}(\mathbf{x}) > 0$ space can be contained in the actual ROA despite their trajectories not being contained in the Lyapunov function.	19
Figure 3.1	Interval Analysis Approach	24
Figure 3.2	ROA Estimation Algorithm	25
Figure 3.3	Schematic depiction of the simplified Cdc2-cyclin B/Wee1 mutually inhibitory system [3]	28
Figure 3.4	The Rb-E2F network [50]: (a) Detailed Rb-E2F signaling network that controls the G1/S transition of mammalian cell cycle. Gray-shaded ovals indicate overlapping or intermediate signaling activities to be lumped. Circled numbers indicate indexes of the regulatory links. (b) A simplified Rb-E2F network. (c) Robust minimal model for resettable bistability.	29
Figure 3.5		33
Figure 3.6	Finding a quadratic Lyapunov function that gives a larger ROA estimate: a) By rotating the initial Lyapunov function (orange), it is possible to find orientations that allow larger level sets of the function to be used as the ROA estimate. The first rotation (purple) shows an estimate that is one level set larger and the second rotation (green) creates an estimate that is two level sets larger than the initial estimate. b) Using the largest ROA estimate from (a), the eigenvalues of the ellipse are scaled to create different shapes. The larger the difference between the eigenvalues (red) creates a smaller estimate, but by decreasing the difference between the eigenvalues (blue), and additional level set of the Lyapunov function is able to be used as the estimate of the ROA, creating a larger estimate. . .	34

Figure 4.1	The simplified Cdc2-cyclin B/Wee1 mutually inhibitory system: (a) Phase portrait of system. Gray curve $\rightarrow \dot{V} = 0$, Shaded gray area $\rightarrow V \leq c^*$. Gray point is the saddle points, and the two black points are the stable points. (b) Several level sets of $V(x)$. a a a	38
Figure 4.2	Trajectories of the system. Solid Black $\rightarrow$ analyzed stable point, Dashed Black $\rightarrow$ other stable pt. Gray dot = saddle point	39
Figure 4.3	The original ROA estimate (blue) from section 4.1 and the largest level sets at three additional orientations of the Lyapunov function around the equilibrium point. The dashed curves mark $\dot{V} = 0$ for each of the associated V curves.	42
Figure 4.4	Three different ROA estimates oriented at $\theta = 15^\circ$. The dashed curves mark $\dot{V} = 0$ for each of the associated V curves.	44
Figure 4.5	The actual ROA of the bistable two state system is the area below the stable manifold (bright green) of the saddle point; compared with the original estimate (blue) and the larger estimate after rotation and scaling (purple). The phase portrait was plotted using PPlane [32].	45
Figure 5.1	Using Free-Form Deformation to Expand an ROA estimate	49

Chapter 1

Introduction

1.1 Overview

Understanding the individual components and the overall dynamic mechanisms associated with biological processes can have promising implications on biofuel production and disease control [5][30]. This understanding cannot be attained through the piecewise qualitative analysis that has traditionally been used to understand biological processes. For this reason, mathematical and computational tools are being developed to quantitatively characterize these systems [20][47]. Biological systems theory is said to be analogous to power and control systems theory. The goal is to apply the analysis and computational tools used to evaluate control systems to models of biological processes. Currently, however, the mechanisms and dynamic interactions in most biological systems remain unknown. The pursuit of effective approaches for modeling the dynamics of biological systems along with computational tools that are capable of characterizing the possible range of model activity remains an open research area.

Methods for modeling biological systems include mass action kinetics, Michaelis-

Menton kinetics [14], S-systems, General Mass Action (GMA) systems, and the power law approach [22]. Of these methods, mass action law and Michaelis-Menten kinetics have been prevalent in the literature [44][33]. These methods, however, require a-priori knowledge of the kinetics associated with the interactions of the system components. This knowledge remains unknown for many biological networks and pathways of practical interest. One approach commonly used to circumvent this lack of knowledge is the use of Boolean models to describe the interactions of components in biochemical processes. Boolean models are discrete models that describe observed dependencies or influences between components, often leaving out detailed kinetics of the interactions [47]. Although these models are a simplification of the process, they have been shown to reproduce the qualitative behavior of a biological process [13][15]. Boolean models in their native form, however, cannot be used to explain or predict the outcome of biological experiments that yield continuous temporal data. This is because they cannot describe continuous concentration levels that change over time. The Hillcube transformation is one method of transforming Boolean models into continuous models that produce continuous dynamics, which can be quantitatively compared to experimental data [47]. This approach has been used to model a variety of biological processes including the spatial gene expression patterns at the murine mid-hindbrain boundary (MHB) to explain the stable maintenance of the MHB [48].

The canonical form of continuous Boolean representations yield a system of equations that gives us the ability to use known computational tools to analyze their characteristics. One such important characteristic of nonlinear systems that is important for biological systems is the region of attraction (ROA) of an equilibrium point. The region of attraction is the set of initial conditions in a system whose trajectories converge to a stable equilibrium point, or steady state. In biological systems, the ROA can be interpreted as

the range of concentrations that will determine if a biological system takes on a certain set of steady state concentrations, often corresponding to a preferred phenotype. Knowledge of the ROA for an equilibrium point could provide knowledge of the sensitivity of a biological system to changes in its environment. It can also be used to determine what perturbations may cause the system to transition to a new phenotype or structure.

For simple 2-dimensional systems, regions of attraction can be easily seen from plots of phase portraits [37]. For higher dimensions, however, it is difficult to impossible to intuitively determine the ROA for a given steady state. For this reason, there has been significant research performed in the last couple decades focusing on methods to quantitatively estimate the ROA for nonlinear systems [18][45]. ROA estimation methods for polynomial models have dominated the literature, with many of these methods using Lyapunov functions to define the region. A common method for determining a Lyapunov function involves linearizing the system and then using Linear Matrix Inequalities (LMIs) to find the coefficients of the functions [7][39]. Various methods have been proposed to find the largest ROA for a given Lyapunov function, which include sums-of-squares programming [41], polytopic regions [27], and neural networks [16]. Chesi and Tibken presented an approach for increasing the accuracy of ROA estimates using the union of several Lyapunov functions [11][40]. Unlike the models used in many of these approaches, many biological models are non-polynomial systems [47]. Few ROA estimation techniques have been proposed for non-polynomial nonlinear systems but uncertainty still remains about the general applicability of these approaches to a broader realm of nonlinear models [8][45]. Thus, it is important to assess whether such approaches can be successfully applied to mathematical structures often used to model biological systems, specifically useful models such as the continuous Boolean approximations.

1.2 Motivation

The goal of this work is to develop and implement ROA estimation techniques for non-polynomial models that are frequently used to model large biological systems. The approach we develop specifically address the conservativeness of current approaches in the literature. Our long term goal is to develop an approach that can be applied to assess the robustness and sustainability of phenotypes in plant systems, and their sustainability under stressful environmental conditions such as climate change.

1.3 Contributions of this Work

1.3.1 Intellectual Merit

In this document, we present results that affirmatively show that the ROA estimation approach presented in [45] can be successfully applied to biochemical systems modeled using continuous Hillcube approximations. We explore the potential limitations of this ROA estimation technique on this class of nonlinear systems and assess the quality of the resulting ROA estimate. We also identify a method that selects from an infinite number of quadratic Lyapunov functions one that could provide an improved estimate of the ROA to decrease the conservativeness of the current ROA estimation algorithms.

1.3.2 Broader Impacts

The research developed in this thesis is applicable to areas of environmental concern ranging from developing cleaner fuel derived from sustainable and renewable resources to international food stability. The ROA estimation techniques described in this document provides the foundation for techniques that will be developed for larger biological

systems, such as biochemical compounds formed by plants. Our goal is to use ROA estimation to determine the robustness of biological systems under environmental stress. This information can then be used to develop methods to combat the changes that are begin caused and to preserve dwindling ecosystems.

Trees as a Sustainable Biofuel Source

On-going research in ROA estimation will provide the mathematical foundation for assessing steady state phenotypes of models of lignin biosynthesis. Through this we will be able to assess the ROA of favorable lignin structures that improve the efficiency of biofuel production for plant biomass. Understanding how changes in the ROA of lignin in *P. trichocarpa* in response to environmental stressors will enhance our ability to use variations of North American pine trees as sustainable sources of biofuel. The genus *Populus* is composed of 25 to 35 species of deciduous plants, including poplar, aspen, and cottonwood trees, that are native to the Northern Hemisphere. Hybrid poplars are among the fastest growing hardwood trees in North America, growing 3-5 feet every year, and can be designed to yield desirable traits for applications such as biofuels production [35][38]. Hybrid poplars, commonly classified as short-rotation woody crops can be grown on forestlands or on economically marginal croplands [35]. The high cellulose and moderate hemicellulose and lignin content in poplar trees make them a desirable source for the production of biofuel. Additionally, poplar wood has lower sulfur content when compared to switchgrass, which is advantageous in terms of the strict environmental regulations limiting the sulfur content of transportation fields [35].

Sustainability of Global Food Sources

ROA estimation can additionally be used to aid in identifying the sustainability of

global food sources by providing the computational framework needed to understand how changes in the environment impact plant systems such as commodity crops. Environmental stress factors such as drought, elevated temperature, salinity and elevated CO_2 concentration affect the robustness of plants and pose an increasing threat to the sustainability of agriculture in the presence of significant global population growth [1]. Plants are evolving to cope with these climate changes, but we are generally unable to predict how well plants are adapting to these challenges. Insight into how plants are adapting are important to breeding programs that produce transgenic varieties of crops with an enhanced tolerance to environmental stress factors.

1.4 Organization

In chapter 2 we discuss a brief background of modeling biochemical systems including the law of mass action, Michaelis-Menten kinetics, and Boolean modeling. We also provide an explanation of some of the mathematical techniques used, including interval arithmetic and linear matrix inequalities. We then provide an outline of the advantages and disadvantages of current approaches for estimating regions of attraction using the Lyapunov stability theorem relates to that. Finally, we describe the interval analysis ROA estimation approach presented by [45] that we utilized in this research.

Chapter 3 sets up the interval analysis ROA estimation technique for a two and three dimensional biochemical pathway modeled using Hillcube continuous Boolean approximations. We then discuss the methodology of a proposed algorithm that would identify a Lyapunov function from the many possible ones to improve the estimate of the ROA.

In Chapter 4 we present and discuss the results of the ROA estimation algorithm for the two and three state examples. We also present our preliminary results for the

proposed technique to determine an alternative quadratic Lyapunov function.

Chapter 5 presents ideas for further research in developing ROA estimation algorithms for the purpose of determining the robustness of plant systems under environmental stress. We then summarize and conclude the work presented in this document.

Chapter 2

Background

In this chapter, we present background information on the techniques used in this document. Section 2.1 presents a brief background on some of the common system modeling techniques used, followed by a basic overview of the mathematical techniques used in section 2.2. Finally, we will go into an overview of regions of attraction and give details on the previous research that has been conducted in this area in section 2.3.

2.1 Modeling Biological Pathways

2.1.1 Mass Action Kinetics

The Law of Mass Action was introduced in the late 19th century by Guldberg and Waage as a method of mathematically describing and predicting reaction rates in biochemical systems. The law states that the rate of change of a substance is proportional to its concentration. Equation (2.1) describes a simple enzyme conversion from substrate to product. This reaction comprises of a reversible formation of an enzyme-substrate complex, ES , from the free enzyme, E , and the substrate, S , and an irreversible release of the

product, P, from the enzyme E. The mass action rate equations are shown in (2.2), where k_i is the respective rate, or kinetic, constant associated with the reactions. These rate equations show how the concentration of each of the substances changes over time, and can be used to predict the outcome of the reaction depending on the initial concentration levels [21].

$$\begin{aligned} \frac{d[S]}{dt} &= k_{-1}[ES] - k_1[E][S] \\ \frac{d[E]}{dt} &= k_{-1}[ES] + k_2[ES] - k_1[E][S] \\ \frac{d[ES]}{dt} &= k_1[E][S] - (k_{-1} + k_2)[ES] \\ \frac{d[P]}{dt} &= k_2[ES] \end{aligned} \quad (2.2)$$

Mass action ODEs can easily become cumbersome and difficult to solve. In the early 20th century Michaelis and Menten found a method of simplifying the mass action kinetics using assumptions of quasi-equilibrium.

2.1.2 Michaelis-Menten Kinetics

Michaelis-Menten kinetics can be used to simplify mass action ODEs, however, several assumptions must be made [21]. For the system of ODEs in (2.2), we must assume that the association and disassociation rates of E and S are much faster than the conversion of ES to P, in other words, $k_1, k_{-1} \gg k_2$. Additionally if we assume that the initial amount of S in the system is much greater than the initial value of E, $[S(t=0)] \gg [E(t=0)]$ then the enzyme-substrate complex, ES, is said to be in *quasi-steady state*, or $\frac{d[ES]}{dt} = 0$.

The system of ODEs in (2.2) are then expressed as:

$$\begin{aligned}\frac{d[S]}{dt} &= -\frac{V_{\max}[S]}{K_m + [S]} \\ \frac{d[P]}{dt} &= \frac{V_{\max}[S]}{K_m + [S]}\end{aligned}\tag{2.3}$$

where,

$$K_m = \frac{k_{-1} + k_2}{k_1} \quad V_{\max} = k_2[E]_{\text{total}} \quad [E]_{\text{total}} = [E] + [ES].$$

$V_{\max}$ represents the maximum rate, and K_m is the Michaelis constant. K_m is the concentration of S that results in a rate v that is equal to half of the maximum rate $V_{\max}$. Figure 2.1 illustrates these relationships. By using Michaelis Menten kinetics, the number of rate equations has been reduced from four, (2.2), to two, (2.3), and the number of parameters has decreased from three to two [21].

Modeling systems using mass action kinetics or Michaelis Menten kinetics can take a long time and requires knowledge of the detailed chemical reactions as shown in (2.1). When these reactions become more complex, as is the case with different levels on inhibitions, the Michaelis-Menten equations are derived again and the functional form of the equation changes. Problems occur, however, when only interactions are known but the detail chemical reactions are unknown. In these cases, researchers began looking at Boolean modeling for a faster way to model biological systems with less *a priori* information.

2.1.3 Boolean Modeling

For cases where only the interactions are known, simpler models have been shown to provide the correct sequence of states [13]. Boolean models are discrete representations

Figure 2.1: Dependence of reaction rate v , $\frac{d[P]}{dt}$, on substrate concentration S in Michaelis-Menten kinetics. $V_{\max}$ denotes the maximal reaction rate that can be reached for a large substrate concentration. K_m is the substrate concentration that results in a half maximal reaction rate. For low substrate concentrations, v increases almost linearly with S , while for high substrate concentrations v is almost independent of S [21].

Figure 2.2: Sample Boolean Network: A activates C and B inhibits C, $C=A\&!B$

with binary states where each gene, protein, etc. is either considered to be on (1) or off (0). These models have been used to predict specific cell sequence patterns of protein and gene activity as observed in living cells with significantly fewer parameters [13].

Figure 2.2 shows a simple Boolean model example. Here, gene A activates gene C and gene B inhibits gene C. Logically, this is represented as

$$C = A \& !B \quad (2.4)$$

where $\&$ is the AND function and $!$ is the NOT function. This means that C is only activated (1) when A is active (1) AND B is inactive (0). Table 2.1 is the state table for this interaction.

Table 2.1: Boolean State Table

A	B	C
0	0	0
0	1	0
1	0	1
1	1	0

Boolean models have been applied to the cell-cycle networks of *S. cerevisiae* [24], the mammalian cell-cycle [15], gene regulatory networks determining embryonic segmentation in *D. melanogaster* [2], and the fission yeast *S. Pombe* cell cycle [13]. In [13], Davidich and Bornholdt compared the Boolean model of *S. Pombe* they created with an existing ODE model of fission yeast [43]. Their Boolean model was able to reproduce the biological sequence of protein activation in accordance with the ODE model, but their

model required the use of 47 fewer kinetic parameters than were necessary for the ODE approach.

The main benefit of Boolean models is that they require less knowledge about the detailed kinetics of the interactions of a system. This allows Boolean models to be created much quicker and easier than their ODE counterparts. However, discrete models are not able to describe continuous concentration levels that change over time. This prevents pure Boolean models from being used to explain or predict the outcome of biological experiments that yield quantitative temporal data. Continuous Boolean modeling was developed to combine the comparative ease of Boolean modeling with the ability to produce continuous dynamics that are often needed to evaluate the stability characteristics of biological systems.

Continuous Boolean Modeling

There are several different transformation techniques for creating continuous Boolean models including BooleCubes, min-max and product-sum fuzzy logic, and HillCubes. BooleCubes are created using multivariate polynomial interpolation and HillCubes are obtained by applying Hill functions to BoolCubes. Due to their smooth analytical properties and the accuracy of their transformation, we explored the continuous Boolean model obtained using the HillCube transformation [47]. HillCube transformations are based on the assumption that molecular interactions typically show a switch-like behavior. This response is modeled using sigmoid shaped Hill functions,

$$f(\bar{x}) = \frac{\bar{x}^n}{\tau(\bar{x}^n + k^n)}. \quad (2.5)$$

Here, k corresponds to the threshold, above which one defines the state as being

'on', and n determines the slope of the curve and is a measure of the cooperativity of the interaction [47]. τ represents the life-time of the species. Continuous Boolean models confine the system dynamics over the range $x \in [0, 1]$ given that the Boolean representations confine the component values to either 0 or 1.

Equation (2.6) uses the HillCube transformation to describe the Boolean model in Figure 2.2.

$$\begin{aligned}\frac{dA}{dt} &= \frac{1}{\tau_a}(-A) \\ \frac{dB}{dt} &= \frac{1}{\tau_b}(-B) \\ \frac{dC}{dt} &= \frac{1}{\tau_c} \left(\frac{A^n}{A^n + k_a^n} \cdot \frac{k_b^n}{B^n + k_b^n} \right)\end{aligned}\tag{2.6}$$

Here we can see that that an AND function in the Boolean model is represented as multiplication, and when one state activates another, the HillCube expression is:

$$\frac{A^n}{A^n + k_a^n}.\tag{2.7}$$

When one state inhibits another the HillCube expression is:

$$1 - \frac{B^n}{B^n + k_b^n} = \frac{k_b^n}{B^n + k_b^n}.\tag{2.8}$$

2.2 Interval Arithmetic

Interval arithmetic is commonly used as an approach for bounding measurements and rounding errors in mathematical computation. Typically, arithmetic is used to solve for exact single point solutions. Interval arithmetic, however, computes the upper and lower bounds of a solution. It can provide guaranteed solutions by providing solution sets that include any variation that could be associated with finite arithmetic and round off error. Interval arithmetic is also valuable in the physical and life sciences due to limited accuracy in experimental data measurement [29]. Interval extensions of functions are used when applying interval analysis to traditional arithmetic equations. The interval extension $[f]([\mathbf{x}])$ of a function $f(x)$ bounds the range of the function $f(x)$ over the range $[\mathbf{x}] = [\underline{x}, \overline{x}]$, where $\underline{x}$ is the minimum value of the interval and $\overline{x}$ is the maximum value of the interval. The interval extension of a function provides a result that guarantees that the true solution is contained in the range. In other words, it guarantees that the interval $[f]([\mathbf{x}])$ contains all values of $f(\cdot)$ evaluated at all points over the interval $[\mathbf{x}]$. Figure 2.3 illustrates the interval extension of a function. Steps are often taken to ensure that $[f]([\mathbf{x}])$ is as close to the true solution set, with as little overestimation as possible [28]. For n -dimensional systems, these intervals, $[\mathbf{x}] = [\underline{x}_1, \overline{x}_1] \times \dots \times [\underline{x}_n, \overline{x}_n]$, form hypercubes where the minimum and maximum values of the intervals, $(\underline{x}_i, \overline{x}_i)$, form the vertices of the hypercube.

Interval arithmetic has been used to achieve numerical certification of the kinematic calibration of a parallel robots [12], reliably calculate phase stability and equilibrium in industrial processes [36], global optimization of molecular structures [25], to verify the existence of the Lorenz attractor [42], and to estimate regions of attraction [45].

The degree of satisfaction involved in enclosing a solution can depend strongly on the

Figure 2.3: Interval extensions $[f]([x])$ and $[f]^*([x])$ (minimal) [19]

tightness of the enclosure obtained. For example, if the bounded solution is

$$x \in [0.66666, 0.66667],$$

then the solution, x , is known to 4 decimal places [29].

2.3 Estimating Regions of Attraction

2.3.1 Lyapunov Functions

Lyapunov functions are frequently used to estimate ROAs. Lyapunov functions are used to define a region around a stable equilibrium point where a trajectory will stay in that region once entering it. The following Lyapunov Stability Theorem states how knowledge of this region can be used to provide an estimate of the ROA.

Lyapunov's Stability Theorem:

For a system containing n states, the closed set $\Omega_c \subseteq \text{ROA}$, $\Omega_c \in \mathbb{R}^{n \times 1}$, is an estimate of the ROA of a stable equilibrium point if there exists a Lyapunov function $V(x)$ such

that:

1. $V(\mathbf{x})$ is positive definite on Ω_c
2. $\dot{V}(\mathbf{x})$ is negative definite $\forall \mathbf{x} \neq 0$ on Ω_c and,
3. $\dot{V}(\mathbf{x}) = 0$ when $\mathbf{x} = 0$

The bounded ROA estimate is then:

$$\Omega_c := \{\mathbf{x} \mid V(\mathbf{x}) < c\}. \quad (2.9)$$

This means that once any trajectory enter the region Ω_c that trajectory stays inside the region.

Figure 2.4 illustrates how some level sets of a Lyapunov function are considered estimates of the ROA and others are not. In this image, Ω_1 and Ω_2 are both ROA estimates. Ω_2 , however, is the better estimate since it is the larger of the two. The largest level set that can be considered and estimate of the ROA is the set that is tangent to the curve or surface $\dot{V}(\mathbf{x}) = 0$ but remains entirely in the space $\dot{V}(\mathbf{x}) < 0$. Ω_3 is not an estimate of the ROA, because even though it is tangent to $\dot{V}(\mathbf{x}) = 0$ at one location, it crosses into the space $\dot{V}(\mathbf{x}) > 0$. The trajectories of the initial conditions that lie in the space $\dot{V}(\mathbf{x}) > 0$ exit region Ω_3 . In order for a region to be considered Lyapunov stable no trajectory can leave the region. Therefore, Ω_3 is not a valid estimate of the ROA.

It should be noted that just because an initial condition lies in the $\dot{V}(\mathbf{x}) > 0$ space does not imply it is not part of the ROA. It is not uncommon for there to be trajectories that are part of the ROA but are not contained in a quadratic Lyapunov function. It is impossible, however, to analytically determine which of the initial conditions inside the

Figure 2.4: Lyapunov ROA estimate example [26]

$\dot{V}(\mathbf{x}) > 0$ space eventually do converge to the equilibrium point and which initial conditions have the trajectories that go elsewhere. Figure 2.5 illustrates this characteristic.

Figure 2.5: This figure illustrates how initial conditions in the $\dot{V}(\mathbf{x}) > 0$ space can be contained in the actual ROA despite their trajectories not being contained in the Lyapunov function.

Most nonlinear systems, including most models of biological processes, can be expressed as a system of nonlinear differential equations

$$\dot{\mathbf{x}} = f(\mathbf{x}) = A\mathbf{x} + F(\mathbf{x}) \quad (2.10)$$

where A is the linearization around the equilibrium point and $F(\mathbf{x})$ is a nonlinear function representing the higher order nonlinearities of the process. Lyapunov functions are polynomial functions of order ≥ 2 . Generally, ROA estimation techniques use the quadratic form of the Lyapunov function which defines an ellipsoidal region around the equilib-

rium point. While higher order Lyapunov functions potentially provide more accurate estimates, they are significantly more difficult to compute [11]. A quadratic Lyapunov function is defined as:

$$V(\mathbf{x}) = \mathbf{x}^T P \mathbf{x}. \quad (2.11)$$

The time derivative of the Lyapunov function for the nonlinear system in (2.11) is calculated as

$$\dot{V}(\mathbf{x}) = \mathbf{x}^T (A^T P + P A) \mathbf{x} + 2\mathbf{x}^T P F(\mathbf{x}). \quad (2.12)$$

The Lyapunov matrix P can be found using computational programs such as MATLAB [17] to solve the following linear constraints:

$$P = P^T > 0 \quad A^T P + P A < 0. \quad (2.13)$$

2.3.2 ROA Estimation for Nonlinear Polynomial Functions

In this section we describe several different ROA estimation techniques for polynomial systems that have been presented in the literature.

Research in region of attraction estimation became popular in the mid-80's. Gene-sio, Tartaglia, and Vicino's outlined the, at that time, state of the art and new ROA estimation proposals in [18]. The methods published in this paper formed the basis for todays state of the art and new ROA estimation techniques. [18] splits the ROA estimation methods into two main groups, Lyapunov and non-Lyapunov methods. Since the mid-80's, the majority of research in ROA estimation has made use of Lyapunov methods. It became especially prevalent in the late-90s/early 2000's, when it was determined that linear matrix inequalities could be used to solve for Lyapunov functions. Below is a collection of ROA estimation techniques that have been presented in the last decade.

Graziano Chesi has written several articles on estimating regions of attraction of non-linear systems using Lyapunov functions and linear matrix inequalities (LMIs) [7][6][11][8][9][10]. In [11], Chesi proposes using the union of continuous families of Lyapunov estimates to approximate the region of attraction. By calculating a parameter-dependent lower bound using an Eigenvalue problem, a family of estimates of the ROA and the estimate given by their union can be defined. The advantage of the proposed approach is to provide a continuous family of estimates rather than one estimate as in existing methods. A disadvantage, however, is the larger computational burden in the LMI optimizations which currently limit the use of the proposed method to low order systems and low degree Lyapunov functions. Additionally, this method is limited to continuous-time polynomial systems.

Valmorbida, Tarbouriech, and Garcia propose a method that calculates an ROA estimate for nonlinear polynomial systems using quadratic Lyapunov functions and stability analysis conditions in a “quasi”-LMI form. They also use an LMI-based optimization problem to compute the Lyapunov matrix “maximizing” the estimate of the ROA. This optimization problem, however, does not take into consideration the different level sets of the Lyapunov function, and therefore, larger estimates of the ROA can be calculated by considering additional level sets.

She and Ratschan presented a method that uses an interval based branch and bound algorithm to compute a Lyapunov-like polynomial that defines an ROA estimate [34]. This method allows a Lyapunov-like function to be calculated depending on the system dynamics, which is likely to provide a larger ROA estimate than using LMI constraints to solve for a feasible Lyapunov function. However, this method, requires that a ‘target area’ that is known to be inside the ROA is known before the algorithm can be implemented.

Estimating the ROA using interval analysis [23] is discussed by Lhommeau, Jaulin,

and Hardouin. This method computes the inner and outer boundaries of the ROA by splitting the system into several rectangular intervals that are then either determined to be inside or outside the ROA. An advantage of the interval analysis method is that no special assumptions of the dynamical system are needed and the algorithm provides guaranteed results, however, it is computationally intensive.

2.3.3 ROA Estimation for Nonlinear Non-Polynomial Functions

Chesi has also developed a method to estimate ROAs for non-polynomial systems [8]. This technique similarly uses Lyapunov functions and linear matrix inequalities to estimate the ROA, however, he expresses the non-polynomial terms using a truncated version of their Taylor expansion.

The ROA estimation method presented by Warthenpfuhl, Tibken, and Mayer [45] uses interval arithmetic in conjunction with Lyapunov functions to define a region that estimates of the ROA. This method differs from those presented in Section 2.3.2 due to the fact that it applies to both polynomial and non-polynomial systems. This method will be discussed in further detail in the next chapter.

Despite multiple biological models that are represented as non-polynomial equations, such as Michaelis-Menten kinetics and continuous Boolean approximations, there has been far less research in ROA estimation for non-polynomial systems. For this reason it is necessary to develop ROA estimation techniques that work for these types of models.

Chapter 3

Methodology

In section 3.1, we describe the ROA estimation method presented by Warthenpfuhl, Tibken, and Mayer in [45]. We then apply this method to two biological systems modeled using Hillcube continuous Boolean approximations. The first model is a simplification of the Cdc2-cyclin B complex and Wee1 protein interaction (Figure 3.3) [3]. This simple two state example will be used to illustrate the properties and restrictions of using Lyapunov functions to estimate the ROA. The ROA estimation approach is then applied to a three state model of the Rb-E2F signaling network (Figure 3.4b) [50], highlighting the applicability of this method on larger systems. We then looked at developing an algorithm to find the best quadratic estimate of the ROA.

3.1 Interval ROA Estimation Method

In [45], Warthenpfuhl, Tibken, and Mayer use the interval extensions of a given Lyapunov function $V(\mathbf{x})$ and its time derivative $\dot{V}(\mathbf{x})$ to solve the optimization problem:

$$c^* = \min_{\dot{V}(x)=0, x \neq 0} V(x). \quad (3.1)$$

Figure 3.1: Interval Analysis Approach

Equation (3.1) solves for the maximum level set of the Lyapunov function, c , in (2.9) that meets Lyapunov's stability criteria. The maximum level set that is guaranteed to be Lyapunov stable corresponds to the smallest level set that intersects or is tangent to the surface $\dot{V}(\mathbf{x}) = 0$. For example, in Figure 2.4, both Ω_2 and Ω_3 touch the surface $\dot{V}(\mathbf{x}) = 0$, but Ω_3 , the larger of the two, does not meet the Lyapunov stability criteria since it crosses into the $\dot{V}(\mathbf{x}) > 0$ region. This is why the case of solving for the maximum level set is a

```

1: Initialize List  $L = \{(\underline{V}, x_I)\}$ 
2: Initialize  $\bar{c} = \infty$  and  $\underline{c} = c_{init} > 0$ 
3: while  $\neq$  termination criteria do
4:   Remove first tuple  $(\underline{V}, x)$  from  $L$ 
5:   Bisect  $x \rightarrow y_1, y_2$ 
6:   for  $y_1, y_2$  do
7:      $V(y) = [\underline{V}, \bar{V}]$ 
8:     if  $\underline{V} \geq \bar{c}$  then
9:       Reject  $y$ 
10:    end if
11:    if  $\bar{V} \leq \underline{c}$  then
12:      Reject  $y$ 
13:    end if
14:     $\dot{V}(y) = [\underline{\dot{V}}, \bar{\dot{V}}]$ 
15:    if  $\underline{\dot{V}} > 0$  then
16:      if  $V_m = V(y_{midpt}) < \bar{c}$  then
17:         $\bar{c} = V_m$ 
18:      end if
19:      Reject  $y$ 
20:    end if
21:    if  $\bar{\dot{V}} < 0$  then
22:      Reject  $y$ 
23:    end if
24:    if  $0 \in \dot{V}(y) = [\underline{\dot{V}}, \bar{\dot{V}}]$  then
25:      if  $\dot{V}_m = \dot{V}(y_{midpt}) > 0$  AND  $V_m < \bar{c}$  then
26:         $\bar{c} = V_m$ 
27:      end if
28:      Return tuple  $(\underline{V}, y)$  to list  $L$ 
29:    end if
30:  end for
31:  Sort list  $L$  so that  $\underline{V}_1 < \underline{V}_2 < \dots < \underline{V}_n$ 
32:  if  $\underline{V}_1 > \underline{c}$  then
33:     $\underline{c} = \underline{V}_1$ 
34:  end if
35: end while

```

Figure 3.2: ROA Estimation Algorithm

minimization problem. This level set defines the c that forms the outer boundary of the resulting ROA estimate.

To solve for c^* , from (3.1), the following algorithm, as seen in [45] is implemented:

Initialization: Add the box $[0,1] \times [0,1] \times \dots \times [0,1]$ and the interval extension of $\underline{V}(\mathbf{x})$ to a list. Initialize the bounds on $c = [\underline{c}, \bar{c}]$ so that $\bar{c} = \infty$ and $\underline{c} = c_{init}$ where c_{init} is some value greater than zero.

Step One: Remove the first box from the list and bisect it. For the two resulting boxes, calculate the interval extensions of the $V(\mathbf{x})$ and $\dot{V}(\mathbf{x})$ functions.

Step Two: If the interval extension of $\dot{V}(\mathbf{x}) = [\underline{\dot{V}}, \bar{\dot{V}}]$ contains $\dot{V}(\mathbf{x}) = 0$, i.e. $\underline{\dot{V}} < 0$ and $\bar{\dot{V}} > 0$, then the interval box and the associated $\underline{V}$ is returned to the list. Otherwise the interval will be discarded.

Step Three: Solve for a new upper bound of c . For intervals that are being re-added to the list, if $\dot{V}_m > 0$ AND $V_m < \bar{c}$, where $\dot{V}_m$ and V_m are the values of $\dot{V}$ and V evaluated at the midpoint of the interval, then $\bar{c} = V_m$. If the interval is being rejected and if $\underline{\dot{V}} > 0$ AND $V_m < \bar{c}$ then set $\bar{c} = V_m$. Otherwise $\bar{c}$ does not change.

Step Four: Once all intervals have either been added to the list or discarded, the list is sorted so that the interval with the smallest $\underline{V}$ is first and the interval with the largest $\underline{V}$ is last. If $\underline{V}$ of the first interval in the list is less than $\underline{c}$, then $\underline{c}$ is reset to that value.

Step Five: Repeat steps one through 4 until a user-defined termination criteria, e.g. $\bar{c} - \underline{c} < \epsilon$, is fulfilled.

These steps are further outlined in Figure 3.2.

3.2 Two State Model

The two state model that we examine is a simplification of the Cdc2-cyclin B complex

and Wee1 protein interaction, which is a key interaction in cell cycle regulation. Figure 3.3 shows the simplified mutually inhibitive Cdc2-cyclin B and Wee1 model. This model was chosen because it exhibits bistability characteristics under certain parameters. This is useful for our purpose since there will be a minimum of two biologically relevant stable equilibrium points. This is useful to know, because it allows us to examine the ROA estimation algorithm for an equilibrium point that is not globally stable.

The Cdc2-cyclin B complex and Wee1 protein interaction from Figure 3.3 can be described in continuous Boolean form as shown in (3.2) [3]. Here, x is Cdc2-cyclin B complex and y is Wee1 protein. Here, $\tau_x = 1$, $\tau_y = 10$, $k_x = 0.1$, and $k_y = 0.3$. The calculated Lyapunov function associated with the equilibrium point at $[1 \ 0]^T$ is shown in (3.3). The P matrix associated with (3.3) was found using the Matlab LMI Control Toolbox [17].

$$\begin{aligned}\dot{x} &= \frac{1}{\tau_x} \left(\frac{k_y^4}{y^4 + k_y^4} - x \right) \\ \dot{y} &= \frac{1}{\tau_y} \left(\frac{k_x^4}{x^4 + k_x^4} - y \right)\end{aligned}\tag{3.2}$$

$$V(\mathbf{x}) = \mathbf{x}^T P \mathbf{x} = 0.9315x^2 + 1.7914y^2 - 0.0000825xy.\tag{3.3}$$

The corresponding $\dot{V}(\mathbf{x})$ is:

$$\dot{V}(\mathbf{x}) = -1.863x^2 - 0.358y^2 + 0.00009xy - 0.000008x + 0.3583y\tag{3.4}$$

$$+ B_1(-1.863x + 0.00008y) + B_2(0.000008x - 0.358y)\tag{3.5}$$

where,

$$B_1 = \frac{y^4}{y^4 + k_y^4}$$

$$B_2 = \frac{x^4}{x^4 + k_x^4}.$$

This Lyapunov function and its time derivative were used to implement the interval ROA estimation procedure shown in Figure 3.2 using the Boost C++ Interval Library [4]. Chapter 4 Section 1 describes the results obtained using this method.

Figure 3.3: Schematic depiction of the simplified Cdc2-cyclin B/Wee1 mutually inhibitory system [3]

3.3 Three State Model

The 3 state Hillcube model used in this work comes from a subnetwork of the Rb-E2F signaling pathway shown in Figure 3.4b. The Rb-E2F pathway regulates the initiation of DNA replication and plays a critical role in cell proliferation by acting as a bistable switch that converts graded growth signals into an all-or-none activation of E2F [50]. This pathway has specific importance to the biomedical sciences, since it has been found

that this pathway is deregulated in almost all cancer cells [31][46]. Yao et. al. identified a minimal circuit that generates robust, resettable bistability. Experimental disruption of this circuit abolishes maintenance of the activated E2F state, supporting its importance for the bistability of the Rb-E2F system [50]. In this work, the ROA estimation technique outlined in Figure 3.2 is applied to this minimal circuit of the Rb-E2F pathway, shown in Figure 3.4c. Hillcube approximations are used to describe this network and are shown in (3.6).

$$\begin{aligned}
\frac{d[\text{MD}]}{dt} &= \frac{1}{\tau_m} \left[\left(\frac{[\text{S}]^{n_1}}{[\text{S}]^{n_1} + k_s^{n_1}} \right) - [\text{MD}] \right] \\
\frac{d[\text{RP}]}{dt} &= \frac{1}{\tau_r} \left[\frac{k_{m1}^{n_2}}{[\text{MD}]^{n_2} + k_{m1}^{n_2}} \cdot \frac{k_e^{n_3}}{[\text{EE}]^{n_3} + k_e^{n_3}} - [\text{RP}] \right] \\
\frac{d[\text{EE}]}{dt} &= \frac{1}{\tau_e} \left[\frac{[\text{MD}]^{n_4}}{[\text{MD}]^{n_4} + k_{m2}^{n_4}} \cdot \frac{k_r^{n_5}}{[\text{RP}]^{n_5} + k_r^{n_5}} - [\text{EE}] \right].
\end{aligned} \tag{3.6}$$

Here, $\tau_m = \tau_r = \tau_e = 1$, $k_s = 0.8$, $k_{m1} = 0.7$, $k_{m2} = 0.02$, $k_e = 0.3$, $k_r = 0.03$, the input, S , is fixed to 1, and $n_i = n = 4 \forall 1 \leq i \leq 5$. The system above has three equilibrium points, two stable nodes, and a saddle point. Without loss of generality we examine the ROA of the stable node $\begin{bmatrix} x_e & y_e & z_e \end{bmatrix}^T = \begin{bmatrix} 0.7094 & 0.4866 & 0.000014 \end{bmatrix}^T$. The nature of the Lyapunov function requires that the stable equilibrium occurs at the origin, so it is necessary to translate the system so that the stable node is at the origin. For the remainder of this manuscript, $[\text{MD}]$ will be expressed as x , $[\text{RP}]$ will be expressed as y ,

and [EE] will be expressed as z . The translated system becomes:

$$\begin{aligned}\dot{x} &= \frac{1}{1 + k_s^n} - x \\ \dot{y} &= \frac{k_{m1}^n}{x^n + k_{m1}^n} \cdot \frac{k_e^n}{z^n + k_e^n} - y \\ \dot{z} &= \frac{x^n}{x^n + k_{m2}^n} \cdot \frac{k_r^n}{y^n + k_r^n} - z.\end{aligned}\tag{3.7}$$

For the system shown in (3.7), the following Lyapunov function satisfying (2.13) was calculated using the LMI solver in Matlab's Control Toolbox [17]:

$$\begin{aligned}V(\mathbf{x}) &= 57.85x^2 + 32.25y^2 + 36.81z^2 + 37.32xy \\ &\quad + 0.0014xz + 0.0032yz\end{aligned}\tag{3.8}$$

The time derivative of the Lyapunov function (3.8) is calculated as:

$$\begin{aligned}\dot{V}(\mathbf{x}) &= -115.69x^2 - 64.50y^2 - 73.63z^2 - 74.66xy \\ &\quad - 0.0028xz - 0.0064yz - 18.16x - 31.385y \\ &\quad - 0.0026z + B_1(37.33x + 64.50y + 0.0032z) \\ &\quad + B_2(0.0014x + 0.0032y + 73.63z)\end{aligned}\tag{3.9}$$

where,

$$B_1 = \frac{k_{m1}^n}{x^n + k_{m1}^n} \cdot \frac{k_e^n}{z^n + k_e^n}$$

$$B_2 = \frac{x^n}{x^n + k_{m2}^n} \cdot \frac{k_r^n}{y^n + k_r^n}.$$

This Lyapunov function and its time derivative were used to implement the interval ROA estimation procedure shown in Figure 3.2 using the Boost C++ Interval Library [4]. The results of this algorithm are described in Chapter 4 Section 2.

3.4 Expanding the Calculated Estimate

The available LMI solvers calculate a P matrix that meets the requirements of the LMI constraints, (2.13). The resulting matrix is always feasible, but is not always the “best” Lyapunov function for providing the largest estimate of the ROA. There exists a potentially infinite number of Lyapunov functions that can satisfy the necessary LMI constraints. Each of these quadratic Lyapunov functions represent an ellipse on the phase plane. The underlying flow of the system, however, makes some Lyapunov functions more appropriate than others in terms of identifying the largest possible estimate of the ROA. In our approach we use mechanisms for controlling the elongation and rotation of the corresponding ellipse to identify a larger estimate of the ROA.

This algorithm will first look at rotating the given P matrix to find the orientation that gives the largest estimate of the ROA for that ellipsoid. Then, by adjusting the eigenvalues of the rotated P matrix, the algorithm will change the shape of the ellipsoid to find the one that works best with the system dynamics. The LMI constraints in (2.13) will be used to verify each new P matrix meets the Lyapunov stability requirements. For

Figure 3.5

systems whose trajectories approach the equilibrium point in a circular fashion, a more circular Lyapunov region would provide a larger estimate than an elongated ellipse. For systems like the two state example in section 3.2 represented by (3.2), a longer ellipse would have a larger estimate of the ROA than a circular region. For each change in the ellipsoid, the interval analysis ROA estimation algorithm, Figure 3.2, is re-run to find the largest level set.

To determine how different orientations and shapes of an ellipse can change the size of the ROA estimate, we looked at the bistable 2 dimensional system presented in section 3.2. Using the P matrix returned by the LMI solver, we evaluated several different orientations, including $\theta = 15^\circ$, 30° , and 45° , where θ increases in the clockwise direction from the x-axis as shown in Figure 3.5.

The P matrix is rotated in the clockwise direction using the following formula:

$$P_\theta = R_\theta P R_\theta^T \quad \text{where, } R_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \quad (3.10)$$

Figure 3.6: Finding a quadratic Lyapunov function that gives a larger ROA estimate: a) By rotating the initial Lyapunov function (orange), it is possible to find orientations that allow larger level sets of the function to be used as the ROA estimate. The first rotation (purple) shows an estimate that is one level set larger and the second rotation (green) creates an estimate that is two level sets larger than the initial estimate. b) Using the largest ROA estimate from (a), the eigenvalues of the ellipse are scaled to create different shapes. The larger the difference between the eigenvalues (red) creates a smaller estimate, but by decreasing the difference between the eigenvalues (blue), and additional level set of the Lyapunov function is able to be used as the estimate of the ROA, creating a larger estimate.

Once the original ellipse is rotated θ° , V_θ is used to calculate a new $\dot{V}_\theta(\mathbf{x})$. These functions are then used to calculate the largest level set c_θ^* for that orientation. Figure 3.6a shows three different orientations of the same Lyapunov function. The purple and green ellipses are rotations of the initial ellipse (orange). Figure 3.6a shows that both of these rotations allow for larger level sets of the ellipse to be used as the ROA estimate.

The ratio of the eigenvalues of the V_θ that gives the largest ROA estimate are adjusted to vary the shape of the ellipse. The eigenvalues determine the size of the major and minor axes of the ellipse. Since we are looking at different level sets of the ellipses, the individual values of the eigenvalues are not important. Instead, the ratio between eigenvalues is what we are interested in, as this determines whether the ellipse is circular or long and skinny. Figure 3.6b shows an initial ellipse (green) is the same ellipse which gave the largest ROA estimate from Figure 3.6a. The eigenvalues of this ellipse are adjusted to produce a skinnier ellipse (red) and a more circular ellipse (blue). Figure 3.6b shows that the largest ROA estimate from 3.6a can further be increased by changing the shape of the ellipse.

Chapter 4

Results

In this chapter we present and discuss the results of the interval analysis ROA estimation algorithm on the two and three state biological models presented in Chapter 3. We then look at our preliminary findings from the discussed method of determining better Lyapunov functions in Chapter 3 Section 3. The plots shown in Figures 4.1 -4.4 were produced using Mathematica [49].

4.1 Two State Model

Figure 4.1a shows the phase portrait of the two state system described in (3.2) with sample trajectories. The solid gray line represents the boundary $\dot{V}(\mathbf{x}) = 0$ and the solid black curve (without arrows) defines the boundary, $c = [0.081, 0.082]$, of the ROA estimation within the shifted Boolean space. The gray point that falls on the curve $\dot{V}(\mathbf{x}) = 0$ is a saddle point, and the two black points are stable nodes. The solid black trajectories converge to the stable point $[1 \ 0]^T$ inside the ROA estimate. The dashed trajectories converge to the other stable point at approximately $[0 \ 1]^T$. As seen in Figure 4.1a, there

are several initial conditions that reside in the true ROA but are not included in the calculated estimate. This is a characteristic of Lyapunov based ROA estimates. This conservativeness of the estimate is due, in part, to the requirement that once a trajectory enters a Lyapunov function, it must remain inside the Lyapunov function. It is not uncommon, however, for trajectories to leave and then re-enter the Lyapunov region, but it is impossible to analytically determine which trajectories re-enter the Lyapunov region, and which trajectories are not elements of the actual ROA. Figure 4.1b illustrates this. Two initial conditions are chosen close together and both of the trajectories leave the level set $V(\mathbf{x}) \leq c = 1.1$. One of the trajectories re-enters the region and converges to the desired stable equilibrium point. The other trajectory, however, converges to the second equilibrium point. For this reason, it is necessary to constrain the ROA estimate to the level set that is tangent to line $\dot{V}(\mathbf{x}) = 0$.

4.2 Three State Model

The ROA estimation approach shown in Figure 3.2 returned the following bounds for c^* ,

$$c = [\underline{c}, \bar{c}] = [6.2828, 6.2832]$$

Figure 4.2 shows sample trajectories of system (3.7). The solid black curves are trajectories that converge to the analyzed stable point, and the dashed black curves are trajectories that converge to the second stable point. As seen in this figure, there are initial conditions outside of our estimated ROA that still converge to the equilibrium point. Clearly, the estimated ROA is only a subset of the actual ROA, in the same manner as the 2 state example. Additionally, as was seen in the 2 state example, Figure 4.2 also shows

Figure 4.1: The simplified Cdc2-cyclin B/Wee1 mutually inhibitory system: (a) Phase portrait of system. Gray curve $\rightarrow \dot{V} = 0$, Shaded gray area $\rightarrow V \leq c^*$. Gray point is the saddle points, and the two black points are the stable points. (b) Several level sets of $V(x)$.

Figure 4.2: Trajectories of the system. Solid Black→analyzed stable point, Dashed Black→other stable pt. Gray dot=saddle point

a trajectory that starts in the $\dot{V}(\mathbf{x}) > 0$ region but eventually converges to the selected equilibrium point. Since the stability characteristics of the ROA estimate in the 3 state example are congruent to the characteristics discussed in the 2 state example, it can be extrapolated that these characteristics will remain regardless the size of the system, and therefore, this method is applicable to not only to 2 and 3 dimensional systems, but also higher dimensional systems whose phase plane cannot be visualized.

Knowing the ROA, or at least an estimate of the ROA is an important stability characteristic in nonlinear dynamics. The ROA characterizes all of the initial conditions that converge to a specified equilibrium point. An estimate of the ROA provides a group of initial conditions whose trajectories are guaranteed to converge to the desired stable point. This is important information when a particular stable characteristic is desired. For the Rb-E2F biochemical pathway, it may be of interest to 'manually' turn the E2F 'On' ($[EE] \approx 1$) or 'Off' ($[EE] \approx 0$). This could be achieved by increasing or decreasing the concentration levels of the components until they reside in the known estimated ROA. This would guarantee that the desired steady state of the system is achieved.

The applicability of this approach to higher dimensional systems is an important quality when considering analysis techniques for biological systems, since the majority of biological systems are large and have many states. The examples in this paper were simplifications of larger biological models, and were reduced to 2 and 3 states respectively for visualization purposes. The Lyapunov function shown in (2.11), however, can be defined for $n > 3$ dimensions. This would allow us to obtain the corresponding level set, c , for the ROA estimate in larger dimensions. Parallelization methods would decrease the computation time of the interval approach to such larger dimensions [26].

The better the estimate of the ROA, the better the understanding of which initial conditions will converge to each equilibrium point. For this reason, further techniques

are needed to increase the estimate of ROAs associated with non-polynomial systems. Since the shape of the surface $\dot{V}(\mathbf{x}) = 0$ is dependent on the specific Lyapunov function, one potential method for increasing an ROA estimate is to take the union of several Lyapunov ROA estimates to increase the total ROA estimate of an equilibrium point. Qualitatively similar techniques have been proposed for polynomial systems but have not been applied to non polynomial nonlinear systems [11][40].

4.3 Expanding Calculated Estimate: Preliminary Results

Below are the preliminary results found when looking at rotating and changing the size and shape of the P matrix given by the LMI solver.

4.3.1 Rotate V

We took the calculated Lyapunov function from section 4.1 and rotated it around the equilibrium point calculating the largest level sets for the different orientations. Figure 4.3 shows the initial estimate calculated in section 4.1 and the largest level sets of that Lyapunov function rotated 15° , 30° , and 45° from the x-axis. As seen in this figure, when the Lyapunov function is rotated 15° from its original position, it covers a larger area of the system space than the other orientations. This information is tabulated in Table 4.1, where it is shown that the area of the 15° rotation is 0.0506 compared to the original area of the estimate, 0.0413, a 22.5% increase.

Figure 4.3: The original ROA estimate (blue) from section 4.1 and the largest level sets at three additional orientations of the Lyapunov function around the equilibrium point. The dashed curves mark $\dot{V} = 0$ for each of the associated V curves.

Table 4.1: The area of the ROA estimate when the given P matrix is rotated about the equilibrium point

θ	c^*	Area	P
0°	0.081	0.0413	$\begin{bmatrix} 0.9315 & -4.1277 \times 10^{-5} \\ -4.1277 \times 10^{-5} & 1.7914 \end{bmatrix}$
15°	0.088	0.0506	$\begin{bmatrix} \mathbf{0.9891} & \mathbf{0.2149} \\ \mathbf{0.2149} & \mathbf{1.7338} \end{bmatrix}$
30°	0.075	0.0492	$\begin{bmatrix} 1.1464 & 0.3723 \\ 0.3723 & 1.5763 \end{bmatrix}$
45°	0.0565	0.0414	$\begin{bmatrix} 1.3614 & 0.4299 \\ 0.4299 & 1.3615 \end{bmatrix}$

4.3.2 Alter Shape of V

Next we looked at scaling the eigenvalues of V_{15° to see how changing the shape of the ellipse can change to size of the estimate. We scaled the eigenvalues to create an ellipse that was narrower than the original and one that was more circular. Figure 4.4 shows the three ellipses, where the largest estimate from section 4.3.1, V_{15° , is shown in red. As can be seen, the narrower ellipse, in purple, has a significant increase of area than the others. Table 4.2 lists the specifications of each of the three regions. The largest estimate has an area of 0.1030, which is an increase of 147% from the original estimate found in section 4.1. Figure 4.5 shows the boundary of the actual ROA along with the original estimate (blue) and the larger estimate (purple) found through rotating and scaling the dimensions of the ellipse. In this figure, the actual ROA is the area below the stable manifold (light green) of the saddle point.

Figure 4.4: Three different ROA estimates oriented at $\theta = 15^\circ$. The dashed curves mark $\dot{V} = 0$ for each of the associated V curves.

Figure 4.5: The actual ROA of the bistable two state system is the area below the stable manifold (bright green) of the saddle point; compared with the original estimate (blue) and the larger estimate after rotation and scaling (purple). The phase portrait was plotted using PPlane [32].

Table 4.2: The ROA Estimate with the largest area from Table 4.1 ($\theta = 15^\circ$) and two additional estimates that result from changing the eigenvalues of that Lyapunov function.

Eigenvalues of Ellipse (λ)	c^*	Area	P
$\lambda_1 = 0.9315, \lambda_2 = 1.7914$	0.088	0.0506	$\begin{bmatrix} 0.9891 & 0.2149 \\ 0.2149 & 1.7914 \end{bmatrix}$
$\lambda_1 = \mathbf{2.045}, \lambda_2 = \mathbf{8.2240}$	0.63	0.1030	$\begin{bmatrix} \mathbf{2.4588} & \mathbf{1.5445} \\ \mathbf{1.5445} & \mathbf{7.8102} \end{bmatrix}$
$\lambda_1 = 3.2, \lambda_2 = 5.1$	0.23	0.0431	$\begin{bmatrix} 3.3272 & 0.4749 \\ 0.4749 & 4.9728 \end{bmatrix}$

These preliminary results show that depending on the Lyapunov function that the LMI solver returns, the ROA estimate can vary in size. The larger the estimate that can be found, the better and therefore, we conclude it is worthwhile to proceed with this research and develop an algorithm that will determine better Lyapunov functions for a given system.

Chapter 5

Conclusion

5.1 Conclusion

This thesis addressed the issue of whether the ROA estimation algorithm outlined in Figure 3.2 [45] is suitable for biochemical systems modeled using continuous Boolean Hillcube approximations. We applied the method to two models, a two state Cdc2-cyclin B complex and Wee1 protein interaction and a three dimensional model of the Rb-E2F pathway. We showed that the interval analysis ROA estimation technique provided an accurate estimate of the ROA for both systems. In both cases, the estimated ROA is on par with accuracy of other ROA estimates in the literature that use Lyapunov stability. We conclude that this method is a viable technique for estimating the ROA of biological systems modeled with continuous Boolean Hillcube approximations.

We also began looking at how different Lyapunov functions can affect the size of the estimated ROA. We showed that there can be a significant improvement in the size of the ROA estimate by rotating and changing the shape of the ellipse given by the LMI solver. Our future research will look at developing an algorithm that encompasses these

changes.

5.2 Future Work

5.2.1 Develop Algorithm to Find Best Quadratic Lyapunov Function

Based on the preliminary results of trying to find larger quadratic functions, the next step is to develop an algorithm that will find the best shape and orientation of the quadratic ROA estimate based on the system. Since this algorithm has the potential to be computationally intensive, we will look at using a genetic algorithm to determine which angles and eigenvalues will be used to find the largest quadratic Lyapunov function. The goal of the genetic algorithm will be to narrow in on the best values without having to run through every single value.

5.2.2 Increase ROA Estimate

The long term goal of my research is to be able to determine how climate change affects the stability of physical characteristics in plants. In order for this to be accomplished, however, there must be significant improvement on the current methods of estimating regions of attraction to get an accurate representation of how the regions of attraction of the phenotypes are affected by changes in the environment. The current elliptical estimates found using quadratic Lyapunov functions are generally too conservative for the information to be very useful. One possible method that has been briefly explored by researchers include looking at the union of several quadratic Lyapunov functions [11][40]. While more research into this area is deserved, it is believed that the results will still

be too conservative. Another idea is to use concepts from computational geometry such as free-form deformation and morphing to expand a quadratic region of attraction into a larger region depending on the system flow. This would mean that the ROA estimate would no longer be constrained to a quadratic function allowing for the potential of a much larger estimate of the ROA. Figure 5.1 shows an example of how this expansion might occur using a geometrical mesh and expanding individual points on the boundary of the region until the trajectories begin leaving the region.

Figure 5.1: Using Free-Form Deformation to Expand an ROA estimate

REFERENCES

- [1] I. Ahuja, R.C.H. De Vos, A.M. Bones, and R.D. Hall. Plant molecular stress responses face climate change. *Trends in plant science*, 15(12):664–674, 2010.
- [2] R. Albert and H.G. Othmer. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster. *Journal of Theoretical Biology*, 223(1):1–18, 2003.
- [3] D. Angeli, J.E. Ferrell, and E.D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7):1822, 2004.
- [4] Hervé Brönnimann, Guillaume Melquiond, and Sylvain Pion. The design of the boost interval arithmetic library, 2006.
- [5] E.C. Butcher, E.L. Berg, and E.J. Kunkel. Systems biology in drug discovery. *Nature biotechnology*, 22(10):1253–1259, 2004.
- [6] G. Chesi. Computing output feedback controllers to enlarge the domain of attraction in polynomial systems. *Automatic Control, IEEE Transactions on*, 49(10):1846–1853, 2004.
- [7] G. Chesi. Estimating the domain of attraction for uncertain polynomial systems. *Automatica*, 40(11):1981–1986, 2004.
- [8] G. Chesi. Estimating the domain of attraction for non-polynomial systems via lmi optimizations. *Automatica*, 45(6):1536–1541, 2009.
- [9] G. Chesi. Robustness analysis of genetic regulatory networks affected by model uncertainty. *Automatica*, 47(6):1131–1138, 2011.
- [10] G. Chesi, L. Chen, and K. Aihara. On the robust stability of time-varying uncertain genetic regulatory networks. *International Journal of Robust and Nonlinear Control*, 2011.
- [11] Graziano Chesi. Estimating the domain of attraction via union of continuous families of lyapunov estimates. *Systems; Control Letters*, 56(4):326–333, 2007.
- [12] D. Daney, Y. Papegay, and A. Neumaier. Interval methods for certification of the kinematic calibration of parallel robots. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 2, pages 1913–1918. IEEE, 2004.

- [13] M.I. Davidich and S. Bornholdt. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS One*, 3(2):e1672, 2008.
- [14] P. Érdi and J. Tóth. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press ND, 1989.
- [15] A. Fauré, A. Naldi, C. Chaouiya, and D. Thieffry. Dynamical analysis of a generic boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):e124–e131, 2006.
- [16] E.D. Ferreira and B.H. Krogh. Using neural networks to estimate regions of stability. In *American Control Conference, 1997. Proceedings of the 1997*, volume 3, pages 1989–1993. IEEE, 1997.
- [17] P. Gahinet, A. Nemirovskii, A.J. Laub, and M. Chilali. The lmi control toolbox. In *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, volume 3, pages 2038–2041, dec 1994.
- [18] R. Genesio, M. Tartaglia, and A. Vicino. On the estimation of asymptotic stability regions: State of the art and new proposals. *Automatic Control, IEEE Transactions on*, 30(8):747–755, 1985.
- [19] L. Jaulin. *Applied interval analysis: with examples in parameter and state estimation, robust control and robotics*, volume 1. Springer Verlag, 2001.
- [20] H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.
- [21] E. Klipp. *Systems biology in practice: concepts, implementation and application*. Vch Verlagsgesellschaft MbH, 2005.
- [22] A.K. Konopka. *Systems biology: principles, methods, and concepts*. Crc, 2007.
- [23] M. Lhommeau, L. Jaulin, and L. Hardouin. Capture basin approximation using interval analysis. *International Journal of Adaptive Control and Signal Processing*, 25(3):264–272, 2011.
- [24] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4781, 2004.
- [25] Y. Lin and M.A. Stadtherr. Deterministic global optimization of molecular structures using interval analysis. *Journal of computational chemistry*, 26(13):1413–1420, 2005.

- [26] S. Mayer, B. Tibken, and S. Warthenpfohl. Parallel computation of domains of attraction for nonlinear dynamic systems. In *Parallel Computing in Electrical Engineering (PARELEC), 2011 6th International Symposium on*, pages 87–92. IEEE, 2011.
- [27] A. Merola, C. Cosentino, and F. Amato. An insight into tumor dormancy equilibrium via the analysis of its domain of attraction. *Biomedical Signal Processing and Control*, 3(3):212–219, 2008.
- [28] R.E. Moore. *Interval analysis*, volume 60. Prentice-Hall Englewood Cliffs, NJ, 1966.
- [29] R.E. Moore, R.B. Kearfott, and M.J. Cloud. *Introduction to interval analysis*. Society for Industrial Mathematics, 2009.
- [30] A. Mukhopadhyay, A.M. Redding, B.J. Rutherford, and J.D. Keasling. Importance of systems biology in engineering microbes for biofuel production. *Current opinion in biotechnology*, 19(3):228–234, 2008.
- [31] J.R. Nevins. The rb/e2f pathway and cancer. *Human Molecular Genetics*, 10(7):699–703, 2001.
- [32] J.C. Polking. Rice university, pplane, 2009.
- [33] C.V. Rao and A.P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm. *The Journal of chemical physics*, 118:4999, 2003.
- [34] S. Ratschan and Z. She. Providing a basin of attraction to a target region by computation of lyapunov-like functions. In *Computational Cybernetics, 2006. ICCCC 2006. IEEE International Conference on*, pages 1–5. IEEE, 2006.
- [35] P. Sannigrahi, A.J. Ragauskas, and G.A. Tuskan. Poplar as a feedstock for biofuels: A review of compositional characteristics. *Biofuels, Bioproducts and Biorefining*, 4(2):209–226, 2010.
- [36] M.A. Stadtherr, G. Xu, G.I. Burgos-Solorzano, and W.D. Haynes. Reliable computation of phase stability and equilibrium using interval methods. *International Journal of Reliability and Safety*, 1(4):465–488, 2007.
- [37] Steven H Strogatz. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry And Engineering*. Westview Press, 1994.
- [38] J. Swana, Y. Yang, M. Behnam, and R. Thompson. An analysis of net energy production and feedstock availability for biobutanol and bioethanol. *Bioresource technology*, 102(2):2112–2117, 2011.

- [39] B. Tibken. Estimation of the domain of attraction for polynomial systems via lmis. In *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, volume 4, pages 3860–3864. IEEE, 2000.
- [40] B. Tibken and Y. Fan. Computing the domain of attraction for polynomial systems via bmi optimization method. In *American Control Conference, 2006*, pages 6–pp. IEEE, 2006.
- [41] U. Topcu, A. Packard, and P. Seiler. Local stability analysis using simulations and sum-of-squares programming. *Automatica*, 44(10):2669–2675, 2008.
- [42] W. Tucker. The lorenz attractor exists. *Comptes Rendus de l’Académie des Sciences-Series I-Mathematics*, 328(12):1197–1202, 1999.
- [43] J.J. Tyson, K. Chen, B. Novak, et al. Network dynamics and cell physiology. *Nature Reviews Molecular Cell Biology*, 2(12):908–916, 2001.
- [44] AR Tzafiriri. Michaelis-menten kinetics at high enzyme concentrations. *Bulletin of mathematical biology*, 65(6):1111–1129, 2003.
- [45] S. Warthenpfehl, B. Tibken, and S. Mayer. An interval arithmetic approach for the estimation of the domain of attraction. In *Computer-Aided Control System Design (CACSD), 2010 IEEE International Symposium on*, pages 1999–2004, sept. 2010.
- [46] R.A. Weinberg. *The biology of cancer*, volume 255. Garland Science New York, 2007.
- [47] Dominik Wittmann, Jan Krumsiek, Julio Saez-Rodriguez, Douglas Lauffenburger, Steffen Klamt, and Fabian Theis. Transforming boolean models to continuous models: methodology and application to t-cell receptor signaling. *BMC Systems Biology*, 3(1):98, 2009.
- [48] Dominik M. Wittmann, Florian Blöchl, Dietrich Trümbach, Wolfgang Wurst, Nilima Prakash, and Fabian J. Theis. Spatial analysis of expression patterns predicts genetic interactions at the mid-hindbrain boundary. *PLoS Comput Biol*, 5(11):e1000569, 11 2009.
- [49] Stephen Wolfram. *Mathematica for students*.
- [50] G. Yao, C. Tan, M. West, J.R. Nevins, and L. You. Origin of bistability underlying mammalian cell cycle entry. *Molecular systems biology*, 7(1), 2011.